Understanding and Harnessing Energy-Dependent Proteolysis for Controlled Protein Degradation in Bacteria

by

Joseph H. Davis

B.A. Computer Science
B.S. Biological Engineering
University of California, Berkeley, 2003

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

APRIL 2010

Signature of Author: _____

Joseph H. Davis
Department of Biology
April 12, 2010

Certified by: _____

Robert T. Sauer
Salvador E. Luria Professor of Biology
Thesis supervisor

Accepted by: _____

Stephen P. Bell
Professor of Biology
Co-Chair, Biology Graduate Committee

Understanding and Harnessing Energy-Dependent Proteolysis for Controlled Protein Degradation in Bacteria

by

Joseph H. Davis

Submitted to the Department of Biology
on April 12, 2010 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biology at the Massachusetts Institute of Technology

ABSTRACT

Regulated intracellular protein degradation is critical for cellular viability. In many organisms, degradation controls cell-cycle progression, executes responses to stress-inducing environmental changes, and enables the rapid depletion of unwanted or deleterious proteins. In bacteria, most processive protein degradation is carried out by a family of AAA+ compartmentalized proteases. These molecular machines convert the chemical energy of ATP binding and hydrolysis into mechanical work, forcefully unfolding their substrates as a prelude to proteolysis.

The AAA+ ClpXP protease, recognizes short peptide tags (degrons) in substrate proteins either directly or with the aid of dedicated specificity factors (adaptors). The prior identification and detailed biochemical characterization of an efficient ClpXP degron (the ssrA tag) and cognate adaptor (SspB) serve as powerful tools and enable the mechanistic studies presented here.

In Chapter 2, I describe a collaborative investigation of substrate denaturation and degradation by ClpXP with single-molecule resolution. Detailed kinetic analysis of these experiments revealed homogenous protease activity across the population of enzymes with comparable levels of microscopic and macroscopic ClpXP activity. These experiments required the development of methods to attach ClpXP to surfaces and stabilize the multimeric enzyme at sub-nanomolar concentrations, advances that should be applicable to future single-molecule studies of complex protein machines.

Subsequent chapters describe the development of molecular tools that harness our understanding of targeted proteolysis and enable small-molecule control of degradation. By engineering synthetic substrates, adaptors and proteases, I directly test models previously proposed to explain adaptor function and identify the minimal requirements for adaptor-mediated substrate delivery. Many different configurations of protease and adaptor domains lead to efficient, predictable substrate degradation and demonstrate the highly modular nature of this system. These tools allow for facile, small-molecule controlled protein degradation *in vivo* and should be valuable in basic research and biotechnology. I also describe a family of synthetic insulated promoters that allow predictable, context-independent levels of protein synthesis.

Thesis Supervisor: Robert T. Sauer
Title: Salvador E. Luria Professor of Biology

ACKNOWLEDGMENTS

For their support, friendship and advice I would like to give my heartfelt thanks to:

My advisor, Bob Sauer, for giving me the freedom to pursue projects of my choosing and the guidance, insight and encouragement to help me see them through to the end. Tania Baker, my co-advisor, for her enthusiasm and advice throughout my time at MIT.

My thesis committee members, Phil Sharp and Mike Laub whose input through committee meetings and informal, impromptu discussions has helped guide my thesis. Frank Solomon for his friendship, support and mentorship throughout my time at MIT. Michael Marletta for taking me into his lab as an undergraduate and teaching me how much fun science can be. Michael Eck for serving on my thesis defense committee.

Members of the Sauer lab, past and present, for creating a rich and challenging scientific environment. My friend and colleague, Andreas Martin whose intellect and fortitude drove me to become a better scientist. Kathleen McGinness for her constant friendship and healthy perspective.

My fantastic single-molecule collaborators, Matt Lang, Ricardo Brau, Yongdae Shin, Marie-Eve Aubin-Tam and Adrian Olivares for helping make this project a (ongoing) success.

My friends, classmates and housemates, Kurt Weiss, Carly Cassano, Jadyn Damon, Ryan and Ellie Graeden, Aaron Reinke, Leah Okumura, Jason Kelly, Steven Glynn, Christos Tsokos, Julie Mumford, Victor Lelyveld, Justin Pritchard, and Shahriar Khushrushahi for giving me a reason to leave the pipettes behind and go skiing/climbing/traveling/concerting/camping/drinking/ bantering/diving/footballing/cooking.

My father, Joe Davis, and my friend, Patty Simonson, for their encouragement, their confidence in me, and their warm reminders that no matter how far I go, my home will always be in California.

Finally, my closest friend and confidant, Jessica Chang. I cannot thank you enough for the joy that you bring me on a daily basis.

To my father, for his unwavering support, encouragement, and friendship.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter Five : Small-molecule Controlled Proteolysis in *E. Coli*

Appendix A : Supplementary Information for "Single-molecule Denaturation and Degradation of Proteins by the AAA+ ClpXP Protease"

Appendix B : Single-Molecule FRET Assays for the Study of the AAA$^+$ Protease ClpXP

Appendix C : Measuring the Activity of BioBrick Promoters Using an *in vivo* Reference Standard

Chapter 1

An Introduction to Engineering Energy-Dependent Proteolysis in Bacteria

A cell's ability to maintain proper steady-state protein levels is vital for survival. In some instances, the ratio of one protein to another is critical, whereas in other cases, the absolute concentration is important. Enzymatic activity, for example, is a function of both the cellular concentration of the enzyme and its specific activity. Indeed, cells control total activity both through the regulation of steady-state enzyme concentrations as well as by dynamically modulating enzyme specific activity using post-translational modifications, inhibitors, and activators. Descriptions of such post-translational regulation can be found elsewhere (Cohen, 2002; Mann and Jensen, 2003). This chapter will instead focus on how intracellular protein concentrations are controlled both naturally and experimentally. In particular, I will emphasize known mechanisms of bacterial proteolysis and how such systems can be utilized to control protein degradation in basic research and biotechnology.

A protein's steady-state concentration is a function of both its rate of production and its rate of degradation. Briefly, cellular RNA polymerase transcribes mRNA from the genome. Balance between mRNA production and the combined effects of degradation by RNAses and dilution via growth results in a steady-state pool of mRNA that serves as a template for protein synthesis. New protein molecules, synthesized via translation of mRNA, are eventually degraded or diluted through cell growth and division. This entire process can be modeled using ordinary differential equations as shown in Equations 1.1-1.2. Under steady-state conditions, *i.e.* when the time-dependent changes in mRNA and protein concentrations are zero, the intracellular concentration of a given protein can be expressed as a function of its basic rate constants for transcription, mRNA degradation, growth rate, translation, and protein degradation (see Equation 1.3).

$$\frac{d[mRNA]}{dt} = k_{transcription} - \left(k_{mRNADeg} + k_{growt\,h}\right)[mRNA] = 0 \qquad (1.1)$$

$$[mRNA] = \frac{k_{transcription}}{k_{mRNADeg} + k_{growt\,h}}$$

$$\frac{d[protein]}{dt} = k_{translation}\ [mRNA] - \left(k_{proteinDeg} + k_{growt\,h}\right)[protein] = 0 \qquad (1.2)$$

$$[protein] = \frac{k_{translation}\ [mRNA]}{k_{proteinDeg} + k_{growt\,h}}$$

$$[protein] = \frac{k_{translation}\ \dfrac{k_{transcription}}{k_{mRNADeg} + k_{growt\,h}}}{k_{proteinDeg} + k_{growt\,h}} \qquad (1.3)$$

Each rate in this process (transcription, mRNA degradation, translation and protein degradation) has been carefully tuned by the process of evolution. For example, the rates of transcription and translation are strongly dependent on sequence elements located in the promoter of the gene and ribosome binding site of the mRNA (Pribnow, 1975; Shine and Dalgarno, 1975). To regulate the rates of mRNA degradation, cellular RNAses have coevolved with mRNA coding sequences selecting for the presence or absence of cleavage sites (Kuwano *et al.*, 1977). Proteins and proteases have evolved under similar selective pressure, incorporating protease recognition sites in proteins that must be rapidly degraded (Neher *et al.*, 2006; Chien *et al.*, 2007). Lastly, a protein's thermodynamic stability often correlates with its susceptibility to proteolysis (Parsell and Sauer, 1989), and thus thermodynamic stability may be under selective pressure not only as a determinant of protein folding but also for its role in protein degradation. In *E. coli*, such selective evolution has given rise to a range of intracellular protein concentrations spanning four orders of magnitude (Lu *et al.*, 2007).

Cells must also dynamically regulate protein levels in response to changing environmental conditions. Powerful techniques exist to directly measure a cell's genome-wide transcriptional response to environmental perturbations, allowing for identification of the proteins whose transcription is regulated in response to such environmental perturbations (DeRisi *et al.*, 1997; Causton *et al.*, 2001). The bacterial *lac* operon, for example, is exquisitely sensitive to the presence of lactose. In the absence of this sugar, a transcriptional repressor, LacI, forms a tight complex with the *lac* operator thereby blocking transcription of the downstream genes, *lac*Z, *lac*Y, and *lac*A. When lactose is imported into the cell, it binds to the repressor, and induces a conformational change that results in dissociation of LacI from the promoter. This process allows RNA polymerase to transcribe the genes required for the lactose metabolism (Wilson *et al.*, 2007). In Chapter 5, I demonstrate that targeted degradation of LacI can also be used to induce transcription of the *lac* operon.

In addition to regulation of protein-production, evidence has steadily accumulated that protein degradation also plays a critical role in the cellular response to changing environments. In 1942, Rudolf Scheonheimer observed degradation and recycling of human proteins using ^{15}N labeled amino acids (Ciechanover, 2005). Since then, controlled protein degradation was found to play a critical role many processes including cell-cycle progression, the response to stress-inducing environmental changes, inhibition of viral infection, and the maintenance of protein homeostasis through the removal of unwanted or deleterious proteins (Murray *et al.*, 1989; Jenal and Fuchs, 1998; Dougan *et al.*, 2002; Neher *et al.*, 2003; Sakuma *et al.*, 2007; Abdelmohsen *et al.*, 2009). In each of these instances, protein degradation is used in concert with additional regulatory mechanisms to efficiently control intracellular protein concentrations.

Our understanding of the biological mechanisms that control protein concentrations has also led to the development of experimental methods to change protein levels artificially by altering the rates of protein production or degradation. Such targeted approaches often complement classical "forward-genetics" (*i.e.* generating random mutations, screening for a phenotype, and eventually mapping the mutation responsible for the phenotype) in the study of biological systems. For genes with a proposed function, targeted techniques provide a means to test specific models in an organismal context. For genes of unknown activity, comparing the phenotypes of an experimentally altered strain and its parent strain may illuminate function.

Techniques to experimentally control protein levels have also been applied in biological engineering. Applications include the engineering of synthetic genetic circuits and metabolic pathways (Elowitz and Leibler, 2000; Gardner *et al.*, 2000; Martin *et al.*, 2003; Kobayashi *et al.*, 2004; Ro *et al.*, 2006; Atsumi and Liao, 2008). In each case, optimization of steady-state protein concentrations was critical for successful implementation and was achieved using a combination of techniques to tune steady-state protein concentrations.

Techniques for genomic perturbations

Numerous overexpression techniques can be used to increase intracellular protein concentrations (Lloyd, 2003; Kitagawa *et al.*, 2005; Sopko *et al.*, 2006; Campbell and Brown, 2008). Here, I will focus on the variety of protein knockdown techniques currently available for use in eukaryotic and prokaryotic organisms. Each technique offers different compromises between ease of application, specificity for a given target, efficiency, ability to control the knockdown in a conditional manner, and the speed of the temporal response.

For nonessential targets in genetically tractable organisms, the genetic locus can be directly perturbed (Baba *et al.*, 2006). In bacteria for example, transiently expressed phage proteins facilitate homologous recombination between linear dsDNA and the chromosome. This approach, termed recombineering, allows for more efficient bacterial genetic manipulation than is possible using endogenous *rec*A-mediated recombination (Court *et al.*, 2002). Recombineering makes use of a targeting construct containing the desired alteration (mutation, insertion or deletion) fused to a selectable marker and surrounded by short (~40 bp) overhangs homologous to the region of interest. After induction of the phage recombination factors, the linearized targeting cassette is electroporated into cells and recombination incorporates the entire construct into the genome, replacing the targeted loci between the homologous regions. Cells are then plated on selective media to identify recombinants (Figure 1.1A,B).

If a cassette containing both a selectable and conditionally counter-selectable marker is first introduced at a locus of interest, it is possible to perform a second round of recombineering, which simultaneously introduces a genetic perturbation and removes the initial cassette (Lee *et al.*, 2001b). This technique allows for the "scarless" manipulation of the genome as illustrated in Figure 1.1C. In the *Methods* section of Chapter 5, I describe this technique and the novel use of the marker mPheS to perform the counter-selectable recombineering step. Because the phage recombination machinery facilitates homologous recombination so readily, robust methods have been developed to closely control phage protein expression and thus avoid genomic instability (Yu *et al.*, 2000; Datta *et al.*, 2006).

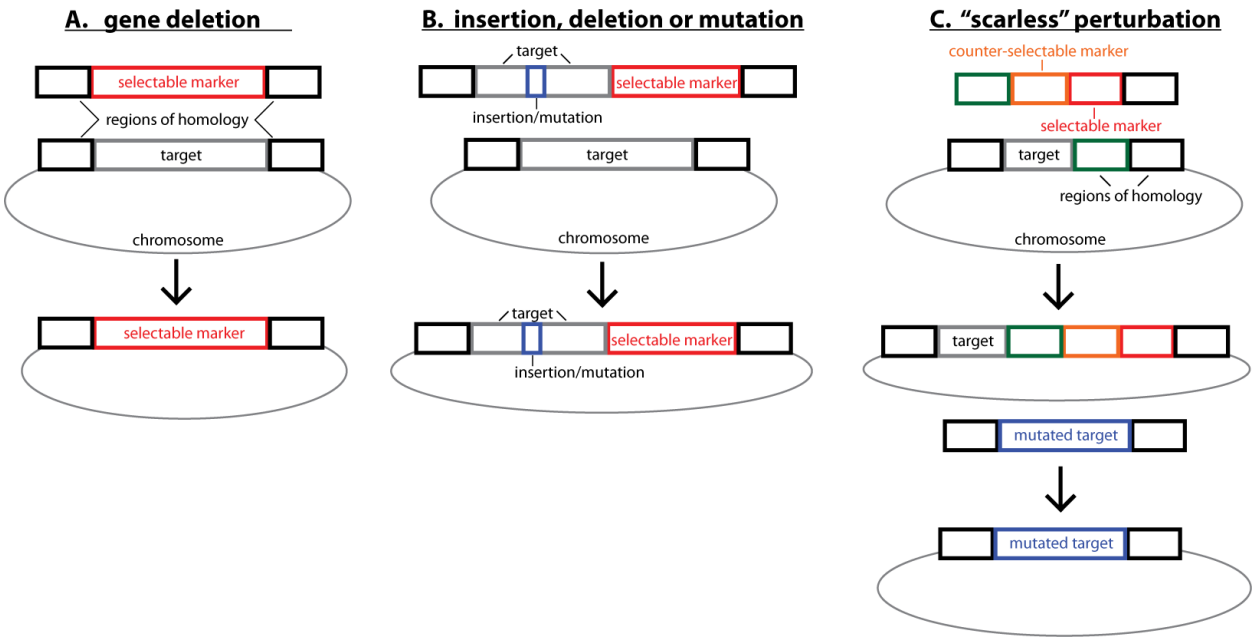

Figure 1.1. Schematic of homologous recombination-based techniques for genomic manipulation. Rectangles represent dsDNA and ovals represent bacterial chromosomes. Black and green squares are regions of DNA with a minimum of 40 base pairs of sequence homology. Recombination events are represented by black arrows. (**A)** Gene deletions are obtained by fusing a selectable marker (an antibiotic resistance cassette for example) to overhangs homologous to sequences residing 5´ and 3´ of the gene target. **(B)** By fusing a mutated variant of the targeted sequence to a selectable marker, mutations, insertions, and deletions can be easily isolated. **(C)** Using sequential rounds of phage-mediated recombination, one can generate "scarless" mutations. First, a cassette bearing both a selectable and conditionally counter-selectable marker is recombined into the genome without disturbing the gene target. Next, this entire cassette is replaced with a desired insertion. After recombination, cells are plated on selective media such that those retaining the counter-selectable marker are killed. Only clones that have successfully undergone this second round of recombination survive.

The utility of this technique is derived from its specificity (a single nucleotide can be mutated in the context of an entire genome) and its general applicability (mutations can target virtually any genetically encoded function). For essential genes, however, this direct manipulation can result in a lethal phenotype which generally precludes further characterization or utilization (Knight and Shokat, 2007). Even for nonessential genes, investigating the mutant phenotypes on short-time scales is challenging, as the isolation of mutated strains generally requires multiple days and many cell generations. During this time, compensatory pathways may be up-regulated or

suppressor mutations may arise, thereby masking the direct effect of the perturbation (Gerdes *et al.*, 2003).

Conditional knockdown techniques

To overcome limitations associated with gene replacement or mutation, conditional knockdown techniques have been developed. Some approaches (such as introducing a regulated promoter, a temperature-sensitive mutation, or a degradation tag) rely on the availability of techniques to manipulate the target organism's genome (Gossen and Bujard, 1992; Couso *et al.*, 1993; Furth *et al.*, 1994; Varadarajan *et al.*, 1996; McGinness *et al.*, 2006; Griffith and Grossman, 2008). Other methods, including RNA interference and chemical inhibitors can often be used in organisms without developed techniques for genetic manipulation. Importantly, however, all of these techniques give researchers temporal control over initiation of the perturbation.

The use of a regulated promoter is a straightforward method to control transcription of a target gene artificially. A number of different systems have been developed. In each, a gene's native promoter is replaced with a transcription-factor regulated promoter whose activity can be controlled by a small molecule, temperature, or light (Sussman and Jacob, 1962; Mieschendahl and Muller-Hill, 1985; Furth *et al.*, 1994; Strickland *et al.*, 2008; Moglich *et al.*, 2009). Tetracycline-controlled systems derived from the bacterial TetR repressor are particularly well developed for use in eukaryotic cells (Bertram and Hillen, 2008). In some variants, addition of a small molecule down-regulates transcription; in others, the small molecule leads to transcriptional activation (Gossen and Bujard, 1992; Gossen *et al.*, 1995). Similar systems have been developed for use in bacteria. For example, synthetic promoters derived from the lactose

operon have been used to control gene expression using the lactose analog, IPTG. As described above, addition of the inducer upregulates transcription by freeing the promoter of the LacI repressor (Jensen *et al.*, 1993).

Despite many reports of success using synthetic, regulated promoters, implementation can be difficult for several reasons. When the native promoter is replaced, researchers must often try many variants of the synthetic promoter to identify a sequence that recapitulates wild-type rates of transcription in the absence of the inducer. Moreover, proteins whose native transcription is dynamic (*e.g.*, cell-cycle regulated genes) are extremely difficult to study using these synthetic promoters, as recreating the complex temporal profile of wild-type transcription is virtually impossible. Lastly, because polycistronic genes are common in bacteria, it can be challenging to specifically perturb one gene in a multi-protein operon. To overcome this, the gene of interest must be knocked out and then reintroduced at another locus. In some cases, the gene can be introduced with its native promoter on a plasmid bearing a temperature-sensitive origin of replication. Cells grown under permissive conditions produce the gene product of interest. Upon shifting to the restrictive temperature, plasmid replication is inhibited, resulting in plasmid dilution and eventual loss (Jasin and Schimmel, 1984; Pyne and Bognar, 1992; Silo-Suh *et al.*, 2009).

As an alternative to controlling transcription, translation can be perturbed using RNA interference (RNAi) and antisense RNA. Unlike the techniques described above, these approaches allow for use of native transcriptional control elements. RNAi occurs when mRNA is specifically targeted for degradation either by injecting or controlling the endogenous

transcription of short RNA sequences that are complementary to the mRNA of interest. Using Watson-Crick base pairing, these short interfering RNAs (siRNAs) hybridize to the mRNA of interest, resulting in an RNA duplex. In protozoa and higher eukaryotes, a conserved response to such duplex RNA is initiated which results in cleavage and degradation of the targeted mRNA (Sharp, 2001; McManus and Sharp, 2002; Dorsett and Tuschl, 2004). As such, translation is precluded, effectively "knocking down" production of the gene product of interest.

Less commonly, translational inhibition techniques have been applied in bacterial species (Ji *et al.*, 2001; Croxen *et al.*, 2007; Gillaspie *et al.*, 2009). Production of short RNAs, which specifically bind to the mRNA translation initiation region and occlude binding of the 30S ribosome to the message, can be used to effectively block translation (Waters and Storz, 2009). Further, because the mRNA is not actively translated, it is more susceptible to degradation, thereby enhancing the silencing effect.

The efficacy of the transcriptional- and translational-repression techniques described above is strongly dependent on the cell's ability to degrade pre-existing protein products. For stable proteins with long half-lives, simply terminating production is not sufficient to give rise to a knockout phenotype. The pre-existing molecules will continue to function despite the termination of synthesis. Thus, several generations may be required for the protein to be diluted to a concentration that prevents function. This fact precludes observing phenotypes on time-scales shorter than the half-life of the protein. In response to these limitations, techniques to directly target the pre-existing protein products have been developed.

Directly targeting pre-existing molecules

For some proteins, mutations can be isolated that greatly decrease the level of activity at one temperature (called a restrictive temperature) but show minimal effect at a different temperature (termed the permissive temperature) (Chakshusmathi *et al.*, 2004). In some instances, the temperature shift causes global protein unfolding, often resulting in the display of protease recognition sites and eventual degradation. For other mutations, the temperature shift impedes catalysis but not folding (Bolhuis *et al.*, 1999). These mutants are reversibly inactivated, recovering activity upon a shift back to the permissive temperature. Still other mutants are only deficient for folding or macromolecular assembly at restrictive temperatures but, if first assembled at the permissive temperature, are completely functional after the temperature shift (Sadler and Novick, 1965; Smith *et al.*, 1980). In such cases, dilution of pre-existing molecules by sustained cell growth and division at the restrictive temperature eventually leads to a loss-of-function phenotype.

Historically, temperature-sensitive mutants (*ts*-mutants) were isolated by randomly mutagenizing an entire genome, screening for temperature-sensitive phenotypes, and mapping the mutations to a particular gene (Edgar and Lielausis, 1964; Hartwell, 1967). To isolate a *ts*-mutant for a particular gene, one can instead mutagenize the isolated locus and assay for temperature-dependent function. The ability to perform such an activity assay requires some understanding of the gene product and is thus inapplicable to the study of genes of unknown function. Even when appropriate assays are available, screening the enormous number of possible mutants can be overwhelming. To expedite this search, computational methods have been developed to help predict *ts*-mutants from primary sequence (Varadarajan *et al.*, 1996; Chakshusmathi *et al.*, 2004).

Once a *ts*-mutant has been identified and characterized, the strain bearing the mutation can be generated and grown at permissive temperatures. After shifting cells to the restrictive temperature, one can directly measure the phenotype associated with the loss of activity from the *ts*-mutant. In many cases, however, the global changes concomitant with a temperature shift can complicate interpretation. In yeast, for example, a temperature shift from 25 °C to 37 °C results in the altered expression of 854 genes, half of which have unknown function (Causton *et al.*, 2001). Further complicating experimental design and interpretation, mutations that phenocopy a wild-type strain at the permissive temperature and a null strain at the restrictive temperature are rare. Despite these shortcomings, successful analysis of *ts*-mutants has guided the understanding of essential gene function in a variety of organisms (Edgar and Lielausis, 1964; Eidlic and Neidhardt, 1965; Hartwell, 1967).

Pharmacology, the use of small molecules to perturb biological function, offers another approach to target pre-existing proteins directly. A large variety of natural and synthetic molecules have been identified that are potent (*i.e.*, efficacious at low concentrations), specific (*i.e.*, alter the activity of only one protein in the context of a complete proteome), and cell-permeable. However, for many proteins, the identification of small-molecule inhibitors and activators is not trivial. Moreover, conclusively ruling out "off-target" effects is a substantial challenge, as such compounds often act on an entire class of similar proteins (Knight and Shokat, 2007).

In response to these limitations, Shokat and colleagues have developed a technique that targets a particular kinase by mutating the active site to allow binding of a complementary inhibitor. Cellular kinases that lack this mutation exclude inhibitor binding, providing genetically encoded

specificity as well as pharmacological temporal control (Bishop *et al.*, 1998). Unfortunately, this "bump-hole" strategy requires substantial engineering and is currently limited to isolated members of the kinase superfamily.

Harnessing the power of directed proteolysis could circumvent some of the aforementioned problems. Various controlled degradation systems have been developed for use in both prokaryotes and eukaryotes. Although the details differ, these systems share several common properties. In most cases, the endogenous protein coding sequence is modified by fusing a degradation tag ("degron") to either the N- or C-terminus of the targeted protein. Ideally, under permissive conditions, this tag neither perturbs the natural function of the protein nor targets it for degradation. Under restrictive conditions, the tag should be recognized by intracellular proteases, leading to degradation. In the following section, I discuss our current understanding of energy-dependent proteolysis and provide examples of how these systems have been utilized to degrade target substrates.

Energy-dependent proteases

The majority of intracellular proteolysis in eukaryotes and prokaryotes is carried out by a group of energy-dependent proteases of the AAA^+ family (Hanson and Whiteheart, 2005). These enzymes form ring oligomers composed of an unfoldase domain (the AAA+ domain) and a protease domain (Sauer *et al.*, 2004). The bacterial Lon and FtsH proteases encode the unfoldase and protease domains as a single polypeptide chain. In contrast, the bacterial ClpXP, ClpAP, HslUV proteases and the eukaryotic 26S proteasome utilize separate proteins for each function

(Gottesman, 1996; Okuno *et al.*, 2006; Murata *et al.*, 2009). In the latter group, subunits of the unfoldase and protease assemble to form the proteolytic complex.

ClpXP, one of the best characterized energy-dependent proteases, consists of a tetradecameric compartmentalized protease, ClpP, which is capped by a hexameric unfoldase, ClpX (Sauer *et al.*, 2004). As shown in Figure 1.2, the cylindrical protease has an axial channel that runs from the top of ClpX to the sequestered active-site residues of ClpP. This pore is small relative to the size of a folded protein, and thus only denatured polypeptides are allowed access to the sequestered proteolytic sites. In this regard, ClpX acts as a gatekeeper, selecting which cellular proteins will be targeted for cleavage by the active sites of ClpP.

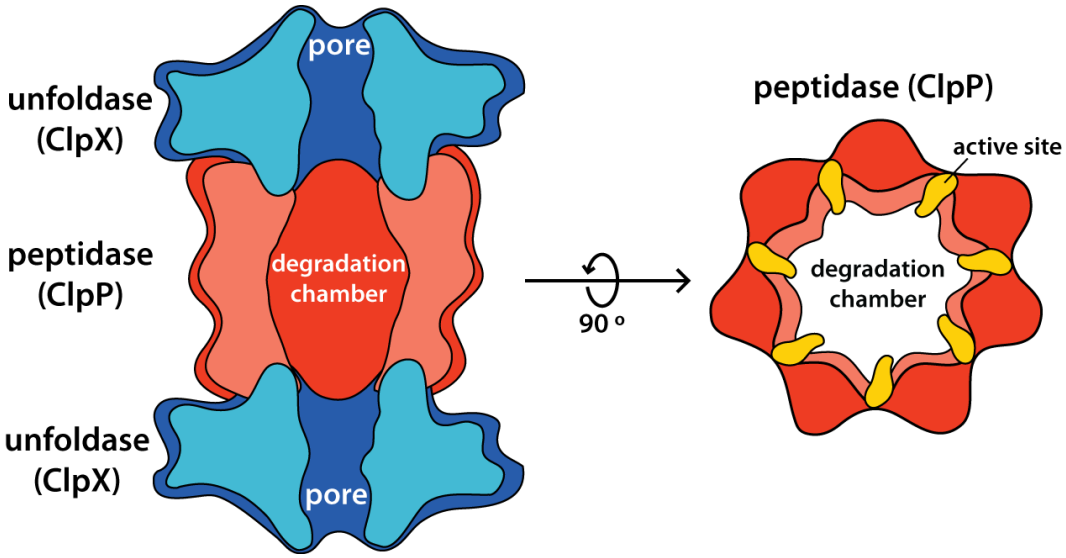

Figure 1.2. Schematic representation of a compartmentalized protease. ClpXP, a representative compartmentalized protease, is composed of the $ClpP_{14}$ peptidase shown in red, and the two $ClpX_6$ unfoldases, shown in blue. The axial pore that runs the length of the complex is too narrow for folded substrates to enter the degradation chamber. Native proteins are first unfolded in an ATP-dependent reaction by ClpX. After translocation to the degradation chamber, substrates are hydrolyzed to short peptides (~10 amino acids) by the active sites of ClpP (yellow). Critically, these active sites are sequestered from solution and thus only act on substrates that have been unfolded and translocated by ClpX.

In some instances, such as the ubiquitin-proteasome system, substrate selection by the proteases can be a complicated process, involving many accessory factors and post-translational modification of the substrate (Hershko and Ciechanover, 1998). In other instances, such as the ssrA-tagging system in bacteria, substrates are targeted for degradation through the simple co-translational addition of a short, C-terminal peptide degradation tag (Keiler *et al.*, 1996). Detailed biochemical studies have shown that ClpX binds to such exposed degradation tags using a set of loops that line the pore of the unfoldase (Farrell *et al.*, 2007; Martin *et al.*, 2008a). ATP binding and hydrolysis are used to drive conformational changes of these loops, resulting in engagement and translocation of the bound substrate down through the pore (Kenniston *et al.*, 2004; Martin *et al.*, 2008a; Glynn *et al.*, 2009). This translocation is thought to pull the folded substrate against the body of ClpX, and thus apply an unfolding force. Successful substrate denaturation results from a combination of the applied pulling force and stochastic changes in local protein stability (Kenniston *et al.*, 2003). For example, if thermally-induced fraying of nearby substrate secondary structure coincides with a pulling event, then complete unfolding of the substrate is more probable. Once globally denatured, the substrate is rapidly translocated, precluding refolding. Interestingly, for some substrates, the process of denaturation is rate-limiting for overall proteolysis, indicating that the probability of a pulling event resulting in denaturation is low. For substrates that are readily unfolded, translocation into the lumen of ClpP is the rate-limiting step, indicating that the probability of an unfolding event is relatively high (Kenniston *et al.*, 2004). The process of substrate binding, engagement, denaturation, translocation, and eventual proteolysis is depicted in Figure 1.3.

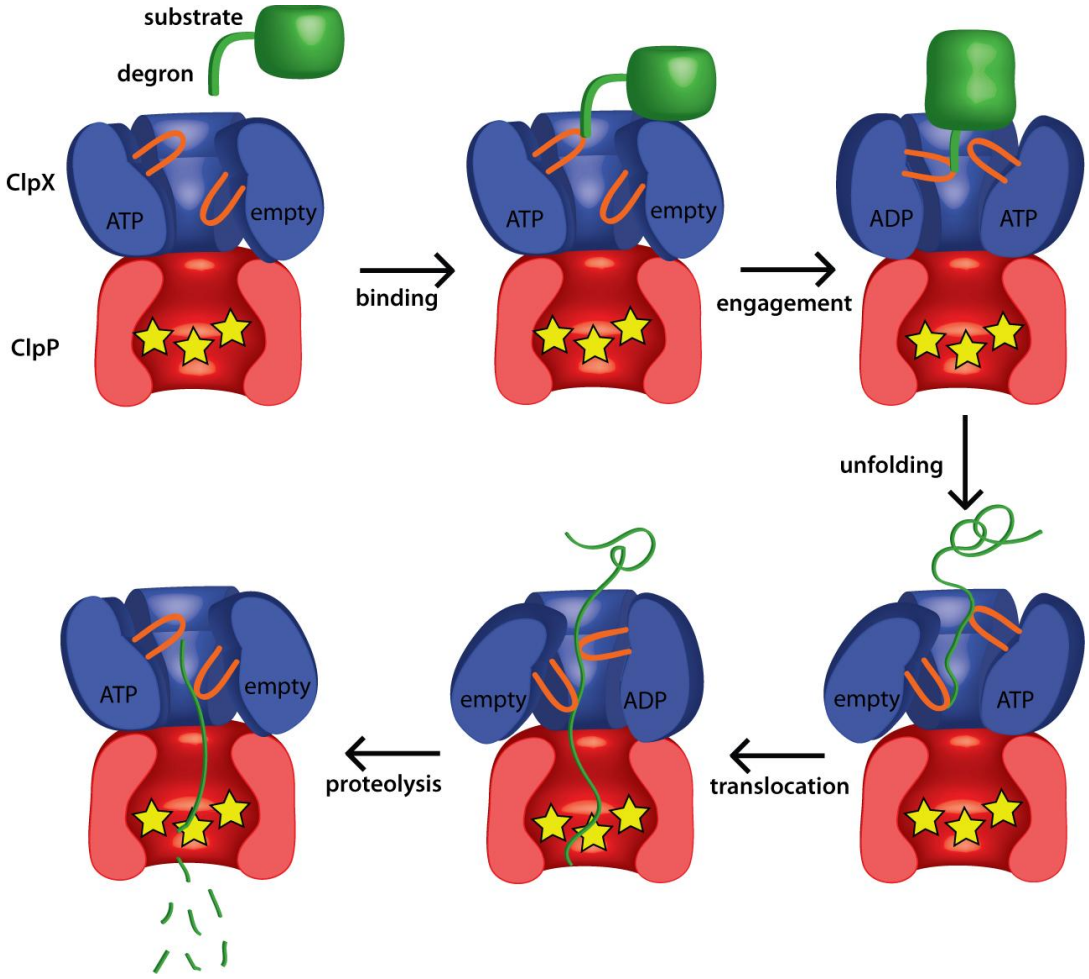

Figure 1.3. ClpXP-mediated protein degradation cycle. Pore loops (orange) extending from ClpX (blue) bind to the substrate's exposed degradation tag (green). Successive rounds of ATP hydrolysis drive conformational change in ClpX, which eventually result in substrate unfolding. The denatured polypeptide is translocated to the sequestered active sites (yellow) of ClpP (red) and cleaved. For ease of representation, only one ClpX hexamer per ClpP 14-mer is shown. Substrate denaturation and translocation can require hundreds of rounds of ATP hydrolysis by multiple subunits, a limited number is shown for simplicity.

Although the details of substrate recognition differ for each AAA+ protease (ClpXP, ClpAP, HslUV, Lon, FtsH, and the 26S proteasome), all are thought to unfold, translocate and degrade substrate using a mechanism similar to that described for ClpX (Kim *et al.*, 2000; Lee *et al.*, 2001a; Reid *et al.*, 2001; Liu *et al.*, 2006). Indeed, biochemical studies have directly demonstrated energy-dependent substrate unfolding and translocation by ClpAP, Lon and HslUV (Weber-Ban *et al.*, 1999; Kwon *et al.*, 2004; Burton *et al.*, 2005; Gur and Sauer, 2009). Further,

like ClpX, each member of this class has been shown to form ring oligomers composed of AAA+ domains which cap compartmentalized proteases (Kopp *et al.*, 1986; Lowe *et al.*, 1995; Bochtler *et al.*, 1997; Groll *et al.*, 1997; Wang *et al.*, 1997; Grimaud *et al.*, 1998; Sousa *et al.*, 2000; Glynn *et al.*, 2009). Lastly, ClpX loops known to be critical for substrate translocation are conserved throughout this class of enzymes and, in the cases of HslUV and ClpAP, have been shown to mediate substrate translocation (Siddiqui *et al.*, 2004; Hinnerwisch *et al.*, 2005; Park *et al.*, 2005; Martin *et al.*, 2008b).

Interestingly, ClpX translocates polypeptides with almost no significant sequence specificity, implying that initial substrate binding and engagement are the critical steps in regulating the potentially destructive activity of ClpXP (Barkow *et al.*, 2009). Consistent with this hypothesis, accessory specificity factors (also known as adaptor proteins) have been isolated that affect substrate selection by ClpXP (Levchenko *et al.*, 2000; Zhou *et al.*, 2001).

The simplest mechanism to explain adaptor function is for the adaptor to simultaneously bind to the substrate and the protease, thereby increasing the substrate's effective concentration (Baker and Sauer, 2006). Much biochemical evidence has accumulated that the *E. coli* adaptor, SspB, delivers ssrA-tagged substrates using this "tethering" mechanism (Levchenko *et al.*, 2000). In a series of detailed *in vitro* experiments, Wah *et al.* demonstrated that limited treatment of SspB with the endoprotease subtilisin resulted in production of a stably folded N-terminal domain (residues 1-117). The C-terminal tail (residues 118-165), by contrast, was protease sensitive and appeared to be highly flexible (Wah *et al.*, 2003). The isolated N-terminal domain of SspB was competent for adaptor dimerization, and bound to ssrA tags as well as full-length SspB. This

domain, however, was not capable of delivering ssrA-tagged substrates to ClpXP and actually inhibited their degradation. Direct binding studies combined with gel-filtration experiments showed that the C-terminal tail of SspB bound to ClpX and facilitated formation of a ternary complex consisting of SspB, the ssrA-tagged substrate, and ClpX. Critically, this work also demonstrated that the substrate binding domain and tail of SspB must be physically linked for adaptor function, as addition of the individual components failed to deliver substrates for degradation. These observations led to the tethering model shown in Figure 1.4A. Crystal structures have been solved of the ssrA tag bound to the substrate-binding domain of SspB and of the C-terminal SspB tail in complex with an isolated N-terminal domain of ClpX (Levchenko *et al.*, 2003; Song and Eck, 2003; Park *et al.*, 2007). Taken together, these studies conclusively demonstrate that tethering is necessary for substrate delivery by SspB.

As predicted by the concentration-dependent tethering mechanism described above, ssrA-tagged substrates are degraded in absence of SspB, but the K_M for degradation is substantially decreased by addition of the adaptor (Levchenko *et al.*, 2000). Moreover, mutations in the terminal three amino acids of ssrA tag (illustrated in Figure 1.4B) have been identified that do not affect adaptor binding but vastly decrease ClpX's affinity for the tag (Flynn *et al.*, 2001; McGinness *et al.*, 2006). Substrate delivery by SspB can compensate for this decreased protease affinity, thereby facilitating robust degradation (McGinness *et al.*, 2006; Griffith and Grossman, 2008). The controlled degradation systems described in Chapters 3 and 5 rely on these adaptor-dependent degradation tags (Davis *et al.*, 2009).

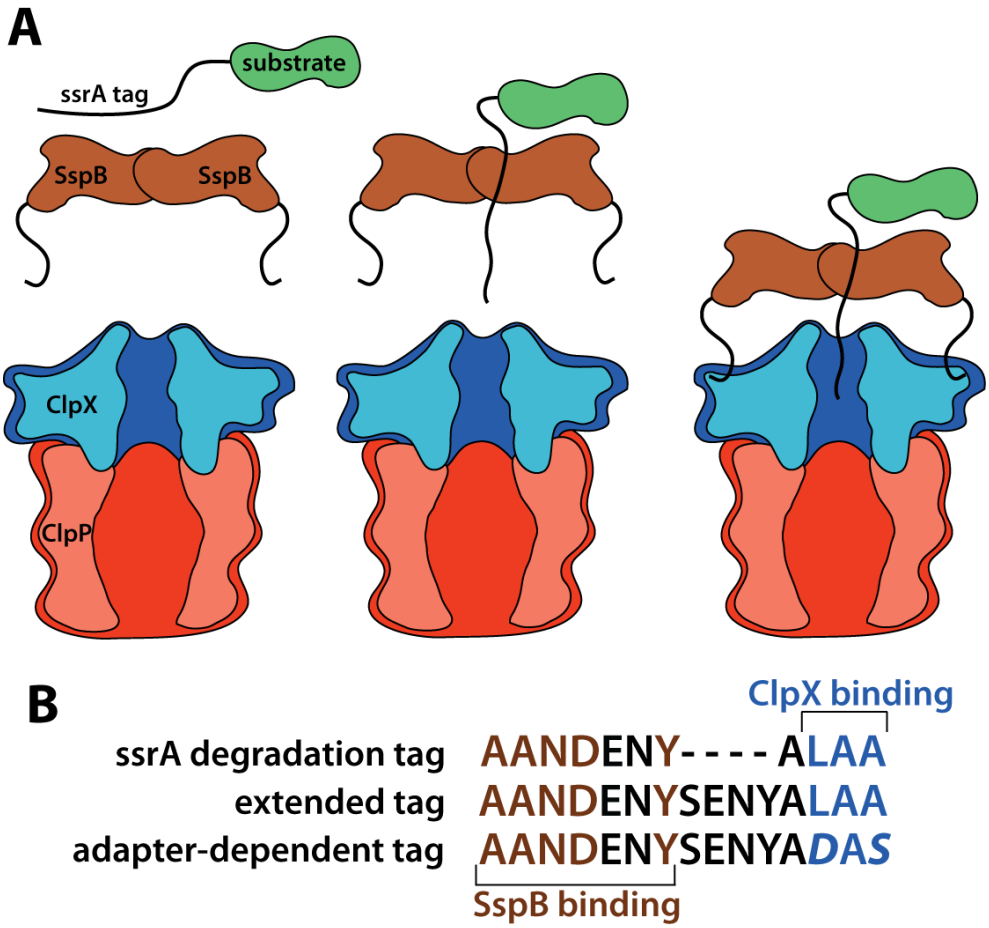

Figure 1.4. Adaptor-mediated delivery of an ssrA-tagged substrate. (A) The ssrA tag of the substrate binds to the N-terminal domain of SspB while the adaptor's flexible C-terminal tails simultaneously tether the substrate-adaptor complex to ClpX. Once tethered, substrate is engaged by ClpX, eventually leading to unfolding and proteolysis. **(B)** Biochemical studies have elucidated distinct, non-overlapping SspB and ClpX binding sites on the ssrA tag that allow protease and adaptor to simultaneously bind to the tag, albeit with slightly diminished affinity (Flynn *et al.*, 2001; Bolon *et al.*, 2004). Insertion of a short linker between the binding sites relieves this inhibition resulting in a more efficiently degraded "extended tag" (Hersch *et al.*, 2004; McGinness *et al.*, 2006). Mutations in the C-terminal residues of the ssrA tag have been identified which decrease affinity for ClpX without affecting affinity for SspB. Degradation of substrates bearing these "adapter-dependent tags" is strongly dependent on adaptor-mediate tethering to ClpX (McGinness *et al.*, 2006; Davis *et al.*, 2009).

In tethering models, the exact geometry of the delivery complex can greatly affect the efficiency of substrate delivery and thus it is not necessarily true that any molecule that tethers a substrate to a protease will act as an adaptor (McGinness *et al.*, 2007). Indeed, more complicated models for adaptor function have been proposed. For example, Thibault *et al.* proposed that SspB

function is mediated, in part, by its ability to direct large conformational changes in the N-terminal domain of ClpX (Thibault *et al.*, 2006). I have demonstrated, however, that the ClpX N-terminal domain is dispensable for SspB-mediated substrate delivery and can be replaced by synthetic tethering domains (Chapter 3). The fact that substrate delivery is as efficient with these synthetic constructs as with the natural adaptor argues against the aforementioned conformational-change model. Furthermore, by analyzing degradation using synthetic adaptors, proteases and substrates, the work in Chapter 3 demonstrates that in many instances tethering-alone is both necessary and sufficient for SspB-mediated substrate delivery. Indeed, substrates bearing multiple ClpX-binding motifs can "auto-tether", thereby mimicking SspB-mediated delivery and degradation.

For some adaptors, tethering alone is not sufficient for function. Recently, ClpS mutants have been identified that simultaneously bind N-end rule substrates and the ClpAP protease but fail to facilitate degradation (Hou *et al.*, 2008). This result indicates that more complicated interactions between this adaptor and protease are required to ensure substrate delivery. Still other adaptors are required for the oligomerization and assembly of active ClpC protease, demonstrating the wide range of biochemical activities performed by this class of regulatory proteins (Kirstein *et al.*, 2006).

Controlled protein degradation systems

Controllable degradation systems using AAA+ proteases and their substrate targeting machinery have been developed. In eukaryotes, it is known that aberrantly folded proteins are recognized and, through the action of a series of protein ligases, are modified by the covalently addition of

multiple ubiquitin molecules (Haas *et al.*, 1982; Hershko and Ciechanover, 1998). The fused polyubiquitin chain then acts like an adaptor, targeting the substrate to the proteasome where it is unfolded and eventually hydrolyzed to short peptides (Thrower *et al.*, 2000; Goldberg, 2003; Prakash *et al.*, 2004). A protein of interest can be experimentally targeted for degradation by fusing it to another protein that is conditionally misfolded. In one example of this technique, a misfolded mutant of the FKBP12 protein is fused to a target protein. Shield, an analog of the small-molecule rapamycin, binds to FKBP12 and stabilizes the folded protein, effectively blocking degradation. After removal of Shield, FKBP12 is destabilized, eventually leading to degradation of the entire fusion construct (Banaszynski *et al.*, 2006). Although not directly demonstrated by the authors, this degradation may occur as a result of substrate ubiquitination and subsequent targeting to the proteasome. Unfortunately, this technique is not particularly rapid, requiring many hours for dissociation and sufficient dilution of Shield.

In prokaryotes, proteins with an N-terminal Phe, Leu, Trp or Tyr residue are degraded by ClpAP (Tobias *et al.*, 1991; Erbse *et al.*, 2006; Wang *et al.*, 2007; Schmidt *et al.*, 2009). Target proteins can be expressed with an N-terminal domain bearing an endoprotease recognition site, which reveals one of these N-terminal residues following cleavage. As a proof of principle, the Sumo domain has been appended to target proteins in a way that allows induction and cleavage by the cognate Ulp1 protease, generating an exposed N-end rule residue that targets the protein for ClpAP degradation (Wang *et al.*, 2007).

Other controlled degradation systems in bacteria rely on the production of adaptors that aid in protease recognition of suitably tagged substrates. In these systems, the adaptor gene is knocked

out of the chromosome and resides on a vector plasmid under the control of an inducible promoter (McGinness *et al.,* 2006). Specific proteins are targeted for degradation through the C-terminal fusion of an adaptor-dependent degradation tag (Griffith and Grossman, 2008). To deplete the cell of the target proteins, one simply induces production of the adaptor.

Controlled degradation systems directly target pre-existing molecules, providing an improved temporal response relative to transcriptional and translational knockdowns. Because they require a genetically-encoded degradation tag these methods are also highly specific. Unlike chemical inhibitors or *ts*-mutations, targeted degradation should, in principle, be directly applicable to most intracellular proteins without extensive screening and optimization. In Chapter 5, I describe a controlled degradation system, using SspB and ClpXP, that employs the small-molecule rapamycin to control degradation. Because rapamycin diffuses into cells and shows no off-target effects in bacteria, this approach promises to be generally applicable in a variety of prokaryotic systems.

Research Approach

In subsequent chapters, I describe a series of experiments that improve our understanding of bacterial energy-dependent proteases and harness the power of these enzymes for controlled intracellular degradation. Chapter 2 details the collaborative development of a single-molecule assay to probe the kinetics of ClpXP-mediated substrate denaturation and degradation (Supplemental experiments can be found in Appendix A). This work, the first inspection of ClpXP activity at the single-molecule level, provides additional evidence that degradation of ssrA-tagged substrates is processive, proceeding from the C- to N-terminus. Further, our detailed

comparison of degradation kinetics, either of single molecules or in bulk solution, indicates similar overall levels of microscopic and macroscopic ClpXP activity, with no indication of non-uniform enzyme activity. During the development of this assay, we identified suitable methods to attach ClpXP to surfaces, to stabilize the hexameric form of the enzyme at extremely low concentrations, and to pre-engage substrates in a way that allows synchronous degradation. These technical advances promise to be valuable in future single-molecule studies of ClpXP. Building on this work, we generated a FRET-based assay to probe conformational changes in ClpXP with single-molecule resolution. The development and preliminary results of this assay are detailed in Appendix B. Results include a novel method to site-specifically attach fluorescent probes to individual ClpX subunits in the context of a hexamer. Using these fluorescently labeled molecules, we can directly observe nucleotide-dependent conformational changes either in bulk solution or at the single-molecule level.

As described above, the minimal biochemical activities sufficient for adaptor function had not been defined. In Chapter 3, I describe experiments that further test the "tethering" model. I generated synthetic adaptors, proteases, and substrates that interact through non-natural protein interfaces and performed degradation assays using these components. Consistent with the "tethering" model, these results show that specific contacts between the SspB adaptor and the ClpXP protease are dispensable for substrate delivery. Moreover, I demonstrate that tethering alone is both necessary and sufficient for synthetic adaptor function. Importantly, tethering in these constructs can be controlled using a small molecule, providing a means to regulate protein degradation with great specificity and temporal control. To facilitate the expression of these degradation components *in vivo*, I designed and characterized a library of σ70-dependent

constitutive bacterial promoters. A detailed analysis of these promoters is presented in Chapter 4. Appendix C describes the collaborative development of a method to measure relative promoter activity.

Chapter 5 details the use of a small-molecule-controlled degradation system *in vivo*. Specifically, I demonstrate rapamycin-dependent degradation of multiple substrates in *E. coli* using the endogenous ClpXP protease, and a synthetic adaptor. I extend this system by engineering a synthetic adaptor-tag pair, which can be utilized in *E. coli* in the presence of wild-type SspB. Use of this system simply requires appending a short degradation tag to a protein of interest, transforming cells with a plasmid driving constitutive production of the synthetic adaptor, and addition of rapamycin.

References

1. Abdelmohsen, K., Srikantan, S., Yang, X., Lal, A., Kim, H. H., Kuwano, Y., Galban, S., Becker, K. G., Kamara, D., de Cabo, R., and Gorospe, M. 2009. Ubiquitin-mediated proteolysis of HuR by heat shock. *Embo J* **28**(9): 1271-1282.

2. Atsumi, S., and Liao, J. C. 2008. Metabolic engineering for advanced biofuels production from Escherichia coli. *Curr Opin Biotechnol* **19**(5): 414-419.

3. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**2006 0008.

4. Baker, T. A., and Sauer, R. T. 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**(12): 647-653.

5. Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G., and Wandless, T. J. 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**(5): 995-1004.

6. Barkow, S. R., Levchenko, I., Baker, T. A., and Sauer, R. T. 2009. Polypeptide translocation by the AAA+ ClpXP protease machine. *Chem Biol* **16**(6): 605-612.

7. Bertram, R., and Hillen, W. 2008. The application of Tet repressor in prokaryotic gene regulation and expression. *Microbial Biotechnology* **1**(1): 2-16.

8. Bishop, A. C., Shah, K., Liu, Y., Witucki, L., Kung, C., and Shokat, K. M. 1998. Design of allele-specific inhibitors to probe protein kinase signaling. *Curr Biol* **8**(5): 257-266.

9. Bochtler, M., Ditzel, L., Groll, M., and Huber, R. 1997. Crystal structure of heat shock locus V (HslV) from Escherichia coli. *Proc Natl Acad Sci U S A* **94**(12): 6070-6074.

10. Bolhuis, A., Tjalsma, H., Stephenson, K., Harwood, C. R., Venema, G., Bron, S., and van Dijl, J. M. 1999. Different mechanisms for thermal inactivation of Bacillus subtilis signal peptidase mutants. *J Biol Chem* **274**(22): 15865-15868.

11. Bolon, D. N., Grant, R. A., Baker, T. A., and Sauer, R. T. 2004. Nucleotide-dependent substrate handoff from the SspB adaptor to the AAA+ ClpXP protease. *Mol Cell* **16**(3): 343-350.

12. Burton, R. E., Baker, T. A., and Sauer, R. T. 2005. Nucleotide-dependent substrate recognition by the AAA+ HslUV protease. *Nat Struct Mol Biol* **12**(3): 245-251.

13. Campbell, T. L., and Brown, E. D. 2008. Genetic interaction screens with ordered overexpression and deletion clone sets implicate the Escherichia coli GTPase YjeQ in late ribosome biogenesis. *J Bacteriol* **190**(7): 2537-2545.

14. Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12**(2): 323-337.

15. Chakshusmathi, G., Mondal, K., Lakshmi, G. S., Singh, G., Roy, A., Ch, R. B., Madhusudhanan, S., and Varadarajan, R. 2004. Design of temperature-sensitive mutants solely from amino acid sequence. *Proc Natl Acad Sci U S A* **101**(21): 7925-7930.

16. Chien, P., Perchuk, B. S., Laub, M. T., Sauer, R. T., and Baker, T. A. 2007. Direct and adaptor-mediated substrate recognition by an essential AAA+ protease. *Proc Natl Acad Sci U S A* **104**(16): 6590-6595.

17. Ciechanover, A. 2005. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* **6**(1): 79-87.

18. Cohen, P. 2002. The origins of protein phosphorylation. *Nat Cell Biol* **4**(5): E127-130.

19. Court, D. L., Sawitzke, J. A., and Thomason, L. C. 2002. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**361-388.

20. Couso, J. P., Bate, M., and Martinez-Arias, A. 1993. A wingless-dependent polar coordinate system in Drosophila imaginal discs. *Science* **259**(5094): 484-489.

21. Croxen, M. A., Ernst, P. B., and Hoffman, P. S. 2007. Antisense RNA modulation of alkyl hydroperoxide reductase levels in Helicobacter pylori correlates with organic peroxide toxicity but not infectivity. *J Bacteriol* **189**(9): 3359-3368.

22. Datta, S., Costantino, N., and Court, D. L. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**109-115.

23. Davis, J. H., Baker, T. A., and Sauer, R. T. 2009. Engineering synthetic adaptors and substrates for controlled ClpXP degradation. *J Biol Chem* **284**(33): 21848-21855.

24. DeRisi, J. L., Iyer, V. R., and Brown, P. O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**(5338): 680-686.

25. Dorsett, Y., and Tuschl, T. 2004. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* **3**(4): 318-329.

26. Dougan, D. A., Mogk, A., and Bukau, B. 2002. Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. *Cell Mol Life Sci* **59**(10): 1607-1616.

27. Edgar, R. S., and Lielausis, I. 1964. Temperature-Sensitive Mutants of Bacteriophage T4d: Their Isolation and Genetic Characterization. *Genetics* **49**649-662.

28. Eidlic, L., and Neidhardt, F. C. 1965. Protein and Nucleic Acid Synthesis in Two Mutants of Escherichia Coli with Temperature-Sensitive Aminoacyl Ribonucleic Acid Synthetases. *J Bacteriol* **89**706-711.

29. Elowitz, M. B., and Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767): 335-338.

30. Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A., and Bukau, B. 2006. ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**(7077): 753-756.

31. Farrell, C. M., Baker, T. A., and Sauer, R. T. 2007. Altered specificity of a AAA+ protease. *Mol Cell* **25**(1): 161-166.

32. Flynn, J. M., Levchenko, I., Seidel, M., Wickner, S. H., Sauer, R. T., and Baker, T. A. 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **98**(19): 10584-10589.

33. Furth, P. A., St Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., and Hennighausen, L. 1994. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proc Natl Acad Sci U S A* **91**(20): 9302-9306.

34. Gardner, T. S., Cantor, C. R., and Collins, J. J. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**(6767): 339-342.

35. Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balazsi, G., Ravasz, E., Daugherty, M. D., Somera, A. L., Kyrpides, N. C., Anderson, I., Gelfand, M. S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M. V., Grechkin, Y., Mseeh, F., Fonstein, M. Y., Overbeek, R., Barabasi, A. L., Oltvai, Z. N., and Osterman, A. L. 2003. Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **185**(19): 5673-5684.

36. Gillaspie, D., Perkins, I., Larsen, K., McCord, A., Pangonis, S., Sweger, D., Seleem, M. N., Sriranganathan, N., and Anderson, B. E. 2009. Plasmid-based system for high-level gene expression and antisense gene knockdown in Bartonella henselae. *Appl Environ Microbiol* **75**(16): 5434-5436.

37. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A., and Sauer, R. T. 2009. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**(4): 744-756.

38. Goldberg, A. L. 2003. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**(6968): 895-899.

39. Gossen, M., and Bujard, H. 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**(12): 5547-5551.

40. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. 1995. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**(5218): 1766-1769.

41. Gottesman, S. 1996. Proteases and their targets in Escherichia coli. *Annu Rev Genet* **30**465-506.

42. Griffith, K. L., and Grossman, A. D. 2008. Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Mol Microbiol* **70**(4): 1012-1025.

43. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C., and Maurizi, M. R. 1998. Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *J Biol Chem* **273**(20): 12476-12481.

44. Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D., and Huber, R. 1997. Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* **386**(6624): 463-471.

45. Gur, E., and Sauer, R. T. 2009. Degrons in protein substrates program the speed and operating efficiency of the AAA+ Lon proteolytic machine. *Proc Natl Acad Sci U S A* **106**(44): 18503-18508.

46. Haas, A. L., Warms, J. V., Hershko, A., and Rose, I. A. 1982. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *J Biol Chem* **257**(5): 2543-2548.

47. Hanson, P. I., and Whiteheart, S. W. 2005. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**(7): 519-529.

48. Hartwell, L. H. 1967. Macromolecule synthesis in temperature-sensitive mutants of yeast. *J Bacteriol* **93**(5): 1662-1670.

49. Hersch, G. L., Baker, T. A., and Sauer, R. T. 2004. SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc Natl Acad Sci U S A* **101**(33): 12136-12141.

50. Hershko, A., and Ciechanover, A. 1998. The ubiquitin system. *Annu Rev Biochem* **67**425-479.

51. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W., and Horwich, A. L. 2005. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121**(7): 1029-1041.

52. Hou, J. Y., Sauer, R. T., and Baker, T. A. 2008. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol* **15**(3): 288-294.

53. Jasin, M., and Schimmel, P. 1984. Deletion of an essential gene in Escherichia coli by site-specific recombination with linear DNA fragments. *J Bacteriol* **159**(2): 783-786.

54. Jenal, U., and Fuchs, T. 1998. An essential protease involved in bacterial cell-cycle control. *Embo J* **17**(19): 5658-5669.

55. Jensen, P. R., Westerhoff, H. V., and Michelsen, O. 1993. The use of lac-type promoters in control analysis. *Eur J Biochem* **211**(1-2): 181-191.

56. Ji, Y., Zhang, B., Van, S. F., Horn, Warren, P., Woodnutt, G., Burnham, M. K., and Rosenberg, M. 2001. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**(5538): 2266-2269.

57. Keiler, K. C., Waller, P. R., and Sauer, R. T. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**(5251): 990-993.

58. Kenniston, J. A., Baker, T. A., Fernandez, J. M., and Sauer, R. T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**(4): 511-520.

59. Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A., and Sauer, R. T. 2004. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* **146**(1-2): 130-140.

60. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. 2000. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**(4): 639-648.

61. Kirstein, J., Schlothauer, T., Dougan, D. A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., and Turgay, K. 2006. Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *Embo J* **25**(7): 1481-1491.

62. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. 2005. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res* **12**(5): 291-299.

63. Knight, Z. A., and Shokat, K. M. 2007. Chemical genetics: where genetics and pharmacology meet. *Cell* **128**(3): 425-430.

64. Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T. S., Cantor, C. R., and Collins, J. J. 2004. Programmable cells: interfacing natural and engineered gene networks. *Proc Natl Acad Sci U S A* **101**(22): 8414-8419.

65. Kopp, F., Steiner, R., Dahlmann, B., Kuehn, L., and Reinauer, H. 1986. Size and shape of the multicatalytic proteinase from rat skeletal muscle. *Biochim Biophys Acta* **872**(3): 253-260.

66. Kuwano, M., Ono, M., Endo, H., Hori, K., Nakamura, K., Hirota, Y., and Ohnishi, Y. 1977. Gene affecting longevity of messenger RNA: a mutant of Escherichia coli with altered mRNA stability. *Mol Gen Genet* **154**(3): 279-285.

67. Kwon, A. R., Trame, C. B., and McKay, D. B. 2004. Kinetics of protein substrate degradation by HslUV. *J Struct Biol* **146**(1-2): 141-147.

68. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M., and Matouschek, A. 2001a. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol Cell* **7**(3): 627-637.

69. Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A., and Copeland, N. G. 2001b. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**(1): 56-65.

70. Levchenko, I., Seidel, M., Sauer, R. T., and Baker, T. A. 2000. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**(5488): 2354-2356.

71. Levchenko, I., Grant, R. A., Wah, D. A., Sauer, R. T., and Baker, T. A. 2003. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol Cell* **12**(2): 365-372.

72. Liu, C. W., Li, X., Thompson, D., Wooding, K., Chang, T. L., Tang, Z., Yu, H., Thomas, P. J., and DeMartino, G. N. 2006. ATP binding and ATP hydrolysis play distinct roles in the function of 26S proteasome. *Mol Cell* **24**(1): 39-50.

73. Lloyd, A. 2003. Vector construction for gene overexpression as a tool to elucidate gene function. *Methods Mol Biol* **236**329-344.

74. Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. 1995. Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science* **268**(5210): 533-539.

75. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**(1): 117-124.

76. Mann, M., and Jensen, O. N. 2003. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**(3): 255-261.

77. Martin, A., Baker, T. A., and Sauer, R. T. 2008a. Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol Cell* **29**(4): 441-450.

78. Martin, A., Baker, T. A., and Sauer, R. T. 2008b. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat Struct Mol Biol* **15**(11): 1147-1151.

79. Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotechnol* **21**(7): 796-802.

80. McGinness, K. E., Baker, T. A., and Sauer, R. T. 2006. Engineering controllable protein degradation. *Mol Cell* **22**(5): 701-707.

81. McGinness, K. E., Bolon, D. N., Kaganovich, M., Baker, T. A., and Sauer, R. T. 2007. Altered tethering of the SspB adaptor to the ClpXP protease causes changes in substrate delivery. *J Biol Chem* **282**(15): 11465-11473.

82. McManus, M. T., and Sharp, P. A. 2002. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet* **3**(10): 737-747.

83. Mieschendahl, M., and Muller-Hill, B. 1985. F'-coded, temperature-sensitive lambda cI857 repressor gene for easy construction and regulation of lambda promoter-dependent expression systems. *J Bacteriol* **164**(3): 1366-1369.

84. Moglich, A., Ayers, R. A., and Moffat, K. 2009. Design and signaling mechanism of light-regulated histidine kinases. *J Mol Biol* **385**(5): 1433-1444.

85. Murata, S., Yashiroda, H., and Tanaka, K. 2009. Molecular mechanisms of proteasome assembly. *Nat Rev Mol Cell Biol* **10**(2): 104-115.

86. Murray, A. W., Solomon, M. J., and Kirschner, M. W. 1989. The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. *Nature* **339**(6222): 280-286.

87. Neher, S. B., Flynn, J. M., Sauer, R. T., and Baker, T. A. 2003. Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes Dev* **17**(9): 1084-1089.

88. Neher, S. B., Villen, J., Oakes, E. C., Bakalarski, C. E., Sauer, R. T., Gygi, S. P., and Baker, T. A. 2006. Proteomic profiling of ClpXP substrates after DNA damage reveals extensive instability within SOS regulon. *Mol Cell* **22**(2): 193-204.

89. Okuno, T., Yamanaka, K., and Ogura, T. 2006. An AAA protease FtsH can initiate proteolysis from internal sites of a model substrate, apo-flavodoxin. *Genes Cells* **11**(3): 261-268.

90. Park, E., Rho, Y. M., Koh, O. J., Ahn, S. W., Seong, I. S., Song, J. J., Bang, O., Seol, J. H., Wang, J., Eom, S. H., and Chung, C. H. 2005. Role of the GYVG pore motif of HslU ATPase in protein unfolding and translocation for degradation by HslV peptidase. *J Biol Chem* **280**(24): 22892-22898.

91. Park, E. Y., Lee, B. G., Hong, S. B., Kim, H. W., Jeon, H., and Song, H. K. 2007. Structural basis of SspB-tail recognition by the zinc binding domain of ClpX. *J Mol Biol* **367**(2): 514-526.

92. Parsell, D. A., and Sauer, R. T. 1989. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J Biol Chem* **264**(13): 7590-7595.

93. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E., and Matouschek, A. 2004. An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **11**(9): 830-837.

94. Pribnow, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A* **72**(3): 784-788.

95. Pyne, C., and Bognar, A. L. 1992. Replacement of the folC gene, encoding folylpolyglutamate synthetase-dihydrofolate synthetase in Escherichia coli, with genes mutagenized in vitro. *J Bacteriol* **174**(6): 1750-1759.

96. Reid, B. G., Fenton, W. A., Horwich, A. L., and Weber-Ban, E. U. 2001. ClpA mediates directional translocation of substrate proteins into the ClpP protease. *Proc Natl Acad Sci U S A* **98**(7): 3768-3772.

97. Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**(7086): 940-943.

98. Sadler, J. R., and Novick, A. 1965. The Properties of Repressor and the Kinetics of Its Action. *J Mol Biol* **12**305-327.

99. Sakuma, R., Noser, J. A., Ohmine, S., and Ikeda, Y. 2007. Rhesus monkey TRIM5alpha restricts HIV-1 production through rapid degradation of viral Gag polyproteins. *Nat Med* **13**(5): 631-635.

100. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

101. Schmidt, R., Zahn, R., Bukau, B., and Mogk, A. 2009. ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. *Mol Microbiol* **72**(2): 506-517.

102. Sharp, P. A. 2001. RNA interference--2001. *Genes Dev* **15**(5): 485-490.

103. Shine, J., and Dalgarno, L. 1975. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**(5495): 34-38.

104. Siddiqui, S. M., Sauer, R. T., and Baker, T. A. 2004. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev* **18**(4): 369-374.

105. Silo-Suh, L. A., Elmore, B., Ohman, D. E., and Suh, S. J. 2009. Isolation, characterization, and utilization of a temperature-sensitive allele of a Pseudomonas replicon. *J Microbiol Methods* **78**(3): 319-324.

106. Smith, D. H., Berget, P. B., and King, J. 1980. Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail-spike protein. I. Fine-structure mapping. *Genetics* **96**(2): 331-352.

107. Song, H. K., and Eck, M. J. 2003. Structural basis of degradation signal recognition by SspB, a specificity-enhancing factor for the ClpXP proteolytic machine. *Mol Cell* **12**(1): 75-86.

108. Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S. G., Cyert, M., Hughes, T. R., Boone, C., and Andrews, B. 2006. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21**(3): 319-330.

109. Sousa, M. C., Trame, C. B., Tsuruta, H., Wilbanks, S. M., Reddy, V. S., and McKay, D. B. 2000. Crystal and solution structures of an HslUV protease-chaperone complex. *Cell* **103**(4): 633-643.

110. Strickland, D., Moffat, K., and Sosnick, T. R. 2008. Light-activated DNA binding in a designed allosteric protein. *Proc Natl Acad Sci U S A* **105**(31): 10709-10714.

111. Sussman, R., and Jacob, F. 1962. [On a thermosensitive repression system in the Escherichia coli lambda bacteriophage.]. *C R Hebd Seances Acad Sci* **254**1517-1519.

112. Thibault, G., Tsitrin, Y., Davidson, T., Gribun, A., and Houry, W. A. 2006. Large nucleotide-dependent movement of the N-terminal domain of the ClpX chaperone. *Embo J* **25**(14): 3367-3376.

113. Thrower, J. S., Hoffman, L., Rechsteiner, M., and Pickart, C. M. 2000. Recognition of the polyubiquitin proteolytic signal. *Embo J* **19**(1): 94-102.

114. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. 1991. The N-end rule in bacteria. *Science* **254**(5036): 1374-1377.

115. Varadarajan, R., Nagarajaram, H. A., and Ramakrishnan, C. 1996. A procedure for the prediction of temperature-sensitive mutants of a globular protein based solely on the amino acid sequence. *Proc Natl Acad Sci U S A* **93**(24): 13908-13913.

116. Wah, D. A., Levchenko, I., Rieckhof, G. E., Bolon, D. N., Baker, T. A., and Sauer, R. T. 2003. Flexible linkers leash the substrate binding domain of SspB to a peptide module that stabilizes delivery complexes with the AAA+ ClpXP protease. *Mol Cell* **12**(2): 355-363.

117. Wang, J., Hartling, J. A., and Flanagan, J. M. 1997. The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**(4): 447-456.

118. Wang, K. H., Sauer, R. T., and Baker, T. A. 2007. ClpS modulates but is not essential for bacterial N-end rule degradation. *Genes Dev* **21**(4): 403-408.

119. Waters, L. S., and Storz, G. 2009. Regulatory RNAs in bacteria. *Cell* **136**(4): 615-628.

120. Weber-Ban, E. U., Reid, B. G., Miranker, A. D., and Horwich, A. L. 1999. Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**(6748): 90-93.

121. Wilson, C. J., Zhan, H., Swint-Kruse, L., and Matthews, K. S. 2007. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64**(1): 3-16.

122. Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**(11): 5978-5983.

123. Zhou, Y., Gottesman, S., Hoskins, J. R., Maurizi, M. R., and Wickner, S. 2001. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. *Genes Dev* **15**(5): 627-637.

Chapter 2

Single-molecule Denaturation and Degradation of Proteins by the AAA+ ClpXP Protease

This work was published as Yongdae Shin[†], Joseph H. Davis[†], Ricardo R. Brau[†], Andreas Martin[†], Jon A. Kenniston, Tania A. Baker, Robert T. Sauer, and Matthew J. Lang. 2009 *PNAS* 106:19340-19345.

R.R.B and J.A.K initiated this work by constructing the TIRF microscope and investigating the on-surface stability of full-length ClpX respectively. A.M. aided in experimental design as well as the construction of substrates and single-chain ClpX. Y.S. collected and analyzed the majority of the single-molecule datasets and composed the initial manuscript.

[†]These authors contributed equally to this work

Abstract

ClpXP is an ATP-fueled molecular machine that unfolds and degrades target proteins. ClpX, an AAA+ enzyme, recognizes specific proteins, and then uses cycles of ATP hydrolysis to denature any native structure and to translocate the unfolded polypeptide into ClpP for degradation. Here, we develop and apply single-molecule fluorescence assays to probe the kinetics of protein denaturation and degradation by ClpXP. These assays employ a single-chain variant of the ClpX hexamer, linked via a single biotin to a streptavidin-coated surface, and fusion substrates with an N-terminal fluorophore and a C-terminal GFP-titin-ssrA module. In the presence of adenosine 5′-[γ-thio]triphosphate (ATPγS), ClpXP degrades the titin-ssrA portion of these substrates but stalls when it encounters GFP. Exchange into ATP then allows synchronous resumption of denaturation and degradation of GFP and any downstream domains. GFP unfolding can be monitored directly, because intrinsic fluorescence is quenched by denaturation. The time required for complete degradation coincides with loss of the substrate fluorophore from the protease complex. Fitting single-molecule data for a set of related substrates provides time constants for ClpX unfolding, translocation, and a terminal step that may involve product release. Comparison of these single-molecule results with kinetics measured in bulk solution indicates similar levels of microscopic and macroscopic ClpXP activity. These results support a stochastic engagement/unfolding mechanism that ultimately results in highly processive degradation and set the stage for more detailed single-molecule studies of machine function.

Abbreviations: GFP, green fluorescent protein; CFP, cyan fluorescent protein; PEG, polyethylene glycol; TIRF, total internal reflection fluorescence.

Introduction

Molecular machines of the AAA+ (ATPases associated with diverse cellular activities) enzyme superfamily play crucial roles in cellular processes ranging from protein degradation and DNA replication to membrane fusion and the movement of motor proteins along microtubule tracks (Ogura and Wilkinson, 2001; Hanson and Whiteheart, 2005). AAA+ proteases degrade proteins that are damaged, remove proteins that are no longer needed by the cell, and function in regulatory circuits that require proteolysis of specific target proteins. In bacteria, such as *Escherichia coli*, intracellular proteolysis is carried out by multiple ATP-dependent proteases, including ClpXP, ClpAP, HslUV, Lon, and FtsH (Gottesman, 2003). In the ClpXP machine, for example, the AAA+ ClpX component engages protein substrates, unfolds them, and ultimately translocates the denatured polypeptide into an internal chamber of the associated ClpP peptidase for irreversible proteolysis (Sauer *et al.*, 2004). Thus, the overall process of ClpXP degradation involves the operation and coordination of enzymatic machinery for substrate recognition, denaturation, translocation, and degradation.

ClpX functions as an asymmetric ring of six subunits, with the sites for ATP binding and hydrolysis located at subunit interfaces (Kim and Kim, 2003; Glynn *et al.*, 2009). ClpP is also active as a multimer, in which two stacked heptameric rings enclose a degradation chamber containing 14 active sites for peptide-bond cleavage (Wang *et al.*, 1997). A hexameric ClpX ring and a heptameric ClpP ring stack coaxially, creating a central channel that allows translocation of unfolded substrates through the ClpX pore and into the ClpP peptidase chamber (Ortega *et al.*, 2000). ClpX recognizes specific substrates by binding to exposed peptide sequences. For example, appending the ssrA tag (AANDENYALAA) to the C terminus of a protein makes it a

substrate for ClpXP degradation (Gottesman *et al.*, 1998). The ssrA tag initially binds in the axial pore of ClpX (Siddiqui *et al.*, 2004; Martin *et al.*, 2008b). Changes in ClpX conformation, powered by ATP binding and hydrolysis, are then postulated to initiate tag translocation through the pore. Because native proteins are larger than the ClpX pore, continued translocation eventually pulls on the attached protein and results in an unfolding force. For very stable protein domains, hundreds of cycles of ATP hydrolysis can be required on average before unfolding is successful, although single mutations that destabilize the substrate can reduce this value almost 50-fold (Kenniston *et al.*, 2003). Moreover, a hyperstable substrate can dissociate from the enzyme after an unsuccessful denaturation attempt (Kenniston *et al.*, 2005). Thus, any structural perturbations caused by the transient strain of attempted unfolding would almost certainly relax before that substrate was rebound by another enzyme. These facts suggest that successful unfolding results from a combination of the applied pulling force and stochastic changes in protein stability. For example, fluctuations in the distribution of thermal energy in the protein could result in occasional fraying of secondary structure or in partial unfolding that then allows a single ClpX pulling event to cooperatively denature the entire protein domain (Kenniston *et al.*, 2003). It is also possible that all ClpXP enzymes in bulk solution hydrolyze ATP but only a small fraction is active in denaturation. By this model, rare encounters between "active" ClpXP and substrate could lead to efficient denaturation in a single turnover, whereas the vast majority of substrate-enzyme interactions would not. In this case, average properties calculated assuming that all ClpXP enzymes in bulk solution are equally active could be very different from the actual properties of individual enzymes.

Since the first detection of single fluorophores at cryogenic temperatures (Moerner and Kador, 1989), single-molecule fluorescence has become a powerful technique for exploring the nanoscale behavior of individual molecules. For example, such studies have revealed important information regarding mechanoenzyme motility, unassisted and chaperone-mediated protein folding, and enzyme dynamics (Lu *et al.*, 1998; Deniz *et al.*, 2000; Yildiz *et al.*, 2003; Ueno *et al.*, 2004). Here, we develop and apply a single-molecule fluorescence assay to probe the kinetics of ClpXP-mediated substrate denaturation and degradation. Our results provide important support for mechanistic conclusions based on ensemble experiments and set the stage for more detailed single-molecule studies of ClpXP function.

Results

Experimental design

We had to overcome several problems during assay development. For example, wild-type *E. coli* ClpX bound nonspecifically and nonfunctionally to surfaces, apparently because hexamers dissociated and the isolated subunits were prone to denaturation. In addition, some methods of surface attachment precluded ClpX binding to ClpP. Eventually, we used a single-chain ClpX pseudohexamer lacking the N-domain (Martin *et al.*, 2005), which is not needed to degrade ssrA-tagged substrates (Singh *et al.*, 2001) but seemed to contribute to surface inactivation. This single-chain variant ($ClpX^{SC}$) was cloned with a sequence that allowed enzyme-mediated covalent attachment of one biotin molecule to each pseudohexamer. The biotinylated $ClpX^{SC}$ enzyme was as active as wild-type ClpX in ClpP-mediated degradation of a GFP-ssrA substrate (Figure A.S1, Appendix A).

Three substrates were used for single-molecule studies (Figure 2.1A). Each had an N-terminal cysteine, which we modified with a Cy3 fluorophore, and a common GFP-$titin^{V15P}$-ssrA unit, consisting of a GFP domain, the I27 domain of titin bearing the destabilizing V15P mutation, and a C-terminal ssrA tag (Kenniston *et al.*, 2003; Martin *et al.*, 2008a). In the presence of Mg^{2+} and adenosine 5′-[γ-thio]triphosphate (ATPγS), an ATP analog which ClpX hydrolyzes slowly, ClpXP degrades the $titin^{V15P}$-ssrA portion of these substrates but stalls when it reaches the GFP domain Figure 2.1B; (Martin *et al.*, 2008a). These stalled complexes are stable after chelation of Mg^{2+} by excess EDTA, which prevents nucleoside-triphosphate hydrolysis, and after replacing ATPγS/EDTA with ATP/EDTA (Figure A.S2, Appendix A). Subsequent addition of ATP/Mg^{2+} then initiates ATP hydrolysis and degradation of the GFP domain and downstream portions of

the substrate (Martin *et al.*, 2008a). Substrates were incubated with biotinylated $ClpX^{SC}$, ClpP, and ATPγS/Mg^{2+} to form stalled complexes, which were then introduced into a flow cell, immobilized on a glass slide coated with a mixture of covalently attached PEG and PEG-biotin-streptavidin (Joo *et al.*, 2006), and quantified by objective-side TIRF imaging (Figure 2.1C). We included an oxygen scavenging system (Yildiz *et al.*, 2003) and minimized the intensity and/or duration of laser excitation to reduce photobleaching. The Cy3 dye served as a marker for stalled complexes and remained bound to the surface-attached enzyme until completion of degradation.

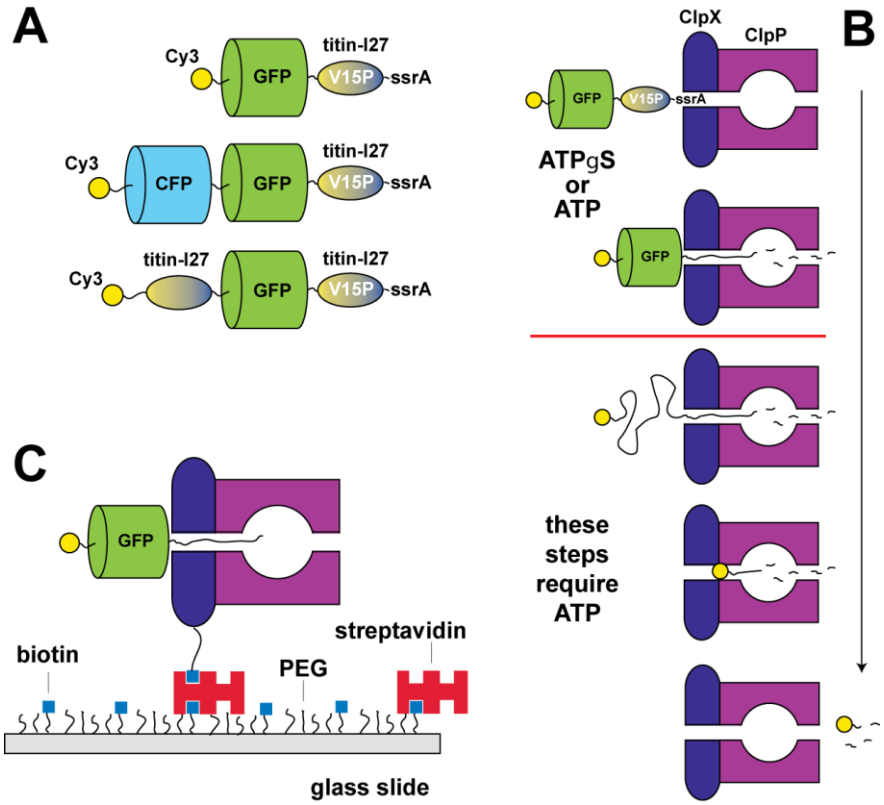

Figure 2.1. Substrates and methods used for single-molecule assays of ClpXP degradation. (A) Substrates contained an N-terminal Cy3 fluorophore, a GFP domain, a titin-I27 domain with the V15P mutation, and a C-terminal ssrA tag. **(B)** In the presence of ATPγS, ClpXP degrades the $titin^{V15P}$-ssrA portion of substrates but stalls because it cannot denature GFP (Martin *et al.*, 2008a). Exchange into ATP then permits GFP denaturation and completion of degradation. **(C)** For single-molecule experiments, pre-engaged complexes of substrates, biotinylated $ClpX^{SC}$, and ClpP were formed in the presence of ATPγS, tethered to a PEG-coated glass surface via PEG-biotin-streptavidin, and visualized by TIRF microscopy. Degradation reactions were initiated by exchange of ATP for ATPγS.

Initial controls were performed by using the Cy3-GFP-titin^{V15P}-ssrA substrate. Under standard conditions with biotinylated ClpX^{SC}, ClpP, and ATPγS/Mg^{2+}, we observed 99.8 ± 7.2 fluorescent spots per field of view (Figure 2.2A). Several experiments established that most spots corresponded to complexes of the substrate and the immobilized protease. (*i*) When biotinylated ClpX^{SC} was omitted (Figure 2.2B) or nonbiotinylated ClpX^{SC} was used, 9 ± 2 spots were observed. (*ii*) When ATPγS (Figure 2.2C) or Mg^{2+} was omitted from the preincubation, 13 ± 3 spots were observed. (*iii*) When Cy3-labeled substrate was omitted, no more than two spots were observed. (i*v*) In the continual presence of ATPγS/Mg^{2+}, Cy3-labeled substrate photobleached in an exponential process with a time constant of 330 s (Figure A.S3A, Appendix A), similar to the photobleaching times of Cy3-labeled molecules bound to control surfaces. (*v*) When buffer with Mg^{2+} but no ATPγS nucleotide was flowed over the surface, most spots disappeared with a time constant of ≈14 s (Figure A.S3B, Appendix A), consistent with studies showing that nucleotide is required for ClpX to maintain an active grip on the substrate (Martin *et al.*, 2008b).

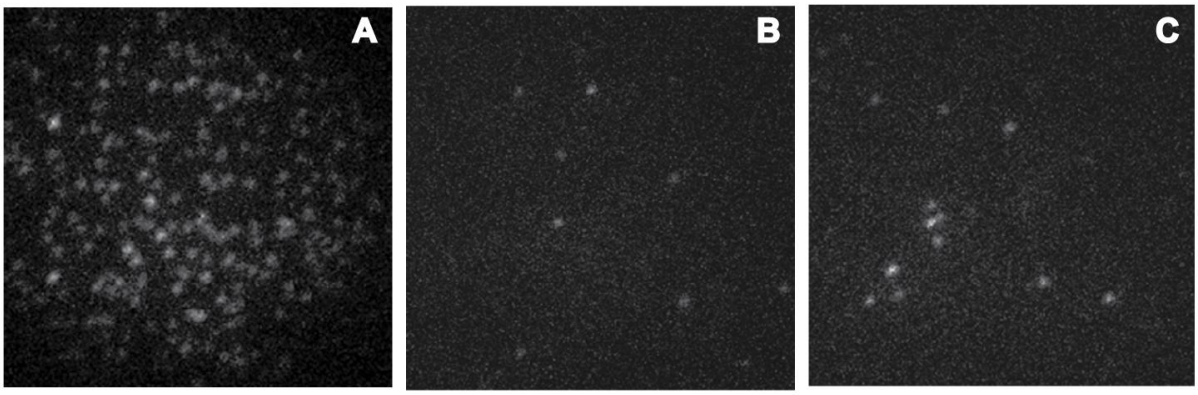

Figure 2.2. Single-molecule TIRF images of the Cy3-GFP-titin^{V15P}-ssrA substrate bound to a glass surface coated with PEG and PEG-biotin-streptavidin. (A) Approximately 100 fluorescent spots were detected after preincubation of the substrate (1 μM) with biotinylated ClpX^{SC} (0.31 μM), ClpP (1 μM), ATPγS (2 mM), and Mg^{2+} (10 mM), and the mixture was flowed over the surface. Far fewer spots were detected when biotinylated ClpX^{SC} **(B)** or ATPγS **(C)** was omitted from the preincubation mix.

Single-molecule degradation/denaturation

After exchanging ATP for ATPγS in the absence of Mg^{2+}, we excited the Cy3 dye, focused, and selected a field of view containing immobilized complexes of $ClpX^{SC}$-ClpP and the preengaged Cy3-GFP-$titin^{V15P}$-ssrA substrate. ATP/Mg^{2+} was then injected into the flow cell to initiate unfolding and degradation. Images were acquired at intervals (typically 3-5 s), and a custom software algorithm was used to quantify the lifetime of each spot. These data were combined for a large number of spots (1,122 spots) observed in 11 independent experiments and are presented as a probability density distribution of spot lifetimes (Figure 2.3A) or as fractional spot populations as a function of time (Figure 2.3B). As discussed below, these data can be corrected for photobleaching and related to the time required by the immobilized protease to unfold, translocate, and degrade substrates.

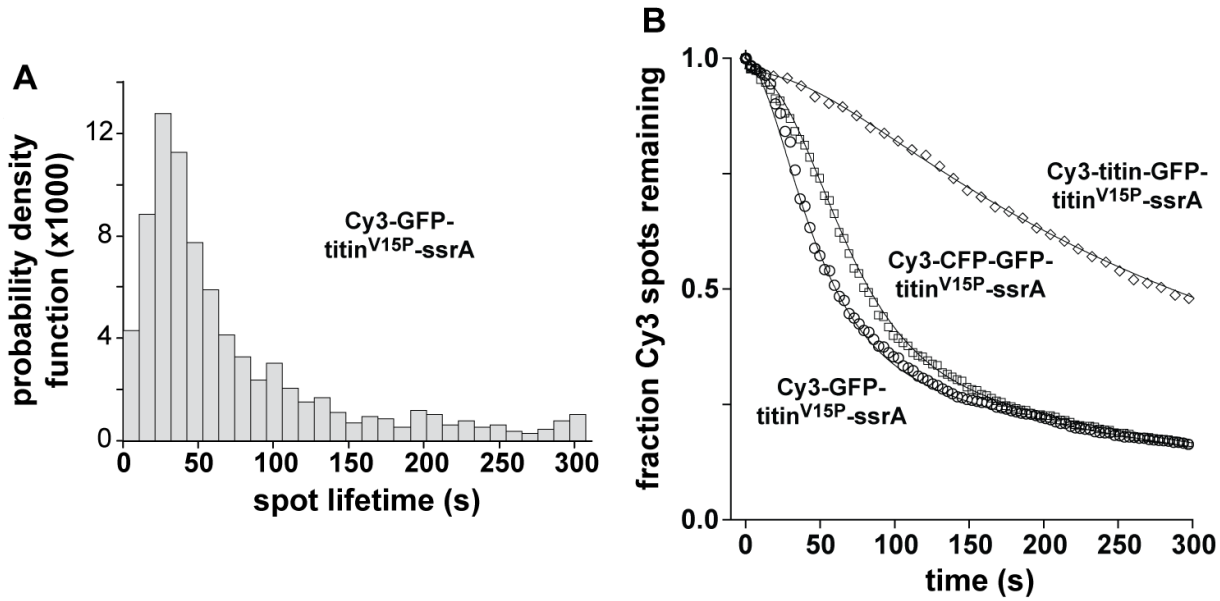

Figure 2.3. Single-molecule degradation. (A) Distribution of spot lifetimes for the Cy3-GFP-$titin^{V15P}$-ssrA substrate. **(B)** Kinetic profiles of single-molecule ClpXP degradation were constructed by summing the number of Cy3 spots in TIRF images taken at 3- to 9- s intervals after initiating degradation at ≈18 °C by addition of ATP/Mg^{2+} and normalizing to the initial value. The total number of initial spots was 1,122 for the Cy3-GFP-$titin^{V15P}$-ssrA substrate, 974 for the Cy3-CFP-GFP-$titin^{V15P}$-ssrA substrate, and 419 for the Cy3-titin-GFP-$titin^{V15P}$-ssrA substrate. The solid lines are fits to reaction models described in *Kinetic Modeling*.

After addition of ATP/Mg^{2+} (1/10 mM), the population of the preengaged Cy3-GFP-titin^{V15P}-ssrA substrate decreased to half the initial value in ≈60 s (Figure 2.3B). Several results indicated that the majority of this population decrease represents active ClpXP degradation. First, spot loss was faster with ATP/Mg^{2+} than with ATPγS/Mg^{2+} (50% loss in 231 s). In the latter experiment, degradation does not occur, and spots are only lost via photobleaching. Second, as expected from solution experiments (Martin *et al.*, 2008a), spot loss slowed when a lower concentration of ATP (0.1 mM) was used (50% loss in 90 s; Figure A.S4A, Appendix A) or when a mixture of ATP/ATPγS (1/0.25 mM) was used (50% loss in 155 s; Figure A.S4B, Appendix A).

We used substrates with additional domains to confirm that the loss of Cy3 fluorescence correlated with the time required for complete ClpXP degradation. For example, single-molecule ClpXP degradation of both Cy3-CFP-GFP-titin^{V15P}-ssrA (50% loss in 80 s) and Cy3-titin-GFP-titin^{V15P}-ssrA (50% loss in 279 s)[1] proceeded more slowly than degradation of Cy3-GFP-titin^{V15P}-ssrA (Figure 2.3B). The relative rates of single-molecule ClpXP degradation of all three substrates were consistent with solution studies (Kenniston *et al.*, 2003; Martin *et al.*, 2008a), in which GFP and CFP domains are degraded at similar rates, but titin is degraded ≈4-fold more slowly because of its exceptional mechanical stability. These results are also consistent with studies that show that degradation of ssrA-tagged substrates proceeds processively from the C to the N terminus (Lee *et al.*, 2001; Kenniston *et al.*, 2005), and thus the Cy3 dye is lost only when degradation is complete.

[1] For this substrate, a 9 s interval between images was used and the photobleaching time constant was 1330 s.

The processive model predicts that GFP denaturation should be an early event in overall degradation (Figure 2.1B). To test this hypothesis, we monitored loss of fluorescent GFP spots after initiating ClpXP denaturation of preengaged GFP-titin^{V15P}-ssrA. As expected, GFP spots were lost at a faster rate than Cy3 spots (Figure 2.4A), indicating that denaturation of the GFP domain occurs earlier than release of the Cy3 fluorophore during ClpXP degradation. Fitting of the GFP data, including a correction for photobleaching, gave a time constant of 19 s for denaturation of this domain.

Kinetic modeling

We globally fit the degradation data for Cy3-GFP-titin^{V15P}-ssrA (called S1) and Cy3-CFP-GFP-titin^{V15P}-ssrA (called S2) to kinetic models with individual steps for unfolding (τ_{unf}) and translocation (τ_{trans}) of each domain (see *Materials and Methods*). Solution studies indicate that ClpXP unfolds and translocates GFP and CFP at similar rates (Martin *et al.*, 2008a), and thus we used the same time constants for both domains. A model including just these kinetic steps predicts that S2 should take twice as long to degrade as S1 and fit the data poorly. Indeed, S2 was only degraded ≈1.5-fold more slowly than S1 in our experiments, suggesting that an extra slow step contributes to both reactions. Fitting to the following reactions, which include a time constant (τ_{term}) for an unspecified terminal step, resulted in good simultaneous fits to both data sets (Figure 2.3B).

$$S1 \xrightarrow{\tau_{unf}} I1 \xrightarrow{\tau_{trans}} I2 \xrightarrow{\tau_{term}} \text{degraded}$$

$$S2 \xrightarrow{\tau_{unf}} I1 \xrightarrow{\tau_{trans}} I2 \xrightarrow{\tau_{unf}} I3 \xrightarrow{\tau_{trans}} I4 \xrightarrow{\tau_{term}} \text{degraded}$$

When unfolding was constrained to the experimental value (τ_{unf} = 19 s), the fitted time constants were τ_{trans} = 6.2 s and τ_{term} = 29.6 s, yielding time constants for overall degradation of 54.8 s for S1 and 80 s for S2 (Table A.S1, Appendix A).

We also fit the data for Cy3-titin-GFP-titin^{V15P}-ssrA degradation by using the τ_{unf} and τ_{trans} values determined for GFP, assuming that τ_{trans} for the titin domain was half the GFP value because it has half as many amino acids, allowing τ_{unf} for titin to vary, and using the τ_{term} value obtained above. This procedure gave a good fit of the experimental data (Figure 2.3B) with a time constant of 347 s for titin unfolding (Table A.S1, Appendix A), indicating that ClpXP unfolding of the titin domain proceeds almost 20-fold more slowly than unfolding of the GFP or CFP domains.

Substrate unfolding and translocation in solution

The single-molecule experiments were performed at ≈18 °C, a temperature where ClpXP activity is not typically assayed. To allow comparisons, we performed a set of bulk experiments in solution at 18 °C. Because some experiments required initial ClpXP binding and degradation of the titin^{V15P}-ssrA portion of substrates, we unfolded the titin^{V15P} domain by carboxymethylation (titin^{CM}) to preclude the need for unfolding (Kenniston *et al.*, 2003). First, we determined K_M (1.9 μM) and V_{max} (0.62 min^{-1}) for degradation of GFP-titin^{CM}-ssrA by biotinylated ClpX^{SC} and ClpP. The corresponding time constant for steady-state degradation (97 s) was longer than for single-molecule degradation but includes the times required for binding, engagement, and translocation of titin^{CM} as well as for denaturation and translocation of GFP.

To study the GFP-unfolding reaction alone, we purified preengaged complexes of GFP-titin^{V15P}-ssrA with ClpX^{SC}-ClpP and initiated degradation by stopped-flow addition of ATP/Mg^{++} (Martin *et al.*, 2008a). In this experiment, the loss of native GFP fluorescence, which accompanied ClpX-mediated unfolding, showed a minor burst phase (amplitude 19%; τ = 0.87 s) and a major unfolding phase (amplitude 81%; τ = 25 s) (Figure 2.4B). Importantly, the kinetics of GFP unfolding in this solution experiment were similar to the kinetics estimated from the single-molecule data (τ = 19 s). Given the lack of precise temperature control and low time resolution in the single-molecule experiments, the differences between these values are probably not significant.

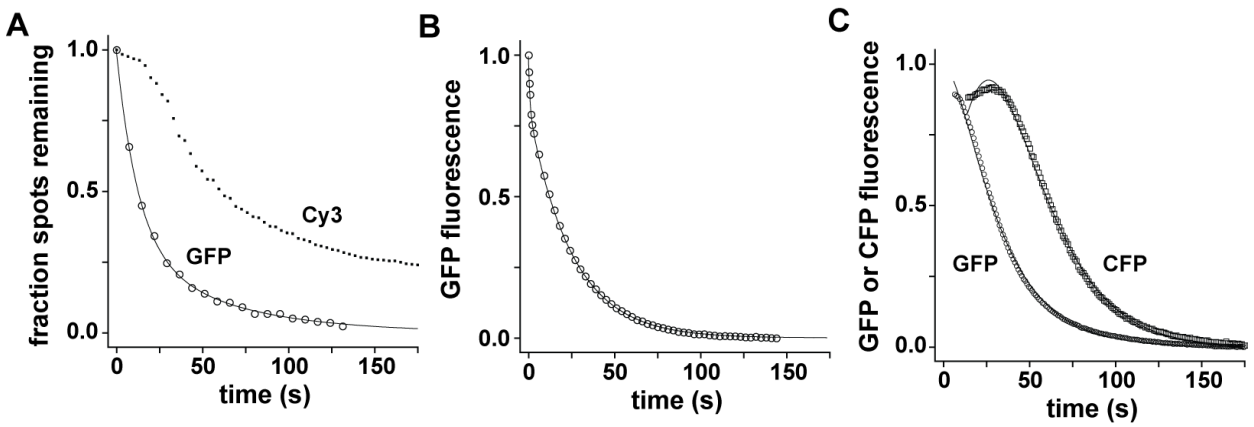

Figure 2.4. Substrate denaturation assayed by single-molecule or solution experiments. (A) Single-molecule ClpXP denaturation of the pre-engaged Cy3-GFP-titin^{V15P}-ssrA substrate. Circles represent a kinetic profile constructed by summing the number of GFP spots in TIRF images taken at 7 s intervals after addition of ATP/Mg^{2+} and normalizing to the initial value (251 spots). The solid-line is a double exponential fit ($y = 0.9 \cdot \exp(-t/17.3) + 0.1 \cdot \exp(-t/197)$). After correcting the time constant for the dominant phase to account for a photobleaching contribution, the time constant for GFP denaturation was 19 s. The squares show the Cy3 degradation data for the same substrate. **(B)** Solution denaturation. Pre-engaged complexes of Cy3-GFP-titin^{V15P}-ssrA, ClpX^{SC}, and ClpP were generated in the presence of ATPγS/Mg^{2+}, purified, and GFP denaturation at 18 °C was initiated by addition of ATP/Mg^{2+} in a stopped flow instrument and monitored by changes in GFP fluorescence (Martin *et al.*, 2008a). The solid-line is a double exponential fit of the data ($y = 0.19 \cdot \exp(-t/0.87) + 0.81 \cdot \exp(-t/25)$). The fast phase may represent diminished fluorescence caused by ClpX extraction of the C-terminal β-strand of GFP (see ref. 21). **(C)** Solution denaturation of the CFP-GFP-titin^{CM}-ssrA substrate. At time 0, the substrate (0.5 μM) was mixed with ClpX^{SC} (1 μM), ClpP (2 μM), the SspB adaptor (0.75 μM), and an ATP regeneration system at 18 °C. Changes in GFP fluorescence or CFP fluorescence were monitored in separate experiments. The solid lines are fits to the model described in *Substrate Unfolding and Translocation in Solution*.

Finally, we monitored loss of GFP or CFP fluorescence under single-turnover conditions with ClpXP in 2-fold excess over CFP-GFP-titin^{CM}-ssrA (Figure 2.4C). In these experiments, ClpXP must bind, engage, and translocate the unfolded titin^{CM}-ssrA portion of the substrate and then denature/translocate the GFP domain followed by the CFP domain. Fitting the GFP data gave a time constant of 10 s for binding/engagement/translocation of titin^{CM}-ssrA and a time constant of 25 s for GFP denaturation, the same value determined above using the preengaged substrate (Figure 2.4C). The CFP data fit well to a model with time constants of 10 s for titin^{CM}-ssrA

binding/engagement/translocation, 20 s for GFP unfolding, 6 s for translocation of unfolded GFP, and 20 s for CFP unfolding (Figure 2.4C). The difference in the GFP unfolding value for the two fits is probably caused by poor fitting of the initial rise in fluorescence that occurs in the CFP trace because of the loss of FRET that occurs upon GFP denaturation (Figure 2.4C). Nevertheless, the kinetic constants obtained from these solution experiments were very similar to the values obtained by fitting the single-molecule data.

Discussion

In our experiments, individual complexes of surface-attached ClpXP enzymes with preengaged substrates containing a Cy3 fluorophore are detected by TIRF microscopy. Under appropriate experimental conditions, these complexes remain fluorescent until the Cy3 dye is removed by the final steps of the degradation reaction or is photobleached. The photobleaching rate can be independently determined, allowing pooled population data to be fitted to determine apparent rates of single-molecule degradation. Multiple experiments indicate that these rates reflect degradation of the complete substrate. First, single-molecule degradation times increased as additional domains were added to the substrate and as these extra domains became more difficult to denature. Second, ClpXP degraded the same substrate more slowly when lower concentrations of ATP or mixtures of ATP/ATPγS were used, as expected from solution studies (Martin *et al.*, 2008a). Third, monitoring GFP instead of Cy3 fluorescence during degradation of GFP-titin^{V15P}-ssrA resulted in faster rates. This result is expected because GFP fluorescence is immediately quenched upon ClpX-mediated denaturation, whereas loss of Cy3 requires additional kinetic steps, including translocation of the unfolded GFP polypeptide, proteolysis by ClpP, and product release. These combined results also provide strong support for a model in which ClpXP degradation of ssrA-tagged substrates proceeds processively from the C terminus to the N terminus (Lee *et al.*, 2001; Kenniston *et al.*, 2005; Martin *et al.*, 2008a).

In our single-molecule experiments, most ClpXP enzymes that engaged the substrate in the presence of ATPγS were then able to denature GFP and to complete the degradation reaction when ATP was added. For example, when we monitored single-molecule GFP denaturation, 90% of the fluorescence loss occurred by the denaturation pathway with 10% following a

photobleaching-only pathway. This ratio is almost exactly that predicted from control experiments, in which ≈10% of the Cy3-GFP-titin^{V15P}-ssrA substrate appeared to be bound nonspecifically to the surface. Moreover, single-molecule experiments and single-turnover solution experiments gave similar time constants for GFP denaturation and, in both cases, the data were fit reasonably well by a single exponential process. These results would not be expected if the individual enzymes in the preparations used for these studies displayed a broad range of denaturation activities, although there could be a population of completely inactive enzymes. Lower estimates of the percentage of active ClpXP enzymes (58-75%) were obtained by fitting the single-molecule degradation data. Because the overall time of laser irradiation was longer in the degradation experiments, some enzyme inactivation may occur during the experiment, potentially as a consequence of oxidative damage.

Intriguingly, good fits to our single-molecule degradation data required a step in addition to substrate denaturation and translocation. In principle, this step could correspond to slow ClpP cleavage of some polypeptide sequences, to slow product release of the Cy3-labeled peptide from ClpP, or if the final step of translocation is slow because the substrate is no longer engaged efficiently by the translocation machinery. Solution studies have shown that steady-state rates of ClpXP degradation at substrate saturation (V_{max}) are slower than those predicted from single-turnover measurements of substrate denaturation and translocation (Martin *et al.*, 2008a). Thus, the additional kinetic step suggested by our single-molecule experiments may also slow the steady-state rate of ClpXP substrate degradation in solution. It is also possible that inactive enzymes in our ClpXP preparations reduce the average activity and lead to an underestimate of V_{max}, which is calculated with the assumption that all enzymes are active. We note, however, that

ClpXP degraded carboxymethylated GFP-titin^{V15P}-ssrA at a steady-state maximal rate that corresponds to a time constant of 97 s for solution degradation. The time constant for single-molecule ClpXP degradation of the GFP portion of this substrate was 55 s. The ratio of these values suggests that at least 57% of the ClpXP enzymes in solution are active. However, this value is likely to be higher because the solution reaction includes additional steps of substrate binding and translocation of the unfolded titin portion of the substrate compared to the single-molecule reaction. Importantly, these results are inconsistent with models in which the solution activity of ClpXP is mediated by a small fraction of active enzymes.

Understanding the operating principles and detailed mechanisms of complex macromolecular machines, like ClpXP, will ultimately require a combination of structural, biochemical, and biophysical approaches. Here, we have developed methods that allow a single-chain variant of ClpX to be tethered to a surface, to bind ClpP, and to carry out denaturation and degradation of specific substrate proteins. Importantly, the summed single-molecule activities of our surface-tethered ClpXP enzymes recapitulate those of a population of free enzymes in bulk single-turnover experiments. It should be straightforward to extend these methods to allow single-molecule measurements of the forces exerted during protein denaturation and/or translocation by ClpXP, to measure detailed rates and step sizes for polypeptide translocation by ClpXP, to perform multiple-color experiments, and to introduce FRET probes that will allow real-time assays of the repetitive ATP-fueled conformational changes that drive the mechanical operations of the ClpXP machine.

Materials and Methods

Protein expression and purification

Detailed methods for the expression and purification of enzymes and substrates are provided in Appendix A (A.S1 text). Prior to labeling substrates with Cy3 maleimide (GE Healthcare), DTT was removed by exchange into 25 mM Hepes (pH 7.2), 50 mM KCl, 1 mM EDTA, 10% glycerol. An aliquot of Cy3 maleimide (2 μL of a freshly prepared 3 mM solution in dimethyl formamide) was then added to 100 μL of the substrate solution (typically 1-5 μM) and the mixture was incubated overnight at room temperature. Unreacted Cy3 dye was removed by chromatography by using three sequential desalting columns. Controls showed that Cy3 labeling was reduced to 6% when substrates lacked a cysteine at the N terminus (Figure A.S5, Appendix A).

Flow cells

Flow cells (30 x 5 mm) had a volume of approximately 15 μL and were made from double-sided sticky tape gaskets sandwiched between a predrilled glass slide and an etched-glass coverslip coated with a mixture of 99% PEG (molecular weight 5,000) and 1% biotin-PEG (Laysan Bio) to minimize nonspecific binding. The holes in the drilled slide were attached to flexible tubing to allow rapid buffer exchange, while minimizing motion of the sample. The flow chamber and tubing were sealed with epoxy. Preengaged substrate-enzyme complexes were formed by incubating substrates (1 μM) with single-chain $ClpX^{SC}$ (0. 31 μM), ClpP-H_6 (1 μM), and ATPγS (2 mM) for 45 min in PD buffer [25 mM Hepes (pH 7.6) , 100 mM KCl, 10 mM $MgCl_2$, 10% glycerol (vol/vol), 0.1% Tween (vol/vol)] at 30 °C (Singh *et al.*, 2001). After treating the flow cell with 20 μl of 0.01 mg/mL streptavidin, preengaged substrate-ClpXP complexes were diluted ≈30-fold, introduced into the flow cell, and incubated for 20 min at room temperature to allow

binding of biotinylated $ClpX^{SC}$ to the streptavidin-biotin-PEG surface. In all subsequent washes, ClpP-H_6 (800 nM) was included to ensure maintenance of the $ClpX^{SC}$-ClpP complex. After the binding step, the flow cell was washed with 2 mM ATPγS and 50 mM EDTA to chelate Mg^{2+} and prevent further hydrolysis. The chamber was then washed with PD buffer (without Mg^{2+}) plus 6 mM EDTA and 1 mM ATP. After identifying a suitable region of the surface for analysis, the shutters were closed and the stage was moved slightly to a nearby field of view. The shutters were reopened to acquire the first image and 100 µL of PD buffer plus ATP was flowed into the cell to initiate the reaction (final ATP/Mg^{2+} concentration normally 1/10 mM). An oxygen scavenging system consisting of 0.8% D(+)-glucose, 165 units/mL glucose oxidase, 2,170 units/mL catalase and 0.1% 2-mercaptoethanol was added in this final buffer to minimize photobleaching (Yildiz *et al.*, 2003).

Single-molecule fluorescence

Single-molecule assays were performed by using a heavily modified inverted microscope outfitted with objective-side total internal reflection fluorescence capabilities (Brau *et al.*, 2006; Tarsa *et al.*, 2007). To minimize photobleaching, the excitation laser (532 nm for Cy3; 488 nm for GFP) was modulated with an acousto-optic modulator and also rapidly toggled with an electronic shutter to extend fluorophore longevity and synchronize image acquisition. The custom TIRF system included a 1.45 N.A. 100X objective and a dichroic mirror. Images were acquired by using an EMCCD camera, which was triggered externally to collect during the 300-ms on-time of the excitation laser, with an average power of 50 µW at the specimen plane. A series of 100 images was taken at 3- to 9-s intervals for degradation experiments; 21 images were taken at 7 s intervals for the denaturation experiment. The excitation zone was typically 300 μm^2

and an average of 99.8 ± 7.2 spots were observed. Images were analyzed to determine the longevity of each fluorescent spot by using custom MATLAB software. Figure A.S6 (Appendix A) shows typical kinetic traces for Cy3 spots, in which fluorescence was lost in a single step. Blinking was observed for some GFP spots (see Figure A.S7, Appendix A), but the software counted only events in which fluorescence was permanently lost as degradation or photobleaching.

Solution assays

Solution unfolding or single-turnover degradation of substrates by $ClpX^{SC}$-ClpP was measured at 18 °C by using methods previously described (Martin *et al.*, 2008a). Assays were performed in PD-1 buffer (25 mM Hepes (pH 7.6), 5 mM $MgCl_2$, 10% glycerol, and 200 mM KCl) by using a creatine-phosphate based ATP-regeneration system. GFP fluorescence (excitation 467 nm; emission 511) or CFP fluorescence (excitation 433 nm; emission 475 nm) were used to monitor unfolding. The GFP-$titin^{V15P}$-ssrA substrate was preengaged using ATPγS, and the enzyme-substrate complex was exchanged into a buffer with ATP but no Mg^{2+} as described (Martin *et al.*, 2008a). GFP unfolding was initiated by mixing one volume of the preengaged enzyme-substrate complex with an equal volume of 2X ATP buffer (8 mM ATP, 32 mM creatine phosphate, and 0.64 mg/ml creatine kinase in PD buffer) and the reaction was monitored by fluorescence by using an Applied Photophysics DX.17MV stopped-flow fluorimeter equipped with a 495-nm cutoff emission filter.

Kinetic fitting

Fitting of experimental data to kinetic models was performed by an iterative nonlinear least-squares algorithm implemented in IGOR PRO 4.07 (WaveMetrics). The probabilities for each kinetic state were derived by solving rate equations with the Laplace transform. In these models (see Appendix A text for more detail), each state either moved on to next state (e.g., from domain unfolding to translocation) or photobleached. The model also included a substrate subpopulation that only lost fluorescence by the photobleaching pathway. The sum of all probabilities except for the final nonfluorescent states corresponds to the expected fractional spot population at any given time. The experimental fractional spot populations for degradation of the S1 and S2 substrates were globally fitted to these models, after fixing the unfolding time constant of GFP (19 s) and the photobleaching time constant of Cy3 (330 s) based on independent measurements.

Acknowledgements

We thank M. Aubin-Tam, C. Castro, J. Damon, A. Khalil, H. Lee, S. Moore, E. Oakes, and A. Olivares for materials, advice, and helpful discussions. T.A.B. is an employee of the Howard Hughes Medical Institute (HHMI). This work was supported by National Science Foundation Career Award 0643745, a Samsung Scholarship from the Samsung Foundation of Culture, National Institutes of Health Grant AI-15706, and HHMI.

References

1. Brau, R. R., Tarsa, P. B., Ferrer, J. M., Lee, P., and Lang, M. J. 2006. Interlaced optical force-fluorescence measurements for single molecule biophysics. *Biophys J* **91**(3): 1069-1077.

2. Deniz, A. A., Laurence, T. A., Beligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S., Dawson, P. E., Schultz, P. G., and Weiss, S. 2000. Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci U S A* **97**(10): 5179-5184.

3. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A., and Sauer, R. T. 2009. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**(4): 744-756.

4. Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**(9): 1338-1347.

5. Gottesman, S. 2003. Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **19**565-587.

6. Hanson, P. I., and Whiteheart, S. W. 2005. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**(7): 519-529.

7. Joo, C., McKinney, S. A., Nakamura, M., Rasnik, I., Myong, S., and Ha, T. 2006. Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**(3): 515-527.

8. Kenniston, J. A., Baker, T. A., Fernandez, J. M., and Sauer, R. T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**(4): 511-520.

9. Kenniston, J. A., Baker, T. A., and Sauer, R. T. 2005. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proc Natl Acad Sci U S A* **102**(5): 1390-1395.

10. Kim, D. Y., and Kim, K. K. 2003. Crystal structure of ClpX molecular chaperone from Helicobacter pylori. *J Biol Chem* **278**(50): 50664-50670.

11. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M., and Matouschek, A. 2001. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol Cell* **7**(3): 627-637.

12. Lu, H. P., Xun, L., and Xie, X. S. 1998. Single-molecule enzymatic dynamics. *Science* **282**(5395): 1877-1882.

13. Martin, A., Baker, T. A., and Sauer, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**(7062): 1115-1120.

14. Martin, A., Baker, T. A., and Sauer, R. T. 2008a. Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat Struct Mol Biol* **15**(2): 139-145.

15. Martin, A., Baker, T. A., and Sauer, R. T. 2008b. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat Struct Mol Biol* **15**(11): 1147-1151.

16. Moerner, W. E., and Kador, L. 1989. Optical detection and spectroscopy of single molecules in a solid. *Phys Rev Lett* **62**(21): 2535-2538.

17. Ogura, T., and Wilkinson, A. J. 2001. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**(7): 575-597.

18. Ortega, J., Singh, S. K., Ishikawa, T., Maurizi, M. R., and Steven, A. C. 2000. Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol Cell* **6**(6): 1515-1521.

19. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

20. Siddiqui, S. M., Sauer, R. T., and Baker, T. A. 2004. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev* **18**(4): 369-374.

21. Singh, S. K., Rozycki, J., Ortega, J., Ishikawa, T., Lo, J., Steven, A. C., and Maurizi, M. R. 2001. Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *J Biol Chem* **276**(31): 29420-29429.

22. Tarsa, P. B., Brau, R. R., Barch, M., Ferrer, J. M., Freyzon, Y., Matsudaira, P., and Lang, M. J. 2007. Detecting force-induced molecular transitions with fluorescence resonant energy transfer. *Angew Chem Int Ed Engl* **46**(12): 1999-2001.

23. Ueno, T., Taguchi, H., Tadakuma, H., Yoshida, M., and Funatsu, T. 2004. GroEL mediates protein folding with a two successive timer mechanism. *Mol Cell* **14**(4): 423-434.

24. Wang, J., Hartling, J. A., and Flanagan, J. M. 1997. The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**(4): 447-456.

25. Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R. 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**(5628): 2061-2065.

Chapter 3

Engineering Synthetic Adaptors and Substrates for Controlled ClpXP Degradation

This work was published as Joseph H. Davis, Tania A. Baker, and Robert T. Sauer. 2009. *Journal of Biological Chemistry* **284**(33):21848-21855.

Abstract

Facile control of targeted intracellular protein degradation has many potential uses in basic science and biotechnology. One promising approach to this goal is to redesign adaptor proteins, which can regulate proteolytic specificity by tethering substrates to energy-dependent AAA+ proteases. Using the ClpXP protease, we have probed the minimal biochemical functions required for adaptor function by designing and characterizing variant substrates, adaptors, and ClpX enzymes. We find that substrate tethering mediated by heterologous interaction domains and a small bridging molecule mimics substrate delivery by the wild-type system. These results show that simple tethering is sufficient for synthetic adaptor function. In our engineered system, tethering and proteolysis depend on the presence of the macrolide rapamycin, providing a foundation for engineering highly specific degradation of target proteins in cells. Importantly, this degradation is regulated by a small molecule without the need for new adaptor or enzyme biosynthesis.

Abbreviations: Ni^{2+}-NTA, Ni^{2+}-nitrilotriacetic acid; CV, column volume; GFP, green fluorescent protein; DHFR, dihydrofolate reductase

Introduction

Targeted proteolytic degradation plays important roles in protein-quality control and in regulating cellular circuitry in organisms ranging from bacteria to humans (Gottesman, 2003; Sauer *et al.*, 2004; Baker and Sauer, 2006; Bukau *et al.*, 2006). In some instances, substrates are recognized directly by a protease enzyme via a degradation tag (Figure 3.1-top) (Gottesman *et al.*, 1998; Neher *et al.*, 2003). In other cases, adaptor proteins or multiple types of substrate sequences are also required to ensure efficient degradation (Figure 3.1-middle, 3.1-bottom) (Flynn *et al.*, 2004; McGinness *et al.*, 2006).

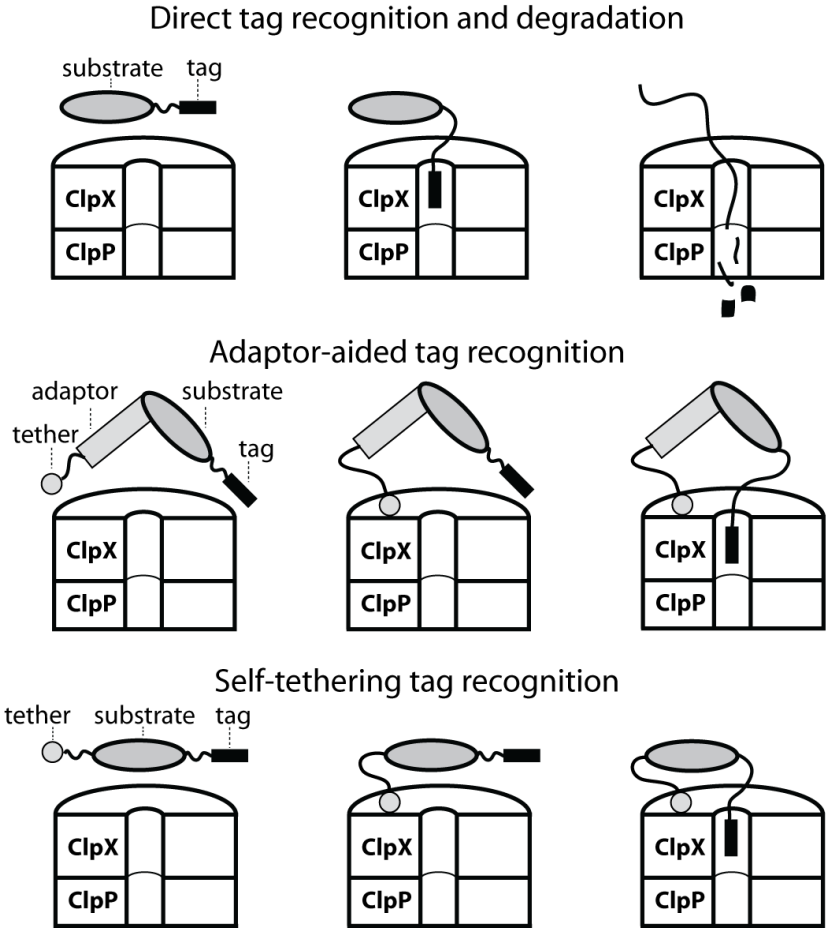

Figure 3.1. The ClpX component of the ClpXP protease recognizes some substrates via a degradation tag, denatures the substrate, and then translocates the unfolded protein into ClpP for degradation (top). Adaptor-assisted binding of a substrate to ClpXP (middle). Self-tethering of a substrate to ClpXP (bottom).

Experimentally induced degradation can be used as a tool to probe the role of specific proteins in cellular processes. For example, a protein that is normally stable can be modified to make its degradation conditionally dependent on the presence of an adaptor, allowing studies of the consequences of depletion after induction of adaptor synthesis (McGinness *et al.*, 2006; Griffith and Grossman, 2008). Such systems complement methods, such as RNAi, that rely upon repressing biosynthesis of the target protein but offer significant advantages when rapid depletion of otherwise long-lived proteins is the goal (Fire *et al.*, 1998; Janse *et al.*, 2004; Banaszynski *et al.*, 2006). We are interested in engineering synthetic adaptor systems to control targeted intracellular degradation.

ClpXP is a AAA+ protease present in bacteria and mitochondria that consists of two components, ClpX and ClpP. Hexamers of ClpX recognize degradation tags in specific substrate proteins, unfold them in a reaction that requires ATP hydrolysis, and then use additional cycles of ATP hydrolysis to translocate the unfolded polypeptide into an interior chamber of ClpP, where proteolysis takes place (Figure 3.1-top). The simplest way in which an adaptor could stimulate degradation is by tethering a specific substrate to a protease, thereby increasing its effective concentration and facilitating proteolysis (Figure 3.1-middle); (Baker and Sauer, 2006). The SspB adaptor, for example, appears to function by this mechanism. SspB enhances ClpXP degradation of certain substrates, including N-RseA and proteins bearing the ssrA-degradation tag (Levchenko *et al.*, 2000; Flynn *et al.*, 2004; Sauer *et al.*, 2004). ClpXP degrades these substrates in the absence of SspB, but K_M for degradation is substantially lower when this adaptor is present. Two features of SspB are consistent with a tethering mechanism. It has a substrate-binding domain with a groove that binds a portion of the ssrA tag or a sequence in N-

RseA, and it contains a flexible C-terminal extension terminating with a peptide motif (XB) that binds to the N-terminal domain of ClpX (Dougan *et al.*, 2003; Levchenko *et al.*, 2003; Song and Eck, 2003; Wah *et al.*, 2003; Levchenko *et al.*, 2005; Park *et al.*, 2007). Mutations that prevent SspB binding to ClpX or block substrate binding to SspB eliminate stimulation of degradation (Levchenko *et al.*, 2000; Bolon *et al.*, 2004; Park *et al.*, 2007).

It has not been rigorously established, however, that tethering *per se* is sufficient for the activity of any adaptor. Based on biochemical experiments, for instance, Thibault *et al*. proposed that the adaptor activity of SspB is mediated, in part, by its ability to direct the movement of the N-terminal domains of ClpX, and thereby to regulate the delivery of tagged substrates to ClpXP (Thibault *et al.*, 2006). For some adaptors, tethering of the substrate to the protease is not sufficient for degradation. For example, the ClpS adaptor tethers N-end-rule substrates to the AAA+ ClpAP protease (Dougan *et al.*, 2002; Erbse *et al.*, 2006; Wang *et al.*, 2007), but some ClpS mutants mediate efficient substrate tethering to ClpAP without facilitating degradation (Hou *et al.*, 2008). In such cases, more complicated transactions between the adaptor and the protease appear to be needed to ensure that the substrate is properly delivered to the protease. Moreover, in some instances, adaptors play roles in substrate delivery but are also required for assembly of the active protease (Kirstein *et al.*, 2006).

The studies reported here were motivated by two major goals. First, we wished to test if a completely synthetic adaptor system could be used to regulate substrate degradation. Second, we sought to design a proteolysis system that could be controlled by the presence or absence of a small molecule. To define the minimal biochemical properties required for adaptor-protein

function, we engineered and characterized synthetic variants of adaptors, substrates, and the ClpXP protease. We reasoned that if specialized interactions between SspB and the N-terminal domain of ClpX were a requisite part of substrate delivery, then replacing either component would preclude efficient degradation. By contrast, we found that rapid degradation of an otherwise poor substrate was possible in the absence of SspB and the N-domain as long as substrate-enzyme tethering was maintained by other interaction domains. These results show that tethering alone is sufficient for synthetic-adaptor function. We were also able to control degradation *in vitro* and *in vivo* using systems in which a small molecule, rapamycin, drives assembly of tethered proteolytic complexes. Thus, targeted degradation can be engineered to depend, in a conditional fashion, on the presence of a small molecule. In principle, degradation under small-molecule control has many of the advantages of chemical genetics (Walsh and Chang, 2006), but should be even simpler and more widely applicable as a method of functional inhibition. In addition, controlling degradation in this fashion is possible even when biosynthesis of new macromolecules is precluded.

Experimental Procedures

Buffers

LB1 buffer (pH 7.6) contained 20 mM Hepes, 400 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, and 10 mM 2-mercaptoethanol. LB2 buffer (pH 8.0) contained 100 mM NaH_2PO_4, 10 mM Tris-HCl, 6 M GuHCl, and 10 mM imidazole. EB1 buffer (pH 7.6) contained 20 mM Hepes, 400 mM NaCl, 100 mM KCl, 200 mM imidazole, 10% glycerol, 10 mM 2-mercapto-ethanol. QB1 buffer (pH 7.0) contained 50 mM $NaPO_4$ and 100 mM NaCl. GF1 buffer (pH 7.6) contained 50 mM Tris-HCl, 1 mM dithiothreitol, 300 mM NaCl, 0.1 mM EDTA, and 10% glycerol. PD-1 buffer (pH 7.6) contained 25 mM Hepes KOH, 5 mM $MgCl_2$, 10% glycerol, and 200 mM KCl. YEG media contained 0.5% yeast extract, 1% NaCl, and 0.4% glucose. 1.5xYT broth (pH 7.0) contained 1.3 % tryptone, 0.75 % yeast extract, and 0.75 % NaCl.

Plasmids and strains

XB-tail-Arc-DAS+4 was cloned into a pET24d vector and consisted of the following sequences from the N- to the C-terminus: (M)GDDRGGRPA <u>LRVVK</u> (XB motif underlined); residues 113-154 of *Escherichia coli* SspB; a H_6 tag; phage P22 Arc repressor; the st11 sequence H_6KNQHD; and a DAS+4 tag (AANDENYSENYADAS). Arc-DAS+4 lacks the XB-tail sequence but is otherwise identical to XB-tail-Arc-DAS+4. Arc-ssrA is identical to Arc-DAS+4 but has a wild-type ssrA tag (AANDENYLAA). FKBP-linker-$ClpX^{\Delta N}$ and FKBP-linker-$[ClpX^{\Delta N}]_3$ were cloned in pACYC vectors and consisted of the N-terminal human FKBP12 protein, followed by residues 139-165 of *E. coli* SspB, a H_6 tag, and either by *E. coli* $ClpX^{\Delta N}$ (residues 61-424) or the covalently linked $[ClpX^{\Delta N}]_3$ trimer (Martin *et al.*, 2005). For studies *in vivo*, FKBP-linker-$[ClpX^{\Delta N}]_3$ was placed under control of a constitutive promoter and ribosome-

binding site (Bba_J23101 and Bba_B0032 respectively, obtained from the Registry of Standard Biological Parts) and expressed from a plasmid bearing a ColE1 origin of replication. The gene for $SspB^{core}$-FRB was generated in a pET vector by fusing the coding sequence for residues 1-113 of *E. coli* SspB to the FRB domain of rat mTOR (residues 2015-2114) with a connecting linker sequence of H_6RGS. Plasmids encoding λO-DHFRII-FRB and λO-titin-I27-FRB were generated by replacing the $SspB^{core}/H_6$ cassette in the $SspB^{core}$-FRB construct with a fragment encoding (M)TNTAKILNFGRS-(DHFRII/titin-I27)-GGSEH_6GS. Standard techniques were used to replace the λO tag in λO-titin-I27-FRB with the sequence MD_6. The GFP-DAS+4, titin-I27-DAS+4 substrates were expressed from pET vectors with N-terminal H_6ID_2LG tags for ease of purification. All growth and degradation experiments *in vivo* were performed in *E. coli* strain X90 [F'*lacI*^{q}*lac' pro'/ara Δ(lac-pro) nalA argE(am) rif*^{R}*thi-1 clpX*$^-$*, recA*$^-$].

Protein expression and purification

E. coli ClpP was expressed and purified as described previously (Kim *et al.*, 2000). Unless noted, all other proteins were over-expressed from IPTG-inducible promoters in *E. coli* strain BLR (BL21 *recA*$^-$ λ(DE3)). Briefly, cells were grown to OD_{600} 0.7 at 37 °C in 1-2 liters of 1.5xYT broth, the cells were chilled to 18 °C, and expression was induced with 1 mM IPTG. Cells were harvested 4 h after induction, resuspended in LB1 buffer (15 mL/liter of culture), and frozen at -80 °C until purification. To aid lysis, cells were subjected to two rounds of freezing and thawing before the addition of 1 mM PMSF, 0.1 mg/mL lysozyme, and 250 Units benzonase nuclease. After a 30-min incubation at 4 °C, lysates were centrifuged at 8000 rpm in a Sorvall SA600 rotor for 40 min, and the supernatant was decanted and saved.

For purification of ClpX and its variants, the lysate supernatant was incubated with 1 mL Ni^{2+}-NTA resin equilibrated with LB1 buffer for 5 min. After two bulk washes with 30 column volumes (CV) of lysis buffer, the slurry was poured into a column, washed with a 20 CVs of LB1 buffer, and the protein was eluted with EB1 buffer (8 x 0.5 CV elutions). Fractions were pooled based on Bradford assays, concentrated using Amicon Ultracel 10k filters, and chromatographed on a Superdex S200 gel-filtration column equilibrated in GF1 buffer. Fractions were analyzed by SDS-PAGE, concentrated, and stored frozen in GF1 buffer at -80 °C.

The GFP-DAS+4, titin-I27-DAS+4, and λO-DHFRII-FRB proteins were purified by Ni^{2+}-NTA chromatography as described above, and stored frozen in EB1 buffer at -80 °C. XB-tail-Arc-DAS+4 was lysed in denaturing buffer LB2. After centrifugation as described above, the soluble fraction was applied to a Ni^{2+}-NTA column, washed with 10 CVs of LB2, 20 CVs of LB1 and eluted as described above. Fractions were pooled based on Bradford analysis, exchanged into QB1 buffer using a spin column, and loaded onto an Amersham 5/50 GL MonoQ column equilibrated in QB1 buffer. The column was washed with five CVs of QB1 buffer and a 10 mL gradient from 0.1 M to 1 M NaCl in QB1 buffer was applied. Appropriate fractions were concentrated, pooled, and stored frozen at -80 °C. For each protein, MALDI-TOF mass spectrometry indicated that the full-length protein without truncations had been purified.

Biochemical assays

Assays were performed in PD-1 buffer at 30 °C using a NADH-coupled colorimetric assay and a plate reader for ATPase assays (Norby, 1988) or an ATP-regeneration system using creatine phosphate for degradation assays (Martin *et al.*, 2008). Degradation assays of XB-tail-Arc-

DAS+4 (10 μM) by 0.3 μM $ClpX_6$ and 0.9 μM $ClpP_{14}$ were quenched by boiling in SDS and monitored by SDS-PAGE. GFP degradation was monitored by loss of fluorescence (excitation 467 nm; emission 511 nm). Degradation of titin-I27-DAS+4 was monitored by release of radioactive peptides soluble in trichloroacetic acid (Gottesman *et al.*, 1998). Log-phase growth rates at 30 °C in YEG media were measured using a plate reader to monitor OD_{600}; the plasmid expressing FKBP-linker-$[ClpX^{\Delta N}]_3$ was maintained using ampicillin (100 μg/mL), and the plasmid expressing λO-DHFRII-FRB was maintained using tetracycline (10 μg/mL). Degradation assays *in vivo* were performed by centrifuging 1 mL of 0.7 OD_{600} cultures, resuspending the pellet in 8 M urea, normalizing each sample by total protein content using a Bradford assay, running SDS-PAGE, transferring by electro-blotting to a PVDF membrane, and probing using a polyclonal anti-DHFRII antibody (a gift from Dr. Elizabeth Howell, U. Tennessee, Knoxville, TN).

Results

Tethering-dependent degradation with no adaptor

SspB uses one part of its structure to bind a substrate and another part (the XB peptide) to bind to the N-domain of ClpX (Dougan *et al.*, 2003; Levchenko *et al.*, 2003; Song and Eck, 2003; Wah *et al.*, 2003; Wojtyra *et al.*, 2003; Park *et al.*, 2007). For a substrate that normally requires delivery by SspB, we reasoned that the need for the substrate-binding portion of this adaptor might be obviated by fusing the XB peptide and the flexible tail of SspB directly to a substrate. In principle, this design would allow the protein to tether itself to the N-domain of ClpX. Efficient ClpXP degradation of such a "substrate" would support a passive tethering model, whereas poor degradation would suggest that the core substrate-binding domain of SspB plays a more active role in delivery.

We constructed and purified a protein consisting of the XB peptide of SspB, the flexible tail of SspB, the Arc repressor protein, and the DAS+4 degradation tag, arranged from the N- to the C-terminus (XB-tail-Arc-DAS+4) The DAS+4 tag (AANDENSENYADAS) is a variant of the ssrA tag that binds SspB normally but binds ClpX with dramatically reduced affinity (McGinness *et al.*, 2006). Proteins containing the DAS+4 tag are degraded in-efficiently by ClpXP, except at very high substrate concentrations or in the presence of SspB. In degradation assays monitored by SDS-PAGE, ClpXP degraded a 33-fold excess of the XB-tail-Arc-DAS+4 protein to near completion over the course of 60 min (Figure 3.2A). By contrast, an otherwise identical variant lacking the XB peptide and tail (Arc-DAS+4) was degraded much more slowly under identical conditions (Figure 3.2B).

Two additional control experiments confirmed that degradation of XB-tail-Arc-DAS+4 depended on tethering of the XB region of the substrate to the N-terminal domain of ClpX. First, $ClpX^{\Delta N}$, a truncated enzyme lacking the N-domain, failed to support ClpP degradation of XB-tail-Arc-DAS+4 (Figure 3.2C) but degraded a tethering-independent substrate (Arc-ssrA) rapidly (Figure 3.2D). Second, addition of free XB peptide substantially slowed degradation of XB-tail-Arc-DAS+4 by wild-type ClpXP (Figure 3.2E) but had no effect on degradation of Arc-ssrA (Figure 3.2F).

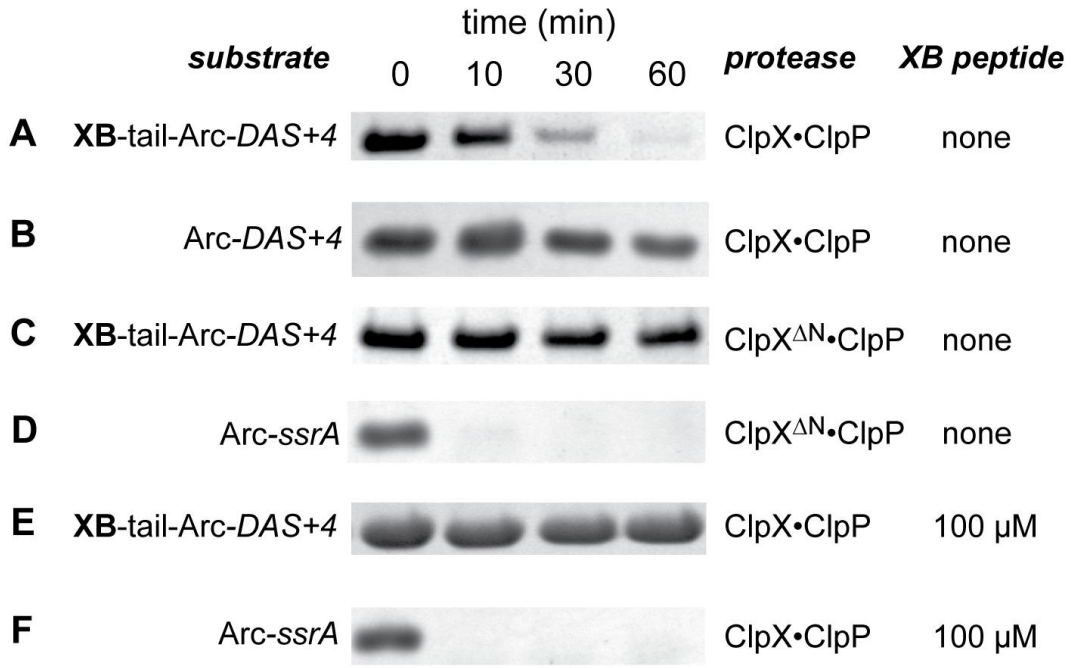

Figure 3.2. SDS-PAGE assays of protein degradation by the ClpXP or $ClpX^{\Delta N}$/ClpP proteases (300 nM ClpX or $ClpX^{\Delta N}$; 900 nM ClpP). The XB-tail-Arc-DAS+4 substrate (10 μM) required auto-tethering to the N-terminal domain of ClpX for efficient degradation. Hence, this substrate was not degraded if the N-domain of ClpX was deleted or if excess XB peptide was present in the proteolysis reaction. Degradation of the Arc-ssrA substrate (10 μM) did not require the ClpX N-domain and was not inhibited by XB peptide.

Taken together, these results suggest that the role of SspB in enhancing substrate degradation is largely one of tethering the substrate to the N-terminal domain of ClpX and increasing its local

concentration. If interactions between SspB and the N-domain are essential for facilitating degradation, then these contacts must be limited to the flexible tail and XB peptide, which comprise the C-terminal 41 residues of SspB.

Artificial tethering supports substrate delivery

Do contacts between the XB peptide and the N-domain of ClpX serve functions other than simple tethering? To address this question, we designed a new binding interface that involved neither the N-domain of ClpX nor the XB peptide. First, we constructed a ClpX variant containing the human FKBP12 protein at the N-terminus, a linker region, and $ClpX^{\Delta N}$ at the C-terminus (FKBP-linker-$ClpX^{\Delta N}$; Figure 3.3A). Second, we fused a SspB variant containing the substrate-binding core domain but lacking the flexible tail and XB motif to the N-terminus of the FRB domain from rat mTor ($SspB^{core}$-FRB). The FKBP12 protein and the FRB domain bind to each other with high affinity only in the presence of the small molecule rapamycin (Choi *et al.*, 1996). Thus, adaptor-mediated tethering in this system is predicted to be rapamycin dependent.

The FKBP-linker-$ClpX^{\Delta N}$ enzyme displayed rates of ATP hydrolysis that increased non-linearly at low protein concentrations (Figure 3.3B). Because N-domain dimerization normally helps stabilize the active hexameric form of ClpX (Grimaud *et al.*, 1998; Wojtyra *et al.*, 2003), replacing this domain with FKBP12 probably resulted in weaker hexamerization. To stabilize the active enzyme, FKBP12 and the linker were fused to the N-terminus of a trimeric form of $ClpX^{\Delta N}$ in which the subunits were connected with a flexible linker (Figure 3.3A) (Martin *et al.*, 2005). Purified FKBP-linker-$[ClpX^{\Delta N}]_3$ exhibited linear ATP-hydrolysis rates at concentrations ranging from 50 nM to 500 nM (Figure 3.3B), consistent with stable pseudo-hexamer formation

at low enzyme concentrations. As observed for wild-type $ClpX_6$ (Kim *et al.*, 2001; Kenniston *et al.*, 2003), ATP hydrolysis by the linked FKBP-linker-$[ClpX^{\Delta N}]_3$ enzyme was slightly repressed by ClpP binding and stimulated by addition of an unfolded substrate (Figure 3.3C). Importantly, FKBP-linker-$[ClpX^{\Delta N}]_3$ and ClpP degraded GFP-ssrA with a K_M of 3.6 μM and a V_{max} of 1.1 min^{-1} enz^{-1} (Figure 3.3D). These steady-state kinetic parameters are similar to those reported for wild-type ClpXP (Kim *et al.*, 2000), showing that the presence of the FKBP12 domain in the linked enzyme does not interfere with degradation of ssrA-tagged substrates.

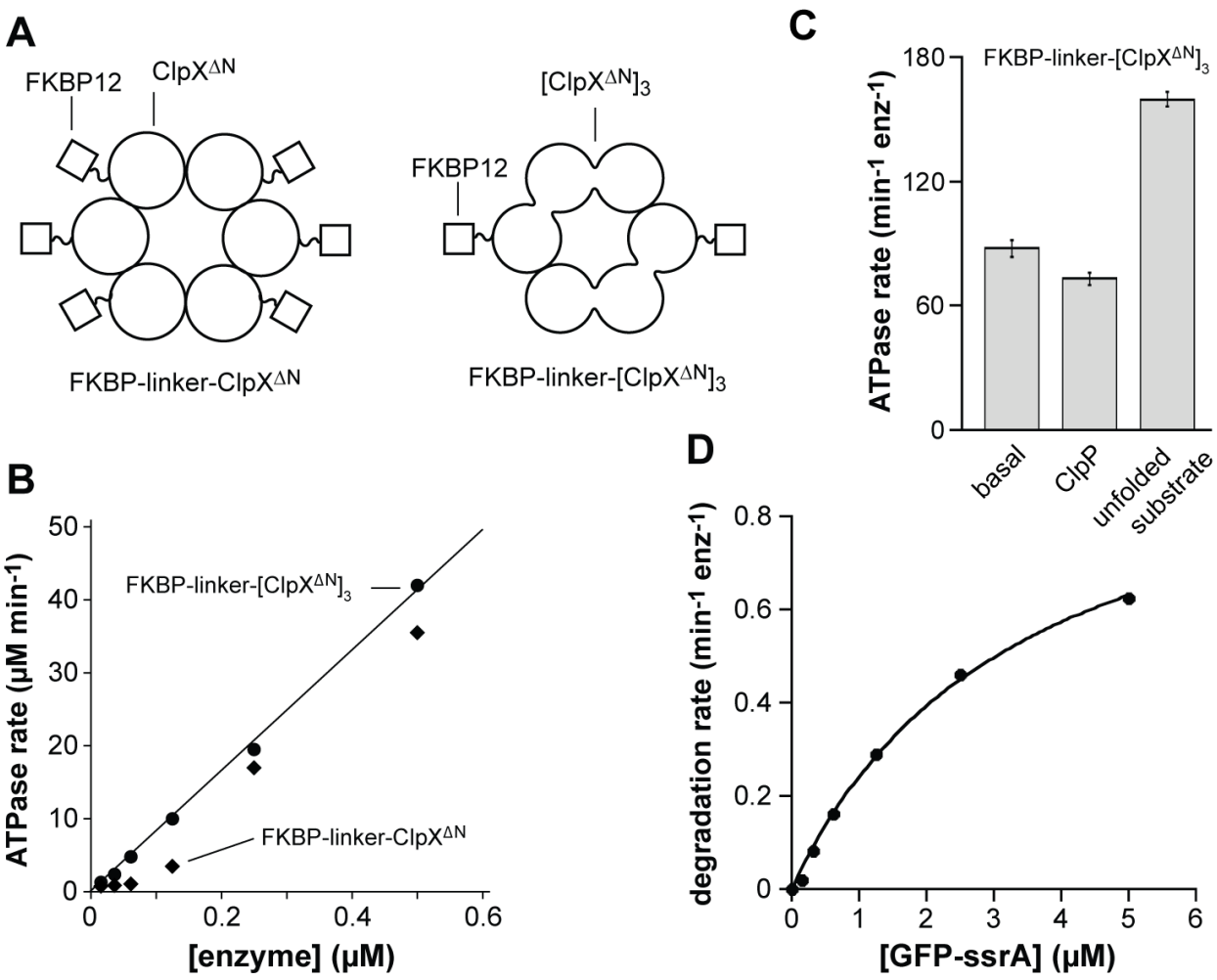

Figure 3.3. (**A**) Cartoon depictions of hexamers formed by the FKBP-linker-ClpX$^{\Delta N}$ or FKBP-linker-[ClpX$^{\Delta N}$]$_3$ enzyme variants. (**B**) Dependence of ATP-hydrolysis rates on enzyme concentration. The rate of ATP hydrolysis by FKBP-linker-[ClpX$^{\Delta N}$]$_3$ was a linear function of enzyme concentration ($R > 0.999$; average rate = 82 ATP min^{-1} enz^{-1}), indicating that this protein forms a stable hexamer at low protein concentrations. ATP hydrolysis by the FKBP-linker-ClpX$^{\Delta N}$ enzyme, by contrast, was highly non linear in a fashion that suggested hexamer dissociation at concentrations below 100 nM. (**C**) The rate of ATP hydrolysis by FKBP-linker-[ClpX$^{\Delta N}$]$_3$ (0.3 μM pseudo hexamer) was slowed by the presence of ClpP (0.9 μM) and enhanced by the presence of an unfolded substrate (1.4 μM), the carboxymethylated-titin-I27-VP15-ssrA protein (Kenniston *et al.*, 2003). (**D**) Michaelis-Menten plot of the substrate dependence of the steady-state rate of degradation of the GFP-ssrA substrate by FKBP-linker-[ClpX$^{\Delta N}$]$_3$ (100 nM pseudo hexamer) and ClpP$_{14}$ (300 nM). The solid line is a non-linear least-squares fit of the experimental data (K_M = 3.6 μM; V_{max} = 1.1 min^{-1} enz^{-1}).

Next, we tested if the artificial $SspB^{core}$-FRB adaptor could deliver substrates to FKBP-linker-$[ClpX^{\Delta N}]_3$ in a rapamycin-dependent fashion (Figure 3.4A). In the presence of this adaptor/enzyme pair, ClpP, and rapamycin, two different DAS+4-tagged substrates were degraded with K_M values near 1 μM and V_{max} values expected based on the resistance of these native proteins to ClpX unfolding (Figure 3.4B & 3.4C) (Kim *et al.*, 2000; Kenniston *et al.*, 2003). In the absence of rapamycin, degradation of both substrates was extremely slow (Figure 3.4B & 3.4C). Similarly, no degradation of untagged GFP or titin-I27 was observed in the presence of $SspB^{core}$-FRB, rapamycin, FKBP-linker-$[ClpX^{\Delta N}]_3$, and ClpP (data not shown). These experiments indicate that artificial tethering mediated by the FRB-rapamycin-FKBP12 complex results in efficient adaptor-dependent degradation. We conclude that simple tethering of substrates to ClpX is sufficient to explain adaptor-mediated enhancement of degradation.

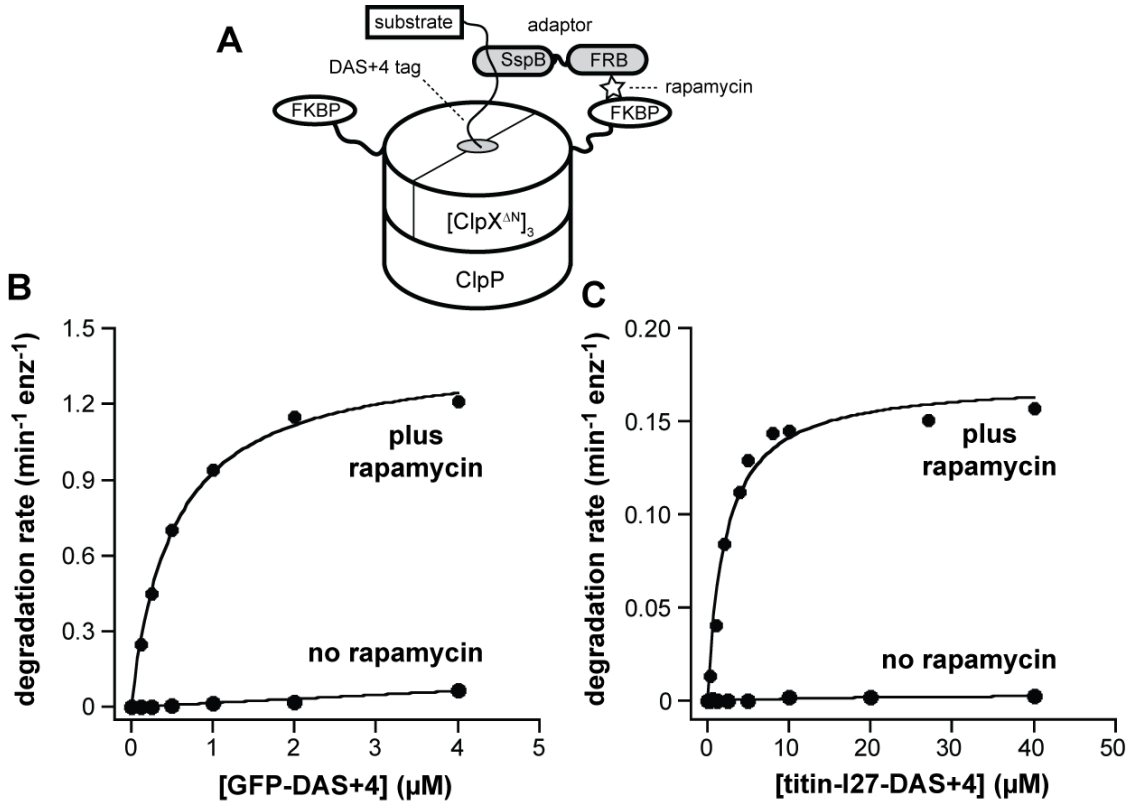

Figure 3.4. **(A)** Cartoon showing delivery of DAS+4 tagged substrate to FKBP-linker-$[ClpX^{\Delta N}]_3$ by the synthetic $SspB^{core}$-FRB adaptor and rapamycin. **(B)** Substrate dependence of GFP-DAS+4 degradation by FKBP-linker-$[ClpX^{\Delta N}]_3$ (100 nM pseudo hexamer) and $ClpP_{14}$ (300 nM) in the presence of $SspB^{core}$-FRB (200 nM) and the presence (1 μM) or absence of rapamycin. The line for the plus-rapamycin curve is a fit to the Michaelis-Menten equation (K_M = 0.51 μM; V_{max} = 1.4 min^{-1} enz^{-1}). The line for the no-rapamycin data is a linear fit. **(C)** Michaelis-Menten plots for titin-I27-DAS+4 degradation by FKBP-linker-$[ClpX^{\Delta N}]_3$ (100 nM pseudo hexamer) and $ClpP_{14}$ (300 nM) in the presence of $SspB^{core}$-FRB (200 nM) and the presence (1 μM) or absence of rapamycin. The fit to the plus-rapamycin data gave K_M = 2.2 μM and V_{max} = 0.17 min^{-1} enz^{-1}. The line for the no-rapamycin data is a linear fit.

Rapamycin-dependent degradation

Artificial tethering mediated by the FRB-rapamycin-FKBP12 interaction requires the interaction of three molecular components. To test the kinetics of assembly, we monitored the fluorescence of the GFP-DAS+4 substrate mixed with $SspB^{core}$-FRB, FKBP-linker-$[ClpX^{\Delta N}]_3$, and ClpP (Figure 3.5A). When rapamycin was added, degradation reached an enhanced steady-state rate within the dead time of the experiment (approximately 20 s; Figure 3.5A). Thus, FRB-rapamycin-FKBP12 binding and subsequent degradation occurs on the time scale of many

biological responses. We also tested the response of the system to rapamycin concentration and found that the degradation rate was unchanged at rapamycin concentrations above the enzyme concentration (Figure 3.5B). This result indicates that near stoichiometric quantities of the small molecule with respect to FKBP-linker-$[ClpX^{\Delta N}]_3$ and $SspB^{core}$-FRB saturate the system. Thus, the response of the system to rapamycin is rapid and sensitive.

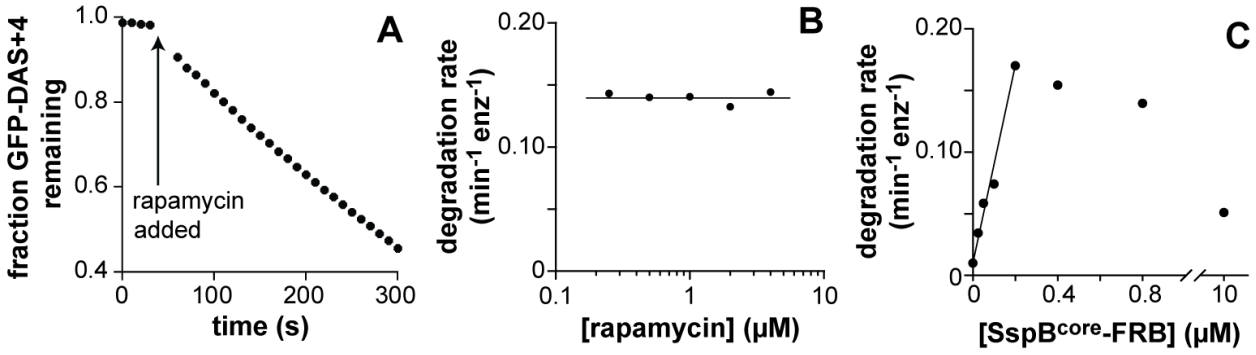

Figure 3.5. **(A)** Rapamycin addition results in rapid degradation. The GFP-DAS+4 substrate (0.5 μM), $SspB^{core}$-FRB (1 μM), FKBP-linker-$[ClpX^{\Delta N}]_3$ (0.5 μM pseudo hexamer), and $ClpP_{14}$ (1.5 μM) were preincubated and slow steady-state degradation was observed by loss of substrate fluorescence. At the time indicated by the arrow, rapamycin (2 μM) was added. Within the dead-time of the experiment (≈ 20 s), degradation of GFP-DAS+4 reached a new and much faster steady-state rate. **(B)** Changing rapamycin concentration over the range shown had almost no effect on the steady-state rate of degradation of ^{35}S-titin-I27-DAS+4 (10 μM) by FKBP-linker-$[ClpX^{\Delta N}]_3$ (100 nM pseudo hexamer), $ClpP_{14}$ (300 nM), and $SspB^{core}$-FRB (200 nM). **(C)** Degradation of ^{35}S-titin-I27-DAS+4 (10 μM) by FKBP-linker-$[ClpX^{\Delta N}]_3$ (200 nM pseudo hexamer) and $ClpP_{14}$ (600 nM) was assayed as a function of the $SspB^{core}$-FRB adaptor concentration in the presence of rapamycin (1 μM).

To evaluate the effects of adaptor concentration, we titrated increasing quantities of the $SspB^{core}$-FRB adaptor against fixed concentrations of FKBP-linker-$[ClpX^{\Delta N}]_3$, ClpP, and rapamycin and assayed degradation of titin-I27-DAS+4. Degradation initially increased linearly and reached a maximal value at a ratio of one $SspB^{core}$-FRB dimer for each FKBP-linker-$[ClpX^{\Delta N}]_3$ pseudo-hexamer. For the wild-type proteins, it is known that one substrate-bound SspB dimer binds to one ClpX hexamer (Wah *et al.*, 2002). At high concentrations of the $SspB^{core}$-FRB adaptor (10

μM), the degradation rate was reduced to approximately 25% of its maximal value (Figure 3.5C). Under these conditions, free adaptor molecules probably compete with ClpX-bound adaptors for substrate binding.

Tethering-dependent delivery of a λO-tagged substrate

The results presented so far show that alternative mechanisms of tethering can facilitate ClpXP proteolysis of substrates with C-terminal degradation tags. Degradation of proteins bearing the N-terminal λO tag normally requires the N-domain of ClpX (Singh *et al.*, 2001), which is missing from the FKBP-linker-$[ClpX^{\Delta N}]_3$ variant. To test the versatility of the artificial tethering system, we constructed and purified a substrate with an N-terminal λO tag (NH_2-TNTAKILNFGR; Flynn *et al.*, 2003), followed by the titin-I27 domain, and then the FRB domain. This λO-titin-I27-FRB substrate contains no sequences from the SspB adaptor but can tether itself via rapamycin to the FKBP-linker-$[ClpX^{\Delta N}]_3$ enzyme.

The purified λO-titin-I27-FRB fusion protein was degraded by FKBP-linker-$[ClpX^{\Delta N}]_3$ and ClpP when rapamycin was present but not when it was absent (Figure 3.6A). As a control, we constructed an otherwise identical fusion protein in which the λO tag was replaced by the sequence MD_6 (MDDDDDD) which we hypothesized would not be recognized by ClpX. No degradation of this protein was observed in the presence or absence of rapamycin (Figure 3.6A). We conclude that an N-terminal λO tag can serve as a degradation signal for substrates which can tether themselves to ClpX.

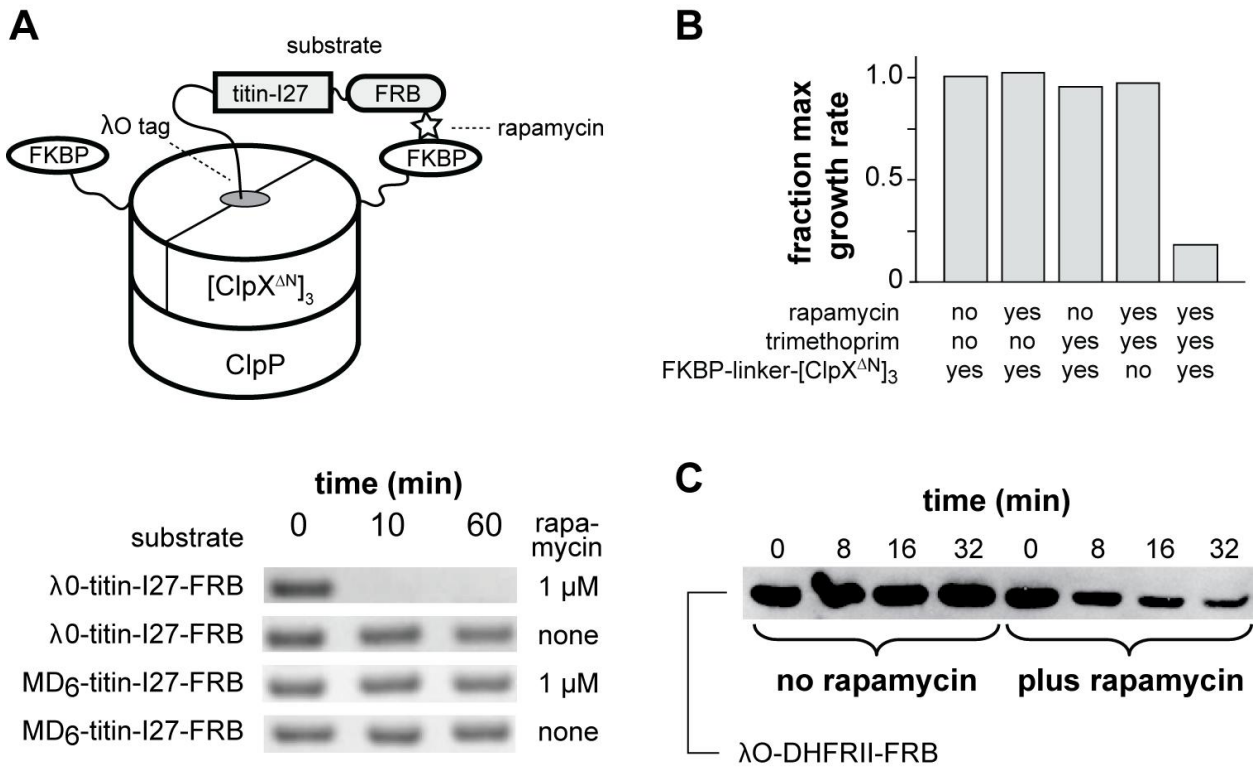

Figure 3.6. (A) Cartoon showing rapamycin-dependent degradation of a substrate with an N-terminal λO tag and C-terminal FRB domain by FKBP-linker-$[ClpX^{\Delta N}]_3$ and ClpP. The *gel below* the cartoon shows that variants of the titin-I27 protein with an N-terminal λO tag and C-terminal FRB domain were degraded in a rapamycin-dependent fashion *in vitro* as assayed by SDS-PAGE. Replacing the λO tag with a MD_6 sequence blocked degradation. All reactions contained substrate (10 μM), FKBP-linker-$[ClpX^{\Delta N}]_3$ (300 nM pseudo hexamer), and $ClpP_{14}$ (900 nM). **(B)** Relative growth rates in M9 minimal medium of *clpX*¯ *E. coli* strains expressing λO-DHFRII-FRB as a function of rapamycin (10 μM, when present), trimethoprim (100 μg/mL, when present), and a plasmid expressing FKBP-linker-$[ClpX^{\Delta N}]_3$. **(C)** Western blotting shows intracellular degradation of λO-DHFRII-FRB after addition of rapamycin (10 μM) but not after a mock addition in *clpX*¯ *E. coli* strains expressing FKBP-linker-$[ClpX^{\Delta N}]_3$. Protein synthesis was not blocked in this experiment. The sample volume in each lane was adjusted to yield the same amount of total cellular protein.

Rapamycin-dependent degradation *in vivo*

For studies of degradation in *E. coli*, we constructed a substrate with an N-terminal λO-tag, followed by the DHFRII enzyme and the FRB domain (λO-DHFRII-FRB). This substrate was degraded in a rapamycin-dependent manner by FKBP-linker-$[ClpX^{\Delta N}]_3$ and ClpP *in vitro* (data

not shown). Expression of DHFRII in *E. coli* results in resistance to trimethoprim, an antibiotic that inhibits the endogenous dihydrofolate reductase enzyme (Fleming *et al.*, 1972; Pattishall *et al.*, 1977; Smith *et al.*, 1979; Stone and Smith, 1979; Krahn *et al.*, 2007). When we expressed λO-DHFRII-FRB in a *clpX*$^-$ strain containing FKBP-linker-[ClpX$^{\Delta N}$]$_3$, the cells grew well in the presence of trimethoprim (Figure 3.6B). When rapamycin was added, however, these cells became trimethoprim sensitive and growth slowed substantially (Figure 3.6B). This result was not caused by rapamycin toxicity, as rapamycin did not affect growth in media lacking trimethoprim or in a strain without FKBP-linker-[ClpX$^{\Delta N}$]$_3$ (Figure 3.6B). Thus, the FKBP-linker-[ClpX$^{\Delta N}$]$_3$ enzyme and ClpP appear to degrade λO-DHFRII-FRB in a rapamycin-dependent fashion *in vivo*. To confirm this inference, we prepared cell lysates at different times after treatment with rapamycin or a mock-addition and subjected them to SDS-PAGE and Western blotting using an anti-DHFRII antibody. The steady-state level of λO-DHFRII-FRB was reduced rapidly upon addition of rapamycin (Figure 3.6C), even though protein synthesis was not blocked in this experiment. Hence, rapamycin-dependent tethering of FRB-fusion substrates to FKBP-linker-[ClpX$^{\Delta N}$]$_3$ can successfully control degradation in the cell.

Discussion

The results presented here demonstrate that tethering of a substrate to a AAA+ protease is sufficient for adaptor function. We found, for example, that the normal tethering function of the SspB adaptor protein could be transferred directly to a substrate by fusing the ClpX-binding peptide and tail of SspB to an otherwise poor substrate for the ClpXP protease. This result shows that the substrate binding domain of SspB is dispensable for adaptor function. Moreover, we were able to engineer a completely artificial tethering system, in which the N-domain of ClpX was replaced with the FKBP12 protein and the normal tail and ClpX-binding peptide of the SspB adaptor was replaced with the FRB domain. In the presence of rapamycin, a small bridging molecule, substrates bearing a weak degradation tag were degraded efficiently in the presence of this artificial adaptor and re-engineered enzyme. Finally, we established a synthetic rapamycin-dependent system in which substrates bearing an N-terminal λO-degradation tag and the FRB domain were efficiently degraded by a FKBP-ClpX variant and ClpP. It is important to note that this substrate contained no parts from the SspB adaptor and ClpX lacked its N-domain, which is required for normal adaptor-mediated delivery of substrates to wild-type ClpX. Because the functions mediated by SspB and the ClpX N-domain can effectively be replaced by other tethering elements, it is possible that the wild-type adaptor system stimulates proteolysis simply by tethering substrates to ClpXP.

Although tethering of substrates to ClpX appears to be sufficient for adaptor function, there clearly must also be other geometric and structural criteria that need to be met (McGinness *et al.*, 2007). For example, a potential substrate that is tethered too rigidly and therefore could not reach the translocation pore would presumably not be degraded efficiently. In the studies reported here,

we tried to account for this factor by engineering a flexible linker between the FKBP12 and ClpX portions of our synthetic enzyme. In some cases, however, it may also be necessary to design flexibility into the adaptor or substrate as well.

Efficient ClpXP degradation mediated by the λO tag normally requires both a multimeric substrate and the N-domain of ClpX. For example, the tetrameric λO protein is degraded well by wild-type ClpXP but is not degraded efficiently by $ClpX^{\Delta N}$ and ClpP (Singh *et al.*, 2001; Wojtyra *et al.*, 2003). Moreover, K_M for ClpXP degradation of a dimeric λO-tagged substrate was substantially lower than that for a monomeric λO-tagged protein (Farrell *et al.*, 2007). Nevertheless, we found that a monomeric substrate, λO-titin-I27-FRB, was degraded efficiently by FKBP-linker-$[ClpX^{\Delta N}]_3$ and ClpP. This result suggests that degradation of λO-tagged substrates does not require substrate multimerization or the ClpX N-domain if alternative mechanisms of tethering to ClpX are available. We propose that λO-tagged substrates normally require multimerization for efficient degradation, because one λO tag in a multimer tethers the substrate to ClpX via the N-domain, allowing a second λO tag to be engaged by the translocation channel of ClpX to initiate degradation. Indeed, this dual-function tag model is supported by reports that both the isolated N-domain of ClpX and $ClpX^{\Delta N}$ itself bind to the λO protein (Singh *et al.*, 2001; Wojtyra *et al.*, 2003).

Multivalent recognition of degradation signals is likely to facilitate the targeted proteolysis of many substrates at low concentrations. For example, monomeric substrates might contain more than one type of degradation tag. Indeed, Flynn *et al.* reported that roughly 25% of identified cellular substrates for ClpXP contained more than one degradation motif (Flynn *et al.*, 2003). For

multimeric substrates, the same sequence signal in two different subunits could be recognized by different parts of the protease as proposed above. We found, for example, that *clpX*$^+$ strains containing λO-DHFRII-FRB were trimethoprim sensitive (data not shown), suggesting that wild-type ClpXP can degrade this substrate. Because DHFRII is a tetramer, it is likely that a λO tag from one subunit tethers the substrate to the N-domain of ClpX, allowing the λO tag from a second subunit to be engaged by ClpX for degradation. Recent studies indicate that ClpX disassembly of the tetrameric MuA transposase also involves recognition of multiple classes of sequence elements (Abdelhakim *et al.*, 2008).

Furthermore, our results demonstrate that adaptor-mediated degradation can be placed under small-molecule control both *in vitro* and *in vivo*. In the cell, the kinetics of rapamycin uptake and the subsequent assembly of the FRB-rapamycin-FKBP12 complex are sufficiently rapid to ensure that degradation starts within minutes of addition of the small molecule. The system described here may not be optimal for many desired uses in controlling intracellular degradation. Nevertheless, our results suggest that it should be possible to design improved systems that combine small-molecule control, fast temporal responses, and the exquisite specificity of a genetically encoded degradation tag. We expect such systems to be useful in the study of protein products, including those encoded by essential genes, for which the phenotype of depletion on a short-time scale needs to be determined. In principle, such a system could be used with either N- or C-terminal tethering-dependent tags, greatly extending their applicability.

Acknowledgements

We thank Kathleen McGinness, Andreas Martin, Kevin Griffith, Eyal Gur, Peter Chien, Shankar Sundar, and Elizabeth Howell for helpful discussions or materials. Supported by NIH grants AI-16892 and GM-48224. T.A.B is an employee of the Howard Hughes Medical Institute.

References

1. Abdelhakim, A. H., Oakes, E. C., Sauer, R. T., and Baker, T. A. 2008. Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA+ unfoldase ClpX. *Mol Cell* **30**(1): 39-50.

2. Baker, T. A., and Sauer, R. T. 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**(12): 647-653.

3. Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G., and Wandless, T. J. 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**(5): 995-1004.

4. Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A., and Sauer, R. T. 2004. Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol Cell* **13**(3): 443-449.

5. Bukau, B., Weissman, J., and Horwich, A. 2006. Molecular chaperones and protein quality control. *Cell* **125**(3): 443-451.

6. Choi, J., Chen, J., Schreiber, S. L., and Clardy, J. 1996. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**(5272): 239-242.

7. Dougan, D. A., Reid, B. G., Horwich, A. L., and Bukau, B. 2002. ClpS, a substrate modulator of the ClpAP machine. *Mol Cell* **9**(3): 673-683.

8. Dougan, D. A., Weber-Ban, E., and Bukau, B. 2003. Targeted delivery of an ssrA-tagged substrate by the adaptor protein SspB to its cognate AAA+ protein ClpX. *Mol Cell* **12**(2): 373-380.

9. Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A., and Bukau, B. 2006. ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**(7077): 753-756.

10. Farrell, C. M., Baker, T. A., and Sauer, R. T. 2007. Altered specificity of a AAA+ protease. *Mol Cell* **25**(1): 161-166.

11. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

12. Fleming, M. P., Datta, N., and Gruneberg, R. N. 1972. Trimethoprim resistance determined by R factors. *Br Med J* **1**(5802): 726-728.

13. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. 2003. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**(3): 671-683.

14. Flynn, J. M., Levchenko, I., Sauer, R. T., and Baker, T. A. 2004. Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes Dev* **18**(18): 2292-2301.

15. Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**(9): 1338-1347.

16. Gottesman, S. 2003. Proteolysis in bacterial regulatory circuits. *Annu Rev Cell Dev Biol* **19**565-587.

17. Griffith, K. L., and Grossman, A. D. 2008. Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Mol Microbiol* **70**(4): 1012-1025.

18. Grimaud, R., Kessel, M., Beuron, F., Steven, A. C., and Maurizi, M. R. 1998. Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *J Biol Chem* **273**(20): 12476-12481.

19. Hou, J. Y., Sauer, R. T., and Baker, T. A. 2008. Distinct structural elements of the adaptor ClpS are required for regulating degradation by ClpAP. *Nat Struct Mol Biol* **15**(3): 288-294.

20. Janse, D. M., Crosas, B., Finley, D., and Church, G. M. 2004. Localization to the proteasome is sufficient for degradation. *J Biol Chem* **279**(20): 21415-21420.

21. Kenniston, J. A., Baker, T. A., Fernandez, J. M., and Sauer, R. T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**(4): 511-520.

22. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. 2000. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**(4): 639-648.

23. Kim, Y. I., Levchenko, I., Fraczkowska, K., Woodruff, R. V., Sauer, R. T., and Baker, T. A. 2001. Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol* **8**(3): 230-233.

24. Kirstein, J., Schlothauer, T., Dougan, D. A., Lilie, H., Tischendorf, G., Mogk, A., Bukau, B., and Turgay, K. 2006. Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *Embo J* **25**(7): 1481-1491.

25. Krahn, J. M., Jackson, M. R., DeRose, E. F., Howell, E. E., and London, R. E. 2007. Crystal structure of a type II dihydrofolate reductase catalytic ternary complex. *Biochemistry* **46**(51): 14878-14888.

26. Levchenko, I., Seidel, M., Sauer, R. T., and Baker, T. A. 2000. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**(5488): 2354-2356.

27. Levchenko, I., Grant, R. A., Wah, D. A., Sauer, R. T., and Baker, T. A. 2003. Structure of a delivery protein for an AAA+ protease in complex with a peptide degradation tag. *Mol Cell* **12**(2): 365-372.

28. Levchenko, I., Grant, R. A., Flynn, J. M., Sauer, R. T., and Baker, T. A. 2005. Versatile modes of peptide recognition by the AAA+ adaptor protein SspB. *Nat Struct Mol Biol* **12**(6): 520-525.

29. Martin, A., Baker, T. A., and Sauer, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**(7062): 1115-1120.

30. Martin, A., Baker, T. A., and Sauer, R. T. 2008. Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol Cell* **29**(4): 441-450.

31. McGinness, K. E., Baker, T. A., and Sauer, R. T. 2006. Engineering controllable protein degradation. *Mol Cell* **22**(5): 701-707.

32. McGinness, K. E., Bolon, D. N., Kaganovich, M., Baker, T. A., and Sauer, R. T. 2007. Altered tethering of the SspB adaptor to the ClpXP protease causes changes in substrate delivery. *J Biol Chem* **282**(15): 11465-11473.

33. Neher, S. B., Flynn, J. M., Sauer, R. T., and Baker, T. A. 2003. Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes Dev* **17**(9): 1084-1089.

34. Norby, J. G. 1988. Coupled assay of Na+,K+-ATPase activity. *Methods Enzymol* **156**116-119.

35. Park, E. Y., Lee, B. G., Hong, S. B., Kim, H. W., Jeon, H., and Song, H. K. 2007. Structural basis of SspB-tail recognition by the zinc binding domain of ClpX. *J Mol Biol* **367**(2): 514-526.

36. Pattishall, K. H., Acar, J., Burchall, J. J., Goldstein, F. W., and Harvey, R. J. 1977. Two distinct types of trimethoprim-resistant dihydrofolate reductase specified by R-plasmids of different compatibility groups. *J Biol Chem* **252**(7): 2319-2323.

37. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

38. Singh, S. K., Rozycki, J., Ortega, J., Ishikawa, T., Lo, J., Steven, A. C., and Maurizi, M. R. 2001. Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *J Biol Chem* **276**(31): 29420-29429.

39. Smith, S. L., Stone, D., Novak, P., Baccanari, D. P., and Burchall, J. J. 1979. R plasmid dihydrofolate reductase with subunit structure. *J Biol Chem* **254**(14): 6222-6225.

40. Song, H. K., and Eck, M. J. 2003. Structural basis of degradation signal recognition by SspB, a specificity-enhancing factor for the ClpXP proteolytic machine. *Mol Cell* **12**(1): 75-86.

41. Stone, D., and Smith, S. L. 1979. The amino acid sequence of the trimethoprim-resistant dihydrofolate reductase specified in Escherichia coli by R-plasmid R67. *J Biol Chem* **254**(21): 10857-10861.

42. Thibault, G., Tsitrin, Y., Davidson, T., Gribun, A., and Houry, W. A. 2006. Large nucleotide-dependent movement of the N-terminal domain of the ClpX chaperone. *Embo J* **25**(14): 3367-3376.

43. Wah, D. A., Levchenko, I., Baker, T. A., and Sauer, R. T. 2002. Characterization of a specificity factor for an AAA+ ATPase: assembly of SspB dimers with ssrA-tagged proteins and the ClpX hexamer. *Chem Biol* **9**(11): 1237-1245.

44. Wah, D. A., Levchenko, I., Rieckhof, G. E., Bolon, D. N., Baker, T. A., and Sauer, R. T. 2003. Flexible linkers leash the substrate binding domain of SspB to a peptide module that stabilizes delivery complexes with the AAA+ ClpXP protease. *Mol Cell* **12**(2): 355-363.

45. Walsh, D. P., and Chang, Y. T. 2006. Chemical genetics. *Chem Rev* **106**(6): 2476-2530.

46. Wang, K. H., Sauer, R. T., and Baker, T. A. 2007. ClpS modulates but is not essential for bacterial N-end rule degradation. *Genes Dev* **21**(4): 403-408.

47. Wojtyra, U. A., Thibault, G., Tuite, A., and Houry, W. A. 2003. The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J Biol Chem* **278**(49): 48981-48990.

Chapter 4

Design, Construction, and Characterization of a Set of Insulated Bacterial Promoters

This work will be submitted for publication with Joseph H. Davis, Adam J. Rubin, and Robert T. Sauer as authors.

J.H.D and A.J.R designed and conducted the experiment described herein. J.H.D and R.T.S. wrote and edited this manuscript.

Abstract

We have generated a series of variable-strength, constitutive, bacterial promoters that act predictably in different sequence contexts, span two orders of magnitude in strength, and contain convenient sites for cloning and the introduction of downstream open-reading frames. Importantly, their design insulates these promoters from the stimulatory or repressive effects of many 5′ or 3′ sequence elements. We show that different promoters from our library produce constant relative levels of two different proteins in multiple genetic contexts. This set of promoters should be a useful resource for the synthetic-biology community.

Introduction

The introduction of novel genetic components and pathways into cells has proven useful in biotechnology and as a tool to study and improve our understanding of natural systems (Elowitz and Leibler, 2000; Ro *et al.*, 2006; Atsumi and Liao, 2008; Stricker *et al.*, 2008). For some applications, achieving the proper steady-state levels of each gene product can be critical in optimizing the function of an entire biosynthetic pathway, whereas, in other cases, assaying the consequences of altered expression levels is important for probing native gene function (Alper *et al.*, 2005; Lutke-Eversloh and Stephanopoulos, 2008; Anthony *et al.*, 2009).

In principle, steady-state protein levels can be controlled by using libraries of variable-strength promoters to change transcription rates, by employing different ribosome-binding sites to alter translation efficiency, and by appending degradation tags to adjust rates of protein turnover (Alper *et al.*, 2005; Wong *et al.*, 2007; Salis *et al.*, 2009). Often, the process of finely-tuning each of these parameters is laborious and relies on trial-and-error, a problem that has led some to utilize directed-evolution to guide the optimization process (Yokobayashi *et al.*, 2002; You *et al.*, 2004; Basu *et al.*, 2005). Others have begun to characterize individual genetic parts rigorously (for example, the transcriptional strength of a promoter) with the hope that such information might guide and expedite the refinement process (Endy, 2005; Canton *et al.*, 2008; Kelly *et al.*, 2009). The success of reusing well-characterized components relies on a critical assumption that such devices are functionally composable, that is the properties of the device in one test context are predictive of those properties in a new context.

Building on the work of others, we sought to design a set of variable-strength constitutive bacterial promoters that are insulated from influences of genomic context. Although it may not be possible to insulate any biological component completely, a wealth of information is available that can be exploited to limit context-dependent behavior. Bacterial transcription can be decomposed into three phases; binding, initiation, and elongation. Extensive biochemical and structural studies have helped elucidate the promoter and polymerase components that control each of these steps (Browning and Busby, 2004). Bacterial RNA polymerase (RNAP) is a heterohexamer composed of a core polymerase (ßß´$\alpha_2\omega$), which is competent for transcriptional elongation, and a σ subunit, which is utilized to define promoter specificity during binding and initiation but is dispensable for elongation. In the initial binding step, the σ subunit of the RNA polymerase holoenzyme (ßß´$\alpha_2\omega$-σ) contacts two hexameric DNA sequences located 10 and 35 base pairs 5´ of the transcription start site (named the -10 and -35 boxes, respectively) (deHaseth *et al.*, 1998). At some promoters, additional contacts are formed between the α_2 subunits and A,T-rich promoter sequences residing as many as 60 base pairs 5´ of the transcription-start site (known as an "UP" sequence). These contacts facilitate polymerase binding and can enhance promoter activity up to 300-fold in manner that depends upon the sequence distance from the core recognition elements (Estrem *et al.*, 1998; Estrem *et al.*, 1999; Ross *et al.*, 2001). Once bound, the enzyme-promoter complex must isomerize from a "closed" complex in which the DNA is double stranded to an "open" complex in which base pairs from approximately positions –10 to +2 melt or separate into single strands (deHaseth *et al.*, 1998). The conformational equilibrium between closed and open complexes depends, in part, on the sequence of the base pairs that melt (Pemberton *et al.*, 2000). From the open complex, the polymerase can undergo a repetitive process, termed abortive transcription, in which it initiates transcription, releases a

short RNA transcript (<10 nt), and then returns to the site of initiation. This process continues in a stochastic fashion until the polymerase clears the promoter, releases the σ subunit, and continues the elongation phase of transcription (often referred to as promoter escape). The identity of the 20 nucleotides downstream of the transcription-start site (defined as +1 to +20) strongly influence the efficiency with which a bacterial polymerase escapes from the promoter and continues elongation (Chan and Gross, 2001; Hsu, 2008).

To respond to external signals, transcription factors have evolved mechanisms to up-regulate or down-regulate each of the steps described above. For example, cyclic-AMP receptor protein (CRP) binds the promoter upstream of the -35 box, simultaneously making favorable contacts with the C-terminal domain of the polymerase α subunits (de Crombrugghe *et al.*, 1984; Lawson *et al.*, 2004). These contacts enhance polymerase binding and increase transcription at promoters where binding is rate limiting (Ebright, 1993). Promoters can also be regulated at the isomerization step of transcription. For example, the merR transcriptional repressor acts by inhibiting conversion of the closed to open complex (Rojo, 1999). Lastly, the efficiency of promoter escape can be regulated. At some promoters, for example, the phage φ29 protein p4 binds tightly to the polymerase α subunit, which slows promoter escape and down-regulates productive transcription (Rojo, 1999). Interestingly, at promoters where polymerase binding is limiting, this same protein can activate transcription by enhancing binding (Mencia *et al.*, 1996). Such non-uniform effects are particularly important when considering the design of synthetic variable-strength promoter libraries.

Many synthetic promoters have been generated and characterized, but they often show activities that vary with the genetic locus or gene transcribed (Jensen and Hammer, 1998; Alper *et al.*, 2005; Hammer *et al.*, 2006; Anderson, 2009; Kelly *et al.*, 2009; Martin *et al.*, 2009). As discussed above, this context dependence is not surprising. For example, increased activity can arise from 5′ UP-sequence elements or as a consequence of read-through from upstream promoters (Estrem *et al.*, 1998; Meng *et al.*, 2001). One solution to the latter problem is to include a transcriptional terminator at the 5′ boundary of the synthetic promoter, but AT-rich terminators share some sequence similarity with UP elements and may themselves increase transcription of downstream genes. Promoter fusions to different genes may also affect transcription efficiency if the 5′ end of the mRNA contains a sequence that changes the rate of promoter escape.

The strength of any minimal promoter, containing sequences from the -35 hexamer to the site of initiation, is likely to vary depending on neighboring sequences. As a consequence, we generated promoters in which adjoining upstream and downstream sequences, which potentially could alter transcription initiation and promoter escape, were included in the promoter cassette (Figure 4.1). These "insulated" promoter cassettes extend from position –105 to +55, a span which includes the majority of transcription-factor binding sites in natural bacterial promoters as well as most elements that affect transcription initiation and promoter escape (Mendoza-Vargas *et al.*, 2009). We find that such promoters exhibit transcriptional activities that are significantly more predictable in varied genetic contexts.

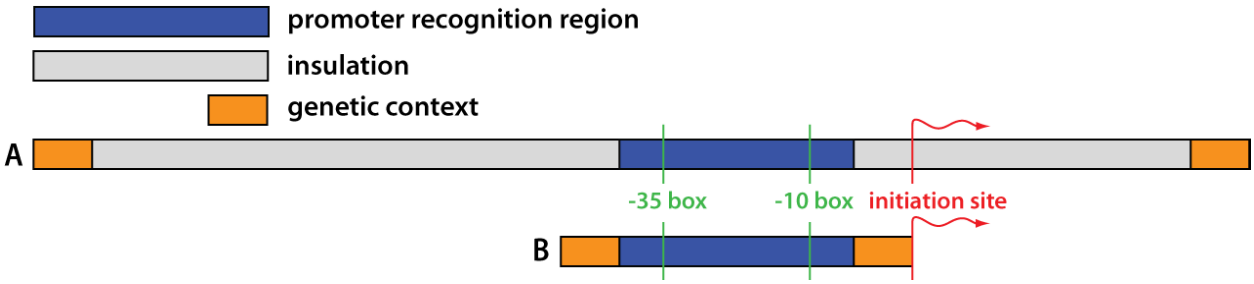

Figure 4.1. Comparison of promoter organization. Schematic of an insulated promoter (**A)** and a minimal, uninsulated promoter (**B)**. The promoter recognition region (PRR) containing the -10 and -35 RNAP binding determinants (green) is shown in blue, the transcription initiation site (+1) is represented by the red line. In the insulated promoter, the surrounding genetic context (orange) is separated from the PRR by insulation sequences (grey). Most elements known to effect transcription initiation and promoter escape are contained within the insulated promoter cassette boundaries. Because of its smaller size, the genetic context surrounding the minimal promoter is more likely to contain sequences that can effect transcription initiation, thereby increasing the possibility of context-dependent activity.

Results

Basic promoter design

Our initial construct was based on *E. coli* rrnB P1, a strong σ^{70}-dependent promoter with near consensus –10 (TATAAT) and –35 (TTtACg) elements. A 17-bp sequence (GGCATGCATAAGGCTCG) separates the -10 and -35 boxes, resulting in the optimal spacing for near-consensus promoters (Aoyama *et al.*, 1983; Rossi *et al.*, 1983; Paul *et al.*, 2004). To provide insulation, we also defined the flanking sequences from position –105 to +55, using elements described in *Materials and Methods*. Our promoter design extends beyond the transcription initiation start site and thus creates a specific and invariant 5´ mRNA terminus. This feature was incorporated to improve the predictability of promoter strength, mRNA stability, and the site of transcriptional initiation, but it may preclude experiments in which this region of the transcript must be of a particular sequence (Win *et al.*, 2009). We note however that the ribosome binding site and translation start codon reside downstream of this element, and thus the resulting protein product is not affected by the insulation sequence. It is also interesting to note that the vast majority of natural transcripts in *E. coli* are predicted to encode a 5´ untranslated region greater than 20 nucleotides in length (Mendoza-Vargas *et al.*, 2009).

We transformed cells with a plasmid vector bearing our first-generation insulated promoter (called proD) driving production of a GFP reporter gene (Bba_E0040, Registry of Standard Biological Parts, www.partsregistry.org) and measured GFP synthesis rates as a proxy for promoter strength. Briefly, cells were grown in culture tubes at 37 °C to mid-log phase, OD_{600} was measured as a surrogate of cell number, and GFP fluorescence was determined. After an

additional 1.25 hours of growth, we again measured OD_{600} and GFP fluorescence. The GFP synthesis rate for the promoter was calculated using equation 4.1 (Kelly *et al.*, 2009).

$$SynthesisRate_x = \frac{GFP(x)_{tp\,2} - GFP(x)_{tp\,1}}{OD_{600}(x)_{avera\ ge}} \qquad (4.1)$$

To determine the strength of proD relative to another promoter, we performed the same procedure using a minimal length, constitutive promoter (Bba_j23101: Registry of Standard Biological Parts), which has been previously characterized (Kelly *et al.*, 2009). In this comparison, proD had greater activity than the reference promoter (Figure 4.2).

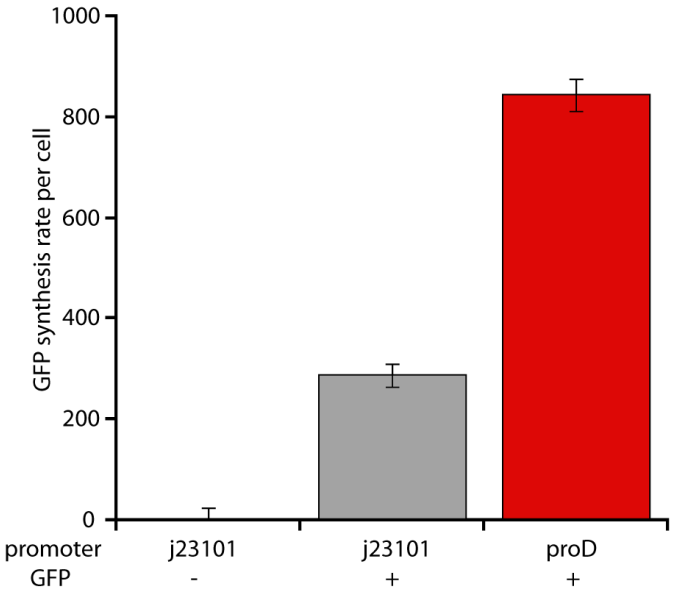

Figure 4.2. Insulated promoters drive production of GFP. GFP synthesis rates per cell were measured for a control construct lacking GFP, a minimal promoter (j23101), or an insulated promoter (proD).

Generation and characterization of an insulated promoter library

Next, we generated a library by using degenerate oligonucleotides to randomize either the –35 or the –10 element of our insulated promoter in the plasmid vector and transformed *E. coli* cells. This library was enriched for active promoters using fluorescence activated cell sorting (data not shown). Measurements of GFP synthesis rates from individual colonies of the enriched pool were then used to identify 10 clones that exhibited varied GFP expression. Measurement of GFP

synthesis rates showed that this set of variants encoded promoter strengths spanning two orders of magnitude. To allow for comparison with previously characterized minimal promoters, we also measured the activity of a set of uninsulated promoters which spanned a similar range of activity (Bba_j23113, Bba_j23150, Bba_j23151 and Bba_j23101 from The Registry of Standard Biological Parts, www.partsregistry.org). As described in Kelly *et al.* (2009), relative measures of promoter activity can greatly reduce assay-to-assay variance. We determined relative promoter strength by normalizing the GFP synthesis rate of each promoter to that of a reference proD promoter, as shown in Equation 4.2. The sequence of each promoter variant as well as the relative promoter strength is listed in Table 4.1 in *Materials and Methods*.

$$RPU_D(x) = \frac{SynthesisRate_x}{SynthesisRate_D} = \frac{\frac{GFP(x)_{tp2} - GFP(x)_{tp1}}{OD_{600}(x)_{average}}}{\frac{GFP(D)_{tp2} - GFP(D)_{tp1}}{OD_{600}(D)_{average}}} \quad (4.2)$$

From our set of ten insulated variants, we selected proD and three additional promoters (proA, proB, and proC), which spanned the activity range, for more detailed characterization (Figure 4.3). These promoters were named in ascending order of activity.

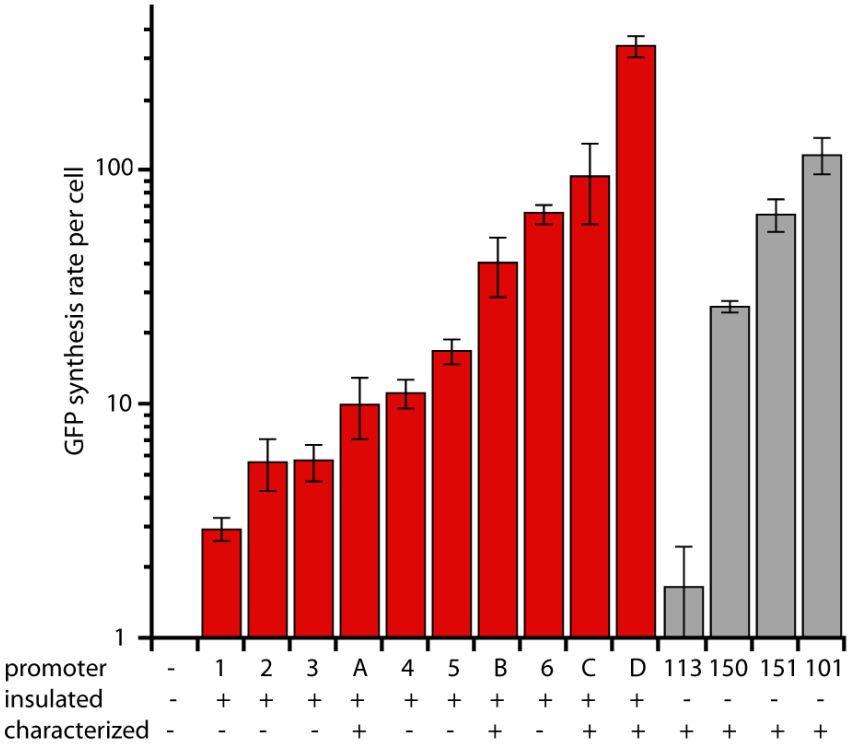

Figure 4.3. Promoter strength of library members. The strength of each promoter was measured in triplicate in *E. coli* DH5α grown in minimal media using GFP as a reporter (Kelly *et al.*, 2009). The GFP synthesis rate is reported on a $\log_{10}$ scale as a surrogate of promoter strength. Promoter strength measurements relative to proD are listed in Table 4.1 in *Materials and Methods*.

Characterization of promoter insulation

The UP element from *rrnB* P1 was chosen to test the efficacy of the 5´ insulation. This 24-nucleotide sequence (AGAAAATTATTTTAAATTTCCTCA) has been shown to activate transcription from some promoters (Meng *et al.*, 2001). We inserted the UP element at the 5´ boundaries of the insulated promoter cassettes (proA, proB, proC, proD) and the noninsulated promoter cassettes (j23113, j23101, j23150, j23151). Using the GFP-reporter assay, we determined the relative strength of each promoter either with or without the UP element (Figure 4.4). Introduction of the UP sequence slightly reduced transcription from each insulated promoter compared to the same promoter with no UP element, whereas it increased transcription from the uninsulated promoters in a highly variable manner (Figure 4.4 inset).

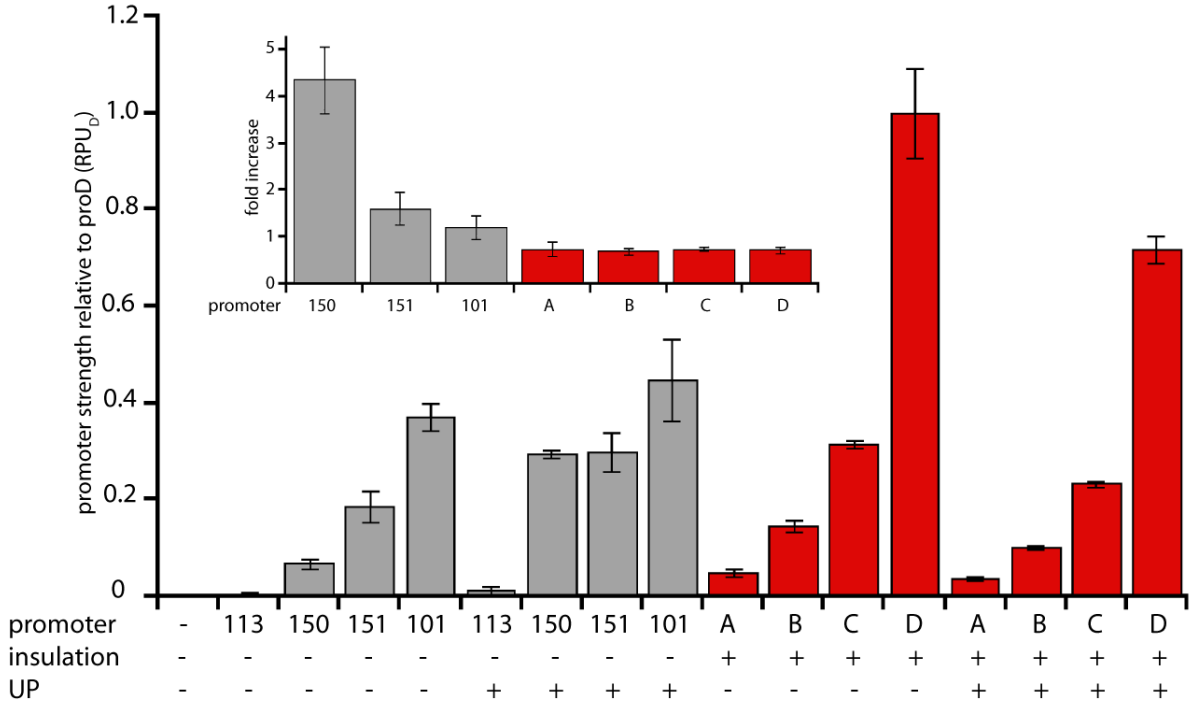

Figure 4.4. Effect of UP sequence on apparent promoter activity. The UP sequence was cloned upstream of either insulated (proA, proB, proC, proD) or uninsulated (j23113, j23150, j23151, j23101) promoters and promoter strength was measured using the GFP-reporter assay. Promoter strength was normalized to the strength of proD resulting in relative promoter units (RPU). The inset shows relative strength of each promoter with the UP sequence normalized to the strength of the parental promoter. A value of 1 indicates no change in promoter strength. The promoter j23113 was excluded from comparative analysis due to its weak promoter strength and relatively large colony-to-colony variation.

Next, we tested downstream insulation by inserting an anti sequence (ATCCGGAATCCTCTGGATCCTC) at the 3´ boundaries of insulated and uninsulated promoters (Figure 4.5). This portable sequence decreases the rate of promoter escape when present at positions +1 to +22 of many transcripts (Chan and Gross, 2001). For each set of promoters, this sequence was inserted using the available restriction sites downstream of the promoter element. This strategy resulted in the same scar between the promoter and the downstream sequence that would be present if the promoter were used to drive the production of a new transcript (see *Materials and Methods* for a description of scar sequences). The position of

the “anti” sequence was +47 to +69 for the insulated promoters and +7 to +31 for the uninsulated promoters. Again, relative GFP synthesis rates were measured as a surrogate for promoter activity. As shown in Figure 4.5, insertion of the anti sequence had almost no effect on the insulated promoters and had variable effects on the uninsulated promoters. The strongest promoter, j23101 was downregulated ~2-fold whereas the weaker promoter, j23150, showed no change. Interpreting activity differences between promoters with and without the anti sequence is difficult because the insertions alter the mRNA and thus could affect mRNA stability and translation efficiency in addition to promoter activity. We note, however, that for each set of insulated or uninsulated promoters, the mRNA transcribed is independent of the particular promoter assayed and thus is constant within that set. When normalized for relative promoter activity as shown in Figure 4.5, it is clear that the insulated promoters are resistant to the effects of inserting this sequence.

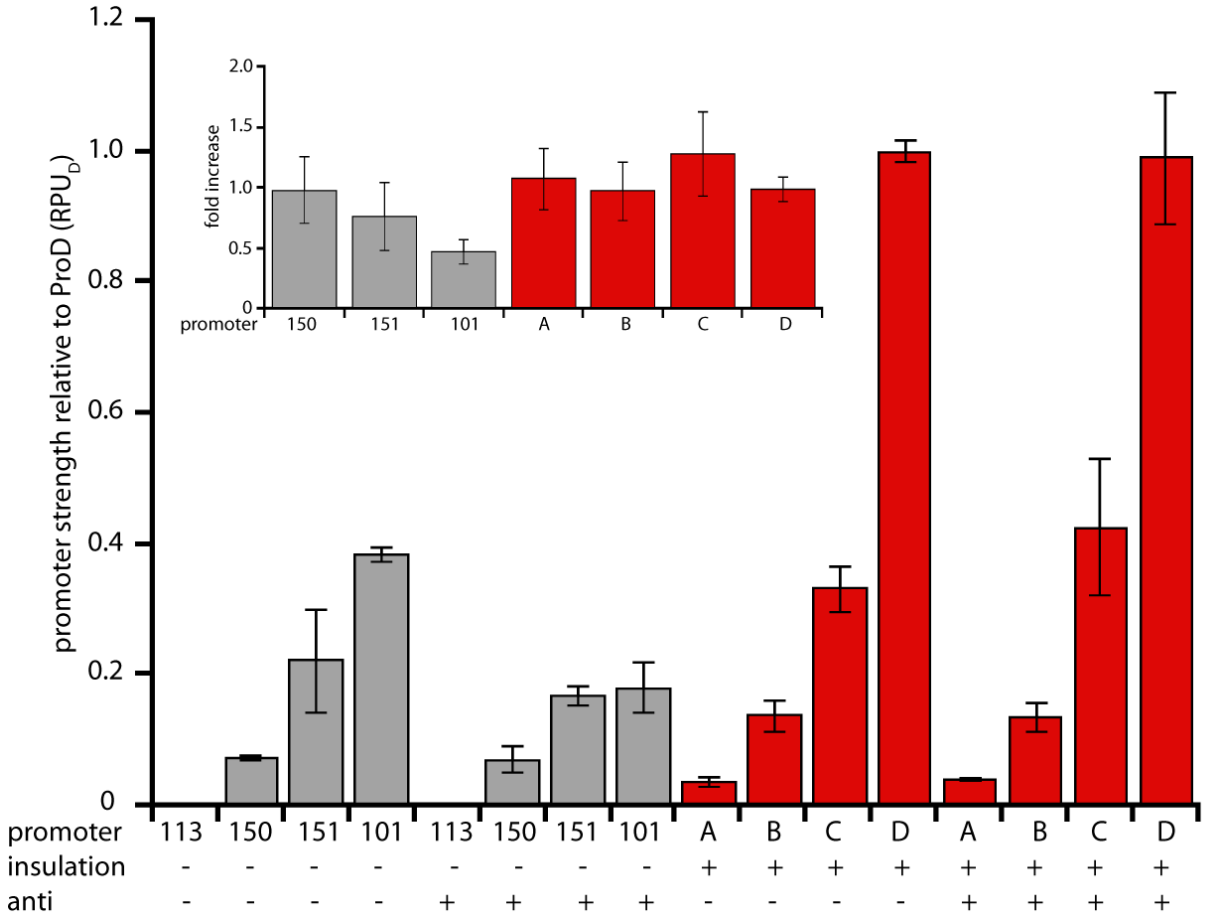

Figure 4.5. Effect of sequences inserted 3´ of the promoter. The "anti" sequence was cloned downstream of either insulated or uninsulated promoters and the promoter strength of each construct was measured using the GFP-reporter assay. For each promoter that could be robustly measured, the inset reports the ratio of promoter strength with the anti sequence to that without the insertion.

Promoter activity when driving disparate open reading frames

Is promoter strength predictable following fusion to different open-reading frames? To address this question, we fused the insulated and uninsulated promoters to sequences encoding two additional fluorescent reporters. The first was Gemini, which contains an N-terminal LacZα sequence and a C-terminal GFP domain (Martin *et al.*, 2009). The second was dsRed, which is an engineered fluorescent protein with little homology to GFP (Campbell *et al.*, 2002). Although it is difficult to directly compare expression levels of GFP with those of dsRed or Gemini in a meaningful way, the relative strengths of different promoters driving expression of each type of

protein should be predictable if there is sufficient insulation from the effects of the initially transcribed sequence.

We first measured the synthesis rate of Gemini for each promoter (proA, proB, proC, proD, j23113, j23150, j23151, j23101) in a manner similar to that described for GFP. Gemini, could be readily assayed using fluorescence from its GFP domain. Importantly, the N-terminus of Gemini (LacZα) is different from that of GFP, and thus one might expect differences in transcription from uninsulated promoters. As reported previously, we observed a decrease in the absolute synthesis rates of Gemini relative to those of GFP (Martin *et al.*, 2009). For each promoter, we calculated the relative promoter strength compared to proD driving the same open reading frame using to equation 4.2. Figure 4.6 plots the relative promoter strength (RPU_{proD}) determined using GFP (x-axis) against that determined using Gemini (y-axis). Promoters not altered by the introduction of Gemini, approximately fall on the line y=x (red). Interestingly, the insulated promoters show a ~1:1 relationship between relative activity measured either by GFP or by Gemini. By contrast, the stronger uninsulated promoters (j23151, j23101) show diminished apparent relative activity when driving production of Gemini (Figure 4.6, right). We do not believe that the decreased activity of the uninsulated promoters occurs as a consequence of translation or Gemini folding becoming rate limiting, because the absolute activity of even the strongest uninsulated promoter is weaker than that of the strongest insulated promoter (proD). The fact that neither proC nor proD show diminished activity when driving Gemini production argues that transcription from the strongest uninsulated promoters becomes limiting in this assay. Again, the uninsulated promoters show greater context-dependent activity.

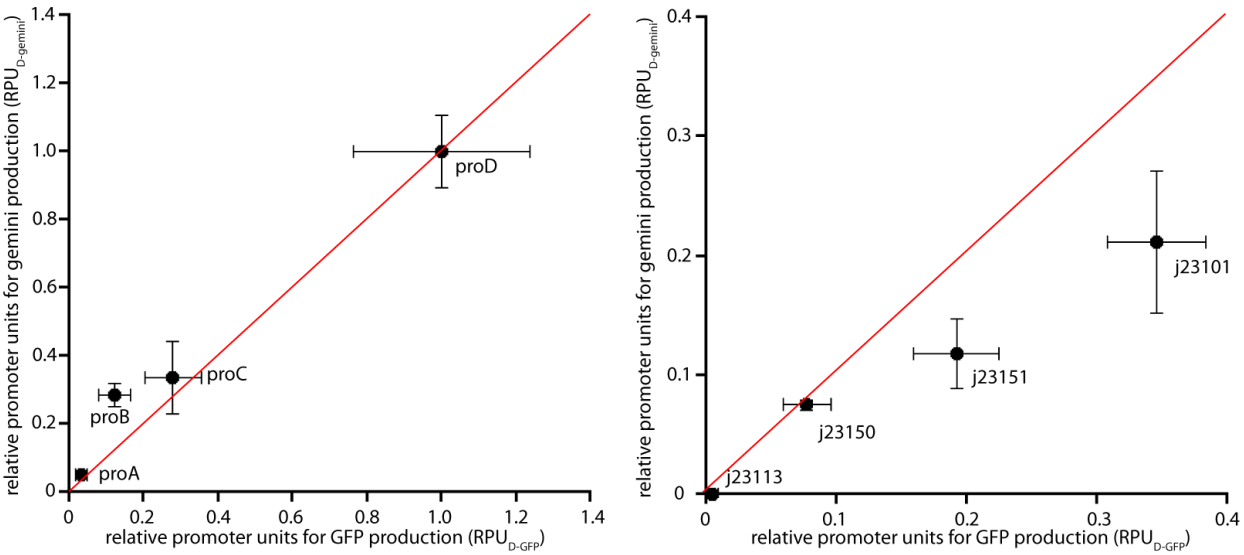

Figure 4.6. Aparent promoter activities driving production of GFP versus Gemini. Promoter strength was measured for a set of insulated (left) or uninsulated (right) promoters driving the production of GFP (x-axis) or Gemini (y-axis). For each open reading frame, aparent synthesis rates were determined by measuring florescence as shown in Equation 4.1. To allow comparison between open reading frames, each synthesis rate was normalized to that of proD driving production of the same open reading frame (Equation 4.2).

We next measured the synthesis rate of dsRed for proA, proB, proC, proD, j23113, j23150, j23151, and j23101 promoters in a manner similar to that described for Gemini (see *Materials and Methods* for details). For each promoter driving production of each open reading frame, we calculated promoter strength relative to j23101 driving production of the same open reading frame. As described above, this analysis allows for comparison between promoter strength determined using either GFP or dsRED. For promoters with weak activity, we observed a 1:1 correspondence between apparent promoter strength measured using GFP and dsRed. For the strongest promoter tested, proD, we saw decreased relative activity when driving production of dsRed (Figure 4.7). As described previously, this could result from altered transcription indicating the insulation is insufficient or could arise because a process other than transcription becomes limiting for this elevated level of dsRed production. Interestingly, for all of the

uninsulated promoters tested, the relative promoter strength was maintained when driving production of dsRed.

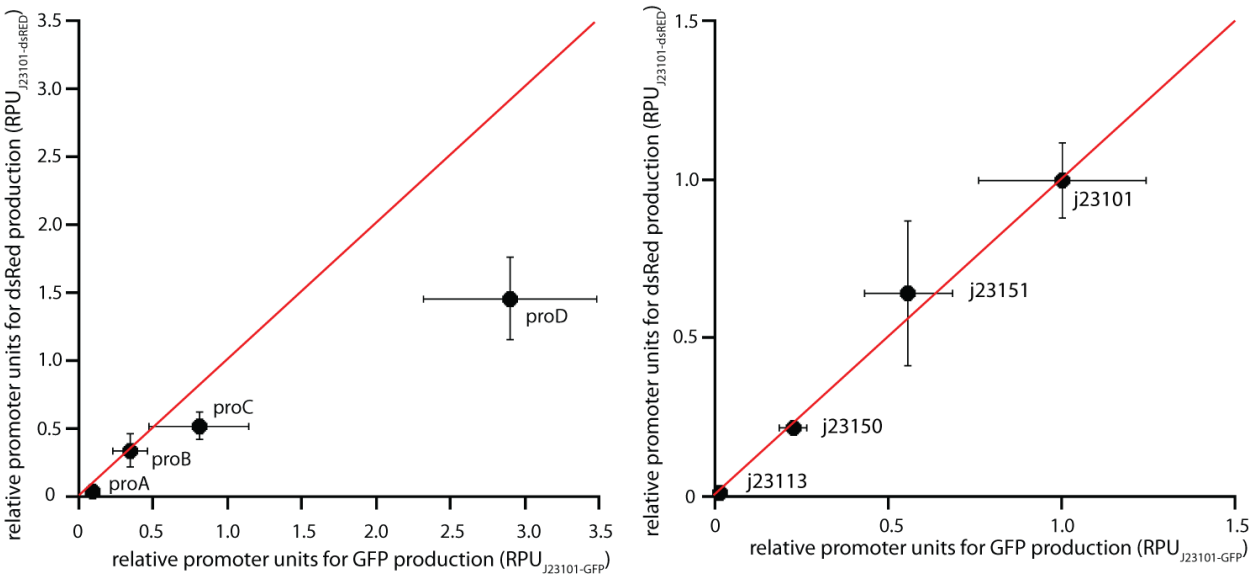

Figure 4.7. Aparent promoter activity when driving production of the protein dsRed. Promoter strength was measured for a set of either insulated (left) or uninsulated (right) promoters each driving the production of either GFP (x-axis) or dsRed (y-axis). For each open reading frame, aparent synthesis rates were determined by measuring florescence as shown in Equation 4.1. To aid in comparison between open reading frames, each synthesis rate was normalized to that of j23101 driving production of the same open reading frame.

Promoter activity from a chromosomal locus

In the experiments described so far, the promoters were carried on the medium copy number plasmid, pSB3C5, which bears a p15A origin of replication (Shetty *et al.*, 2008). To further investigate the efficacy and predictability of these promoters, we moved the insulated proA, proB, proC, proD promoters, the uninsulated j23113, j23150, j23151, j23101 promoters, and a nonfluorescent control cassette to a chromosomal locus. The promoter-RBS-GFP-terminator construct was fused to a kanamycin resistance marker using PCR and the entire cassette was site-

specifically recombined at the chromosomal *tonB* locus using recombineering techniques (see *Materials and Methods*) (Datta *et al.*, 2006).

Promoter activities from the *tonB* locus were measured using the fluorescent signal from the encoded GFP. Figure 4.8A shows the synthesis rates for each promoter. As expected, the absolute activity of each promoter was decreased when placed on the chromosome, presumably a result of decreased copy number. To allow for comparison of promoter strength between these two loci, the synthesis rate was converted to relative promoter units by normalization to the synthesis rate of proD (Equation 4.2). As described previously, this normalization masks the effects of copy number (Kelly *et al.*, 2009). For each promoter, Figure 4.8B shows the relative promoter strength measured from the plasmid vs. the relative promoter strength measured from the chromosome. Interestingly, both sets of promoters exhibited an approximately 1:1 correspondence in relative promoter activity at the chromosomal versus plasmid loci.

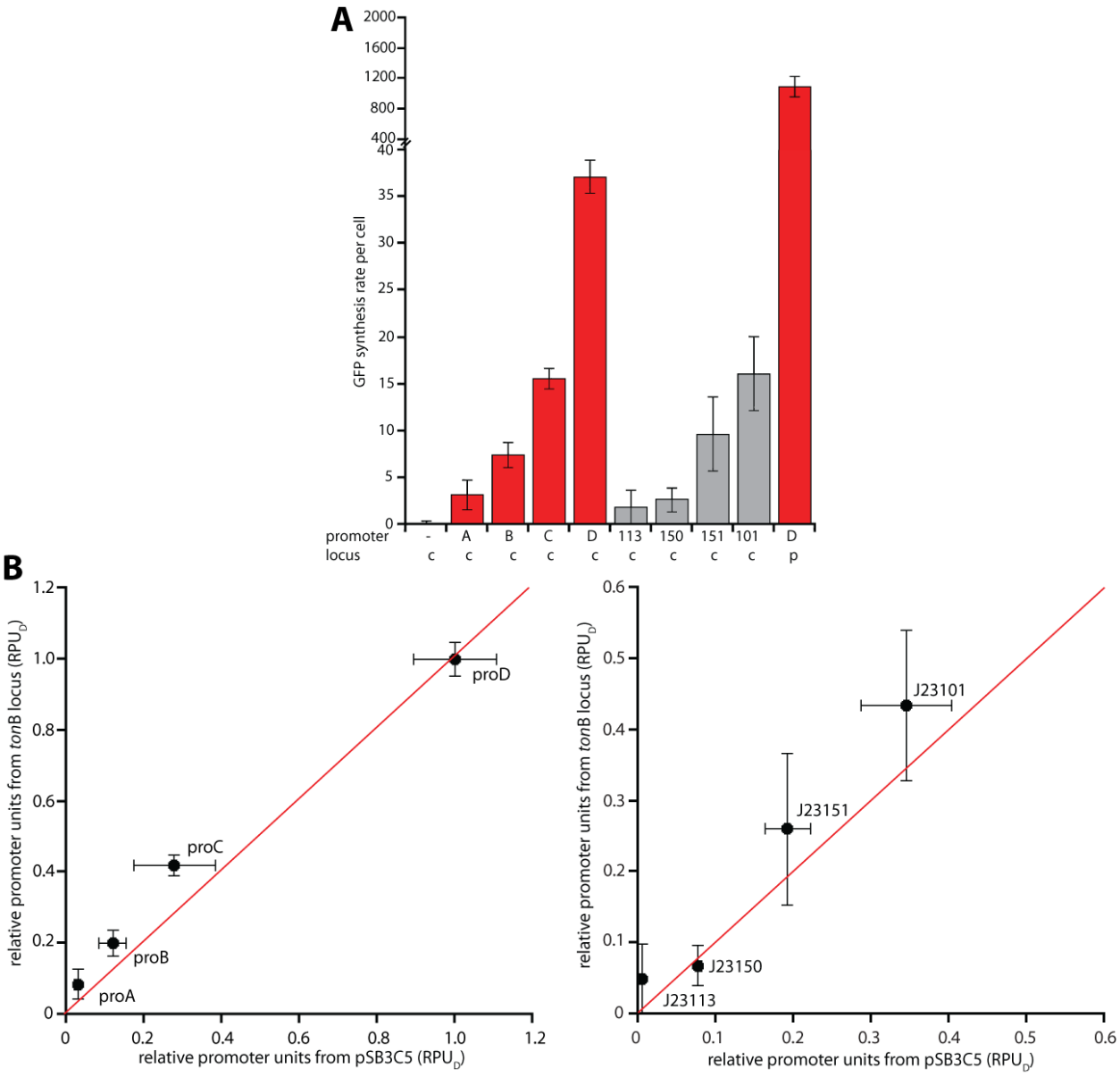

Figure 4.8. Promoter activity from the chromosomal *tonB* locus. Promoters driving the expression of GFP were recombined onto the chromosome downstream of the *tonB* locus and the activity of the promoter was measured using fluorescence from the GFP gene. **(A)** GFP synthesis rate per cell from the *ton*B locus (c) or from the plasmid pSB3C5 (p). **(B)** Plots correlating the relative promoter activity either from the plasmid, pSB3C5 or the chromosome (at the *tonB* locus) for the insulated promoters (left) or the uninsulted promoters (right).

Discussion

We have generated a library of constitutive bacterial promoters whose activity spans two orders of magnitude. These promoters contain sequences extending beyond the core polymerase binding region in both the 5′ and 3′ direction. By testing these promoters in a variety of sequence contexts, both chromosomal and plasmid-based, we have demonstrated that their activity is highly predictable with minimal effects from the surround genetic context. As such, the promoter's activity measured in one context should be predictive of their function in a new context. Such functional composition should facilitate the engineering of biological systems.

Our promoters are 160 bp in length, substantially larger than the minimal 51 bp promoters we have used for comparison. Although our promoters were less affected by the stimulatory and repressive effects of many sequence elements, it should be noted that some long-range regulation by transcription-factors may still affect these promoters. Given that such regulation is probably unavoidable with any amount of insulation, we compromised on a promoter size that shows improved context-independent behavior but remained small enough to be easily incorporated into genetic pathways.

To decrease the possibility of the initially transcribed sequence altering promoter activity, our promoters include insulation extending beyond the transcriptional start site. This design results in several desirable features. First, unlike a minimal promoter, the transcription initiation site is clearly defined and invariant for the library irrespective of the downstream gene. Second, the defined 5′ untranslated sequence may facilitate the measurement and prediction of mRNA degradation rates for transcripts generated by these promoters. Lastly, because some prediction

algorithms for the strength of ribosome binding sites make use of surrounding sequence information, the invariant 5′ termini could improve the prediction of translational initiation rates (Salis *et al.*, 2009).

Although the 5′ termini encoded by our promoters has no effect on the resulting protein product, it will prohibit some applications. For example, natural and synthetic riboregulators located in the 5′ untranslated region allow for small-molecule control of translation (Winkler *et al.*, 2002; Suess *et al.*, 2004). The initially transcribed sequence generated by our promoters may be incompatible with such devices. For most applications, however, this 5′ sequence should pose no problem, as it affects neither the ribosome binding site nor the resultant protein product.

We find that our insulated promoters are not perturbed by the introduction of stimulatory UP sequences 5′ of the promoter, or repressive "anti" sequences 3′ of the promoter. The UP sequence likely has no effect on these promoters because the extended insulation sequence precludes the polymerase from simultaneously forming favorable contacts with the core promoter and the UP sequence. By contrast, we observe strong activation of the weak minimal promoters by the UP sequence, suggesting that this sequence improves polymerase recruitment. For the strongest minimal promoter, the UP element had no effect and thus some process after polymerase recruitment is probably rate limiting. Such non-uniform effects for the minimal promoters limits their predictably in different sequence contexts.

The anti sequence also affects the minimal promoters non-uniformly, down regulating the strongest promoter 2-fold but not affecting other promoters. For these unaffected promoters, a

process other than promoter escape (*e.g.* polymerase binding) is likely to be rate limiting. Interestingly, none of the insulated promoters are affected by the anti sequence, which causes transcriptional repression in a σ^{70}-dependent fashion (Chan and Gross, 2001). It seems likely that the insulated promoters are unaffected, because the σ subunit dissociates from the core polymerase before encountering the anti sequence.

In one instance, we found that the insulated promoters predictably drive production of a different gene product (Gemini), whereas the minimal promoters did not. Again, it appears that for the strongest uninsulated promoters, the initially transcribed sequence strongly affects promoter activity. The fact that even the strongest insulated promoter was not affected implies that transcription is rate limiting for production of GFP and Gemini. However, for the strongest insulated promoter driving dsRed production, we observed decreased apparent promoter activity. In this instance, translation or chaperone-dependent folding may have become limiting for protein synthesis. In such cases, one can no longer assume that protein production will depend linearly on promoter activity and instead, a more complicated transfer function to correlate promoter activity to protein production will have to be determined. When a process other than transcription becomes rate-limiting for production, no amount of insulation will mitigate non-linear effects.

Achieving proper steady-state protein levels can be accomplished using libraries of constitutive promoters as described here, or by using promoters whose activity can be titrated using a small-molecule inducer (Lee and Keasling, 2005; Canton *et al.*, 2008). Inducible systems are particularly appealing for applications in which only one gene product must be regulated and for

those systems in which it is advantageous to regulate gene activity dynamically (*e.g.* induction and repression of essential or toxic genes). The use of insulation sequences to decrease context-dependent promoter activity could easily be extended to regulated promoters.

Materials and Methods

Promoter sequences

The sequences for the promoters used in this work are listed below. The scar sequences generated by standard BioBrick assemblies are in lowercase font, and the expected start site is underlined. For each promoter set (insulted or minimal), only the -35 and -10 hexamers (shown in large, bold font) vary between library members (Table 4.1).

proD (insulated promoter)
ttctagagCACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTGGTTGC
TGGATAAC**TTTACG**GGCATGCATAAGGCTCG**TATAAT**ATATTC**A**GGGAGACCACAACGGTTTCCCTC
TACAAATAATTTTGTTTAACTTTtactagag

j23101 (minimal promoter)
ttctagag**TTTACA**GCTAGCTCAGTCCTAGG**TATAAT**GCTAGCt**a**ctagag

Table 4.1. Promoter Sequences

Promoter	-35 hexamer	-10 hexamer	RPU_D
proA	tttacg	taggct	0.030
proB	tttacg	taatat	0.119
proC	tttacg	tatgat	0.278
proD	tttacg	tataat	1.000
pro1	tttacg	gtatct	0.009
pro2	gcggtg	tataat	0.017
pro3	tttacg	gaggat	0.017
pro4	tttacg	gatgat	0.033
pro5	tttacg	taggat	0.050
pro6	tttacg	taaaat	0.193
j23113	ctgatg	gattat	0.005
j23150	tttacg	tattat	0.077
j23151	ttgatg	acaatg	0.192
j23101	tttaca	tattat	0.345

Plasmids and strains

Experiments with plasmid-borne reporters were performed in *E. coli* strain DH5α (F^-, λ^-, φ80*lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *deo*R, *rec*A1, *end*A1, *hsd*R17(rk^-, mk^+), *pho*A, *sup*E44, *thi*-1, *gyr*A96, *rel*A1), whereas experiments with chromosomally encoded reporters were performed in *E. coli* strain W3110 (F^-, λ^-, IN(*rrn*D-*rrn*E), *rph*-1). With the exception of assays for promoter function on the chromosome, all experiments utilized plasmid pSB3C5 and the construct of interest was cloned via standard procedures between the BioBrick cloning sites (Shetty *et al.*, 2008). Each construct contained the Bba_B0032 ribosome binding site and the Bba_B0015 transcriptional terminator. The sequences for the fluorescent reporter proteins, GFP (Bba_E0040), dsRed (Bba_E1010), Gemini (Bba_E0051), the ribosome binding site (Bba_B0032) and the terminator (Bba_B0015) can be found at the Registry of Standard Biological Parts (www.partsregistry.org).

Plasmid constructs contained the following elements: promoter-TACTAGAG-B0032-TACTAG-ORF(dsRed, GFP, Gemini)-TACTAGAG-B0015, where the ORF was exchanged using standard PCR-based techniques. For chromosomal insertions, test constructs were fused to a kanamycin resistance marker (Bba_P1003, Registry of Standard Biological Parts, www.partsregistry.org) using SOEing PCR (Horton *et al.*, 1990). PCR products were recombined onto the chromosome using the λ-red recombination system, encoded on the plasmid vector pSIM5, as described previously (Datta *et al.*, 2006). After verification of successful cassette insertion by sequencing, pSIM5 was cured from the strain. For the *tonB* locus, the SOEing primers used are as follows (with homology to the locus listed in bold).

tonB-BioBrickPrefix-fwd
AAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTgaattcgcggccgcttctag

BioBrickSuffix-rev
cgaacttttgctgagttgaaggatcagCTGCAGCGGCCGCTACTAGTA

BioBrickSuffix-fwd
TACTAGTAGCGGCCGCTGCAGctgatccttcaactcagcaaaagttcg

P1003-*tonB*-rev
GATCCTGAAGGAAAACCTCGCGCCTTACCTGTTGAGTAATttattagaaaaactcatcga

The UP sequence (GAGAAAATTATTTTAAATTTCCTC) was introduced upstream of the promoter constructs using standard techniques resulting in a BioBrick scar (ACTAGA) between the UP sequence and the promoter. The anti sequence (ATCCGGAATCCTCTGGATCCTC) was introduced in a similar fashion resulting in constructs of the form: promoter-TACTAGAG-**anti**-B0032-TACTAG-GFP-TACTAGAG-B0015.

Two strains were used to control for cellular auto-fluorescence. DH5α transformed with pSB3C5 was used as a negative control for experiments using plasmid-based constructs. For experiments testing promoter function from the *tonB* locus, the kanamycin resistance marker with no reporter construct was recombined downstream of the *tonB* locus using primers *tonB*-BioBrickPrefix-fwd, P1003-*tonB*-rev.

Promoter activity assays

All GFP, dsRed, and Gemini based assays were performed and analyzed as described (Kelly *et al.*, 2009) with the minor modifications listed below. Individual test colonies along with a negative control strain were picked in triplicate and grown overnight in 5 cm culture tubes in LB broth supplemented with 35 μg/mL chloramphenicol (for experiments using pSB3C5) or 10

μg/mL kanamycin (for experiments using chromosomal insertions). Cultures were then diluted 100-fold into M9 media (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% glycerol) supplemented with appropriate antibiotics and grown at 37 °C for four hours. Cultures were aliquoted (150 μL) into a 96-well plate (Greiner Bio-One) in which OD (600 nm) and fluorescence (GFP, Gemini: excitation 467 nm, emission 511 nm; dsRed: excitation 560 nm, emission 590 nm) were read using a SpectraMax M5 fluorescence plate reader (Molecular Devices). Cultures were allowed to continue growing in tubes for an additional 1.25 hours at 37 °C at which time OD_{600} and fluorescence were read again. For each sample, the change in fluorescence signal between the two readings was divided by the average OD_{600}. This measure of promoter activity (per cell synthesis rate) was corrected for background auto-fluorescence by subtracting the per cell synthesis rate of the negative control. The corrected synthesis rate was then normalized to the average synthesis rate of the reference promoter, proD, resulting in relative promoter units (RPU_D) (Kelly *et al.*, 2009). The average synthesis rate of the minimal promoter, j23101, was used to normalize promoter activity in experiments using dsRed as proD displayed diminished activity with this open reading frame. Errors bars shown in all figures represent the standard deviation of triplicate measurements.

Acknowledgements

We thank Jason Kelly, Drew Endy, and the MIT Synthetic Biology Working Group for helpful discussion and materials.

References

1. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. 2005. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**(36): 12678-12683.

2. Anderson, J. C. (2009) in *Registry of Standard Biological Parts*.

3. Anthony, J. R., Anthony, L. C., Nowroozi, F., Kwon, G., Newman, J. D., and Keasling, J. D. 2009. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab Eng* **11**(1): 13-19.

4. Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H., and Ikehara, M. 1983. Essential structure of E. coli promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res* **11**(17): 5855-5864.

5. Atsumi, S., and Liao, J. C. 2008. Metabolic engineering for advanced biofuels production from Escherichia coli. *Curr Opin Biotechnol* **19**(5): 414-419.

6. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature* **434**(7037): 1130-1134.

7. Browning, D. F., and Busby, S. J. 2004. The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**(1): 57-65.

8. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. 2002. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **99**(12): 7877-7882.

9. Canton, B., Labno, A., and Endy, D. 2008. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26**(7): 787-793.

10. Chan, C. L., and Gross, C. A. 2001. The anti-initial transcribed sequence, a portable sequence that impedes promoter escape, requires sigma70 for function. *J Biol Chem* **276**(41): 38201-38209.

11. Datta, S., Costantino, N., and Court, D. L. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**109-115.

12. de Crombrugghe, B., Busby, S., and Buc, H. 1984. Cyclic AMP receptor protein: role in transcription activation. *Science* **224**(4651): 831-838.

13. deHaseth, P. L., Zupancic, M. L., and Record, M. T., Jr. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol* **180**(12): 3019-3025.

14. Ebright, R. H. 1993. Transcription activation at Class I CAP-dependent promoters. *Mol Microbiol* **8**(5): 797-802.

15. Elowitz, M. B., and Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767): 335-338.

16. Endy, D. 2005. Foundations for engineering biology. *Nature* **438**(7067): 449-453.

17. Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. 1998. Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A* **95**(17): 9761-9766.

18. Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W., Niu, W., Ebright, R. H., and Gourse, R. L. 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* **13**(16): 2134-2147.

19. Hammer, K., Mijakovic, I., and Jensen, P. R. 2006. Synthetic promoter libraries--tuning of gene expression. *Trends Biotechnol* **24**(2): 53-55.

20. Horton, R. M., Cai, Z. L., Ho, S. N., and Pease, L. R. 1990. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **8**(5): 528-535.

21. Hsu, L. M. (2008). *Chapter 4.5.2.2, Promoter Escape by Escherichia coli RNA Polymerase. Escherichia coli and Salmonella.* ASM Press, Washington, DC.

22. Jensen, P. R., and Hammer, K. 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64**(1): 82-87.

23. Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* **3**4.

24. Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M., and Ebright, R. H. 2004. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol* **14**(1): 10-20.

25. Lee, S. K., and Keasling, J. D. 2005. A propionate-inducible expression system for enteric bacteria. *Appl Environ Microbiol* **71**(11): 6856-6862.

26. Lutke-Eversloh, T., and Stephanopoulos, G. 2008. Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: identification of enzymatic bottlenecks by systematic gene overexpression. *Metab Eng* **10**(2): 69-77.

27. Martin, L., Che, A., and Endy, D. 2009. Gemini, a bifunctional enzymatic and fluorescent reporter of gene expression. *PLoS One* **4**(11): e7569.

28. Mencia, M., Monsalve, M., Rojo, F., and Salas, M. 1996. Transcription activation by phage phi29 protein p4 is mediated by interaction with the alpha subunit of Bacillus subtilis RNA polymerase. *Proc Natl Acad Sci U S A* **93**(13): 6616-6620.

29. Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., and Morett, E. 2009. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PLoS One* **4**(10): e7526.

30. Meng, W., Belyaeva, T., Savery, N. J., Busby, S. J., Ross, W. E., Gaal, T., Gourse, R. L., and Thomas, M. S. 2001. UP element-dependent transcription at the Escherichia coli rrnB P1 promoter: positional requirements and role of the RNA polymerase alpha subunit linker. *Nucleic Acids Res* **29**(20): 4166-4178.

31. Paul, B. J., Ross, W., Gaal, T., and Gourse, R. L. 2004. rRNA transcription in Escherichia coli. *Annu Rev Genet* **38**749-770.

32. Pemberton, I. K., Muskhelishvili, G., Travers, A. A., and Buckle, M. 2000. The G+C-rich discriminator region of the tyrT promoter antagonises the formation of stable preinitiation complexes. *J Mol Biol* **299**(4): 859-864.

33. Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**(7086): 940-943.

34. Rojo, F. 1999. Repression of transcription initiation in bacteria. *J Bacteriol* **181**(10): 2987-2991.

35. Ross, W., Ernst, A., and Gourse, R. L. 2001. Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**(5): 491-506.

36. Rossi, J. J., Soberon, X., Marumoto, Y., McMahon, J., and Itakura, K. 1983. Biological expression of an Escherichia coli consensus sequence promoter and some mutant derivatives. *Proc Natl Acad Sci U S A* **80**(11): 3203-3207.

37. Salis, H. M., Mirsky, E. A., and Voigt, C. A. 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**(10): 946-950.

38. Shetty, R. P., Endy, D., and Knight, T. F., Jr. 2008. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* **2**5.

39. Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**(7221): 516-519.

40. Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. 2004. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32**(4): 1610-1614.

41. Win, M. N., Liang, J. C., and Smolke, C. D. 2009. Frameworks for programming biological function through RNA parts and devices. *Chem Biol* **16**(3): 298-310.

42. Winkler, W., Nahvi, A., and Breaker, R. R. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**(6910): 952-956.

43. Wong, W. W., Tsai, T. Y., and Liao, J. C. 2007. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3**130.

44. Yokobayashi, Y., Weiss, R., and Arnold, F. H. 2002. Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A* **99**(26): 16587-16591.

45. You, L., Cox, R. S., 3rd, Weiss, R., and Arnold, F. H. 2004. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**(6985): 868-871.

Chapter 5

Small-molecule Control of Protein Degradation Using Split Adaptors

A modified version of this chapter will be submitted for publication with Joseph H. Davis, Tania A. Baker, and Robert T. Sauer as authors.

Abstract

Targeted degradation provides an effective method to study the function of protein products and has applications in biotechnology. One promising approach to control intracellular degradation uses adaptor proteins to specifically target substrates bearing genetically encoded degradation tags. We have developed a degradation system using an engineered split adaptor in which assembly and thus adaptor function is dependent on rapamycin, a small-molecule. This degradation system does not require modification of endogenous proteases, functions robustly over a wide range of adaptor concentrations, and does not require new synthesis of adaptor or protease molecules to induce degradation. In addition, we identify new C-terminal degradation tags that exhibit varied levels of rapamycin-dependent and rapamycin-independent degradation by the ClpXP protease. Together, these reagents should prove useful in controlling protein degradation in bacteria.

Introduction

Targeted proteolysis provides one mechanism to control intracellular protein levels in a dynamic fashion and has applications in basic science and biological engineering (Banaszynski *et al.*, 2006; McGinness *et al.*, 2006; Griffith and Grossman, 2008; Moore *et al.*, 2008; Davis *et al.*, 2009; Taxis *et al.*, 2009). Indeed, the perturbation of protein stability has proven critical in the generation of synthetic cellular circuits (Elowitz and Leibler, 2000; Wong *et al.*, 2007; Stricker *et al.*, 2008). Controlled degradation could also be applied to metabolic engineering. For example, after growing cells to a critical density, essential enzymes catalyzing off-pathway reactions could be targeted for degradation, directing metabolite flux toward the desired product more efficiently. Targeted degradation has also been utilized to investigate loss of function phenotypes (Griffith and Grossman, 2008), providing an alternative and often complementary mechanism to transcriptional knockdowns. A degradation approach is particularly valuable when pre-existing proteins are inherently long-lived, as terminating the synthesis of such molecules results in relatively slow elimination of the gene product (Guzman *et al.*, 1995; Rappleye and Roth, 1997; Fire *et al.*, 1998; Ji *et al.*, 2001; Knight and Shokat, 2007). In these cases, proteins are diluted via cell growth and division, often requiring many cell cycles to reach levels that are low enough to eliminate function. During this time period, cells may up-regulate compensatory pathways or acquire suppressor mutations, often obscuring the phenotype of the perturbation.

When combined with temporal control, targeted proteolysis can be used to investigate essential proteins under control of their native transcriptional and translational control elements. Temperature-sensitive (*ts*) alleles share many of the advantages of targeted proteolysis, but tight *ts*-alleles are difficult to isolate for many gene products. Furthermore, a temperature shift can

result in unintended changes in cellular physiology, complicating interpretation. In yeast, for example, shifting the temperature form 25 to 37 °C alters the expression of nearly 1000 genes, half of which have unknown function (Causton *et al.*, 2001). Small-molecule inhibitors/activators allow perturbations of native gene function with rapid temporal control under many environmental conditions (Stockwell, 2000). Unfortunately, the identification of highly specific, cell-permeable inhibitors is not trivial, and conclusively ruling out "off-target" effects is a substantial challenge.

In *Escherchia coli* and most other bacteria, processive intracellular proteolysis is mediated by a group of energy-dependent AAA+ proteases, including ClpXP (Baker and Sauer, 2006). The ClpXP enzyme degrades substrates via a multi-step process, which begins with the recognition of short peptide sequences (referred to as degradation tags or degrons), which are exposed in the native protein substrate (Flynn *et al.*, 2003; Sauer *et al.*, 2004). Once bound, processive cycles of ATP hydrolysis in ClpX drive translocation of the degradation tag through a narrow axial pore (Sousa *et al.*, 2000; Kenniston *et al.*, 2003; Glynn *et al.*, 2009). Successive pulling events eventually result in global substrate unfolding, allowing translocation of the denatured polypeptide into the lumen of ClpP, where it is cleaved into short peptide fragments (Joshi *et al.*, 2004; Kenniston *et al.*, 2004). ClpX can unfold proteins with a wide range of thermodynamic stabilities and appears to have little sequence specificity in terms of substrate translocation (Kenniston *et al.*, 2005; Barkow *et al.*, 2009). Thus, substrate selectivity is determined by the efficiency of the initial binding event, which can be modulated by accessory factors called adaptors. For example, the SspB adaptor improves ClpXP degradation of ssrA-tagged substrates by binding to the protease and to a portion of the ssrA tag (AANDENYALAA), thereby

increasing the effective concentration of the substrate relative to the enzyme (Levchenko *et al.*, 2000; Wah *et al.*, 2003; Bolon *et al.*, 2004a; McGinness *et al.*, 2007). Indeed, using synthetic degradation components, we found that tethering alone is sufficient for efficient substrate delivery by SspB (Davis *et al.*, 2009).

Adaptor proteins have proven useful in engineering controlled degradation systems. For example, McGinness *et al.* (2006) showed that ClpXP degradation became almost entirely dependent on SspB when the C-terminal residues of the ssrA tag were mutated from LAA to DAS. Importantly, model substrates bearing DAS tags were selectively degraded in cells when transcription of the adaptor from a *lac* promotor was induced using IPTG (McGinness *et al.*, 2006; Wilson *et al.*, 2007; Griffith and Grossman, 2008). Combining a genetically encoded degradation tag and small-molecule inducer has many of the advantages of classical genetics and pharmacology and can be applied to almost any protein target with temporal control provided by the presence or absence of the small molecule.

In the work reported here, we have engineered and characterized a new targeted degradation system in which the assembly and thus activity of a split adaptor is controlled by the small-molecule rapamycin. Experiments with purified components *in vitro* and with model substrates in *E. coli* show that ClpXP degradation of appropriately tagged proteins can be efficiently controlled in a rapamycin-dependent manner without the need for new protein synthesis. This system is simple, generally applicable, and requires few genomic modifications. Moreover, it should be relatively straightforward to port to other ClpXP-containing bacteria such as *Caulobacter crescentus* and *Bacillus subtilis* (Chien *et al.*, 2007b; Griffith and Grossman, 2008).

Results

Rapamycin-dependent control of adaptor function *in vitro*

For initial studies, we used wild-type ClpXP protease, substrates bearing a DAS+4 ssrA tag (AANDENYSENYADAS; McGinness *et al.*, 2006), and split the functionally important portions of SspB into two components. The first adaptor part consisted of the SspB core domain ($SspB^{CORE}$), which binds the ssrA tag, fused to an FRB domain (Figure 5.1, $SspB^{CORE}$-FRB). The second part contained the FKBP12 protein fused to the SspB ClpX-binding tail (Figure 5.1 FKBP12-$SspB^{XB}$). Addition of rapamycin mediates dimerization of FRB and FKB12 (Chen *et al.*, 1995). The FKBP12-$SspB^{XB}$ and $SspB^{CORE}$-FRB constructs exist as physically separate and therefore nonfunctional components in absence of rapamycin but should assemble into a potentially functional adaptor in the presence of rapamycin (Figure 5.1).

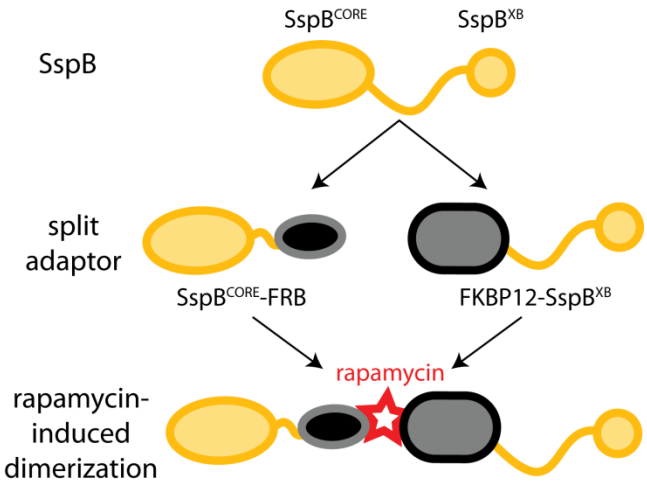

Figure 5.1. Schematic of the split adaptor. The SspB core domain (residues 1-113) was split from the ClpX-binding tail ($SspB^{XB}$, residues 139-156). $SspB^{CORE}$ was fused to FRB and FKBP12 was fused to $SspB^{XB}$. Addition of rapamycin reconstitutes a complete adaptor.

To test for rapamycin-dependent degradation, we incubated 0.3 μM ClpX, 0.9 μM ClpP, 5 μM FKBP12-SspB^{XB}, 5 μM SspB^{CORE}-FRB and an ATP regenerating system at 30 °C and initiated the reaction by addition of 2 μM GFP-DAS+4 substrate in the presence or absence of rapamycin (8 μM). With rapamycin present, GFP-DAS+4 was degraded at an initial rate of 0.58 min^{-1} enz^{-1} (Figure 5.2A). No degradation was detected in the absence of rapamycin. To assay assembly kinetics of the degradation complex, we preincubated substrate, ClpX, ClpP, and the adapter components and monitored the degradation of GFP-DAS+4 before and after addition of rapamycin (Figure 5.2B). A steady-state rate of degradation was reached within the dead-time of the experiment (~20 s), demonstrating that formation of the multi-component degradation complex is rapid.

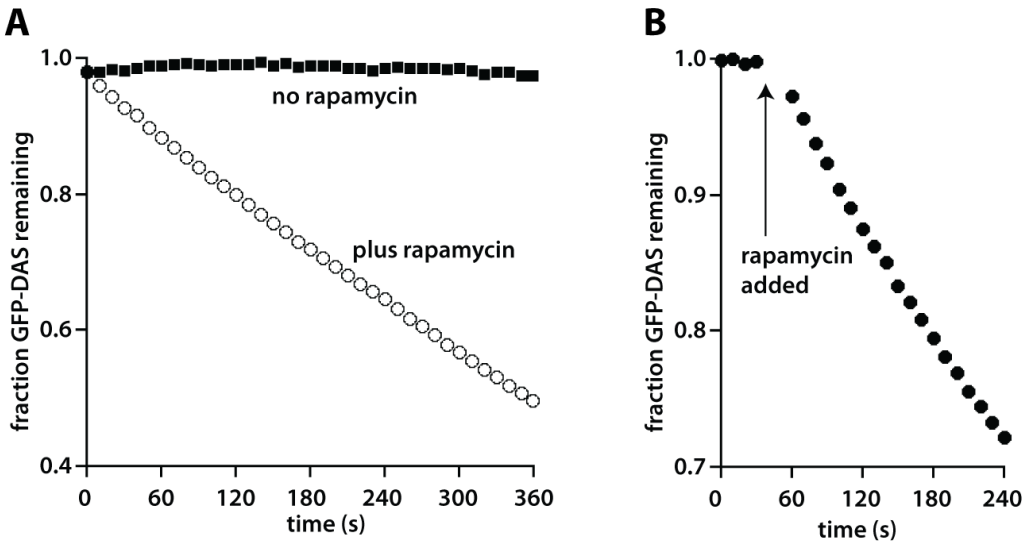

Figure 5.2. Rapamycin-dependent degradation. (A) Incubation of GFP-DAS+4 (2 μM), ClpX (0.3 μM), ClpP (0.9 μM), FKBP12-SspB^{XB} (5 μM), SspB^{CORE}-FRB (5 μM), and rapamcyin (8 μM) resulted in degradation at a rate of 0.58 min^{-1} enzyme^{-1}. No degradation was observed in the absence of rapamycin. **(B)** Addition of rapamycin resulted in rapid assembly of the degradation complex. GFP-DAS+4 (2 μM), ClpX^{WT} (0.3 μM), ClpP (0.9 μM), FKBP12-SspB^{XB} (5 μM), SspB^{CORE}-FRB (5 μM) were preincubated at 30 °C. Rapamycin (8 μM) was added at the time indicated by the arrow.

Next, we varied the concentration of each adaptor component and assayed degradation. First, we titrated equal quantities of the two adaptor components against fixed concentrations of protease and substrate. In this experiment, degradation increased in a hyperbolic fashion with an apparent binding constant of 0.51 μM (Figure 5.3A), consistent with the previously reported affinity between SspB and ClpX (Bolon *et al.*, 2004b). We also determined the efficiency of substrate degradation in experiments in which the protease, substrate, and $SspB^{CORE}$-FRB (5 μM) concentrations were fixed and the concentration of FKBP12-$SspB^{XB}$ was varied (Figure 5.3B). Again, degradation increased in a roughly hyperbolic fashion. By contrast, when the concentration of FKBP12-$SspB^{XB}$ was fixed and $SspB^{CORE}$-FRB was varied, degradation increased initially and then was inhibited at high concentrations of $SspB^{CORE}$-FRB (Figure 5.3C). This inhibition probably results from substrate competition between free $SspB^{CORE}$-FRB and the limited number of active delivery complexes ($SspB^{CORE}$-FRB•rapamycin•FKBP12-$SspB^{XB}$).

Finally, we measured the rapamycin and substrate dependence of ClpXP degradation in the presence of 5 μM FKBP12-$SspB^{XB}$ and 5 μM $SspB^{CORE}$-FRB (Figure 5.3D). Rapamycin strongly enhanced degradation, even at high substrate concentrations. Indeed, in the presence of rapamycin and the split-adaptor components, K_M (1.3 μM) and V_{MAX} (0.92 min^{-1} enz^{-1}) for GFP-DAS+4 degradation demonstrated that the split adaptor delivers substrates to ClpXP efficiently, even at relatively low substrate concentrations. Taken together, these results indicate that most enzyme and adaptor concentration regimes result in robust rapamycin-dependent degradation.

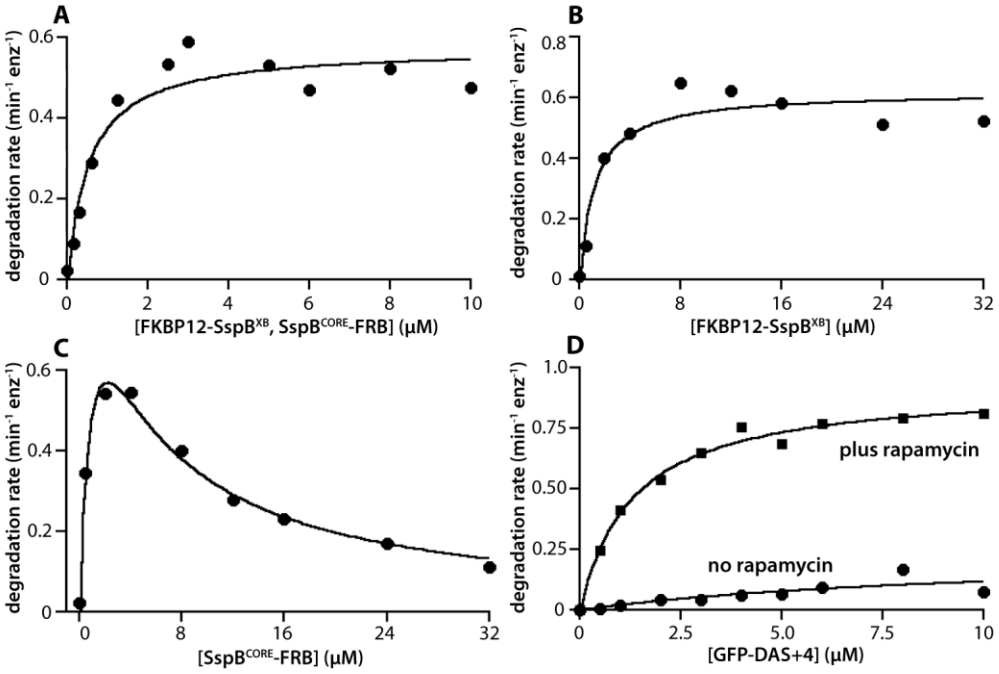

Figure 5.3. Dependence of GFP-DAS+4 degradation on adaptor or substrate concentration. **(A)** Degradation in the presence of increasing equimolar concentrations of FKBP12-SspB^{XB} and SspB^{CORE}-FRB. The solid line is a hyperbolic fit with an apparent binding constant of 0.51 μM. **(B)** Degradation in the presence of a fixed concentration of SspB^{CORE}-FRB (5 μM) and increasing FKBP12-SspB^{XB}. The solid line is a hyperbolic fit. **(C)** Degradation with a fixed concentration of FKBP12-SspB^{XB} (5 μM) and increasing concentrations of SspB^{CORE}-FRB. Data were fit to the equation, $rate = \frac{k_1[SspB^{CORE}-FRB]}{(K_a+[SspB^{CORE}-FRB]+\frac{[SspB^{CORE}-FRB]^2}{K_b}}$ **(D)** Substrate dependence of degradation in the presence of SspB^{CORE}-FRB (5 μM) and FKBP12-SspB^{XB} (5 μM). Fitting the rapamycin data to the Michaelis-Menten equation gave K_M = 1.3 μM, V_{MAX} = 0.92 min^{-1} enz^{-1}. In all panels, the concentrations of $ClpX_6$ and $ClpP_{14}$ were 0.3 and 0.9 μM, respectively. In panels **A**-**C** and the "plus rapamycin" experiment of panel **D**, the rapamcyin concentration was 12 μM. In panels **A**-**C**, the GFP-DAS+4 concentration was 2 μM.

Controlled degradation of a transcriptional repressor

Given the promising results *in vitro*, we developed an assay for rapamycin-dependent degradation *in vivo*. To prevent uncontrolled substrate delivery and degradation, the *ssp*B gene was deleted from the chromosome of *E. coli* strain W3110. Next, we introduced a DAS+4 tag at the C-terminus of the *lac*I transcriptional repressor, using a scarless λ-red based recombination technique (see *Materials and Methods*). This strain was transformed with a plasmid (pJD427)

bearing constitutive promoters driving the production of FKBP12-SspB^{XB} and SspB^{CORE}-FRB. We expected LacI-DAS+4 to repress *lac*Z transcription and thus production of β-galactasidase in the absence of rapamycin (Figure 5.4A), with addition of rapamcycin resulting in LacI-DAS+4 degradation and thus *lac*Z induction, which could be measured by RT-qPCR or by Miller assays for β-galactasidase activity (Miller, 1972; VanGuilder *et al.*, 2008; Clarkson *et al.*, 2010).

LacI-DAS+4 exhibited rapamycin-dependent degradation *in vitro* (Figure 5.4B). To assay for degradation *in vivo*, we grew strain JD704 (W3110 *lac*I-DAS+4, *ssp*B^{-}, pJD427) in M9 media at 37 °C to mid-log phase and added 10 μM rapamycin, 5 mM IPTG as a positive control for induction, or an equal volume of DMSO (the solvent for rapamycin and IPTG). At different times, samples were taken and β-galactosidase activity was assayed and normalized to OD_{600}. Addition of either rapamycin or IPTG led to increased β-galactosidase activity (Figure 5.4C). By contrast, addition of DMSO had no effect. Importantly, otherwise identical strains lacking the LacI degradation tag showed no rapamycin response (Figure 5.4C). Rapamycin-dependent induction of *lac*Z mRNA levels was also observed by RT-qPCR, confirming a direct effect on transcription (Fig 5.4D). Compared with IPTG induction, rapamycin induction showed a lag and somewhat slower kinetics (Figure 5.4C, D). This lag may result from slow diffusion of the drug into the cell or from relatively slow ClpXP-mediated degradation. We note, however, that rapamycin-dependent induction still occurred on the time scale of many biological processes.

Compared to a strain containing wild-type LacI, the strain bearing LacI-DAS+4 exhibited increased β-galactasidase activity, even in the absence of rapamycin (Figure 5.4C). In principle, this derepression could arise for several reasons. For example, LacI-DAS+4 may be degraded in

the absence of adaptor-mediated tethering *in vivo*, either by ClpXP or another cellular protease. Alternatively, the C-terminal DAS+4 tag, either alone or bound to $SspB^{CORE}$-FRB might reduce the ability of LacI to repress transcription. Finally, introduction of the DAS+4 tag might disrupt the *lac* operon at the DNA level. To help determine the contribution of each of these alternatives, we added rapamycin or a mock control (DMSO), waited 1 h, and assayed β-galactasidase in strains containing wild-type LacI, LacI-DAS+4, or LacI-LDD+4, a tag that prevents ClpXP degradation (McGinness *et al.*, 2006) (Figure 5.4E). To control for adaptor effects, strains were also grown in the presence or absence of pJD427. In the absence of rapamycin and adaptor, introduction of the DAS+4 and LDD+4 tags led to small increases in β-galactasidase levels, with a somewhat larger effect for DAS+4. With adaptor but no rapamycin, the β-galactasidase activities of LDD+4 and DAS+4 strains increased, suggesting that additional derepression may arise from binding of the $SspB^{CORE}$-FRB adaptor component to the tag. Finally, in presence of rapamycin plus adaptor, β-galactasidase levels increased substantially in the *lac*I-DAS+4 strain but did not increase in the *lac*I-LDD+4 strain. This result confirms that LacI-LDD+4 is not degraded by ClpXP.

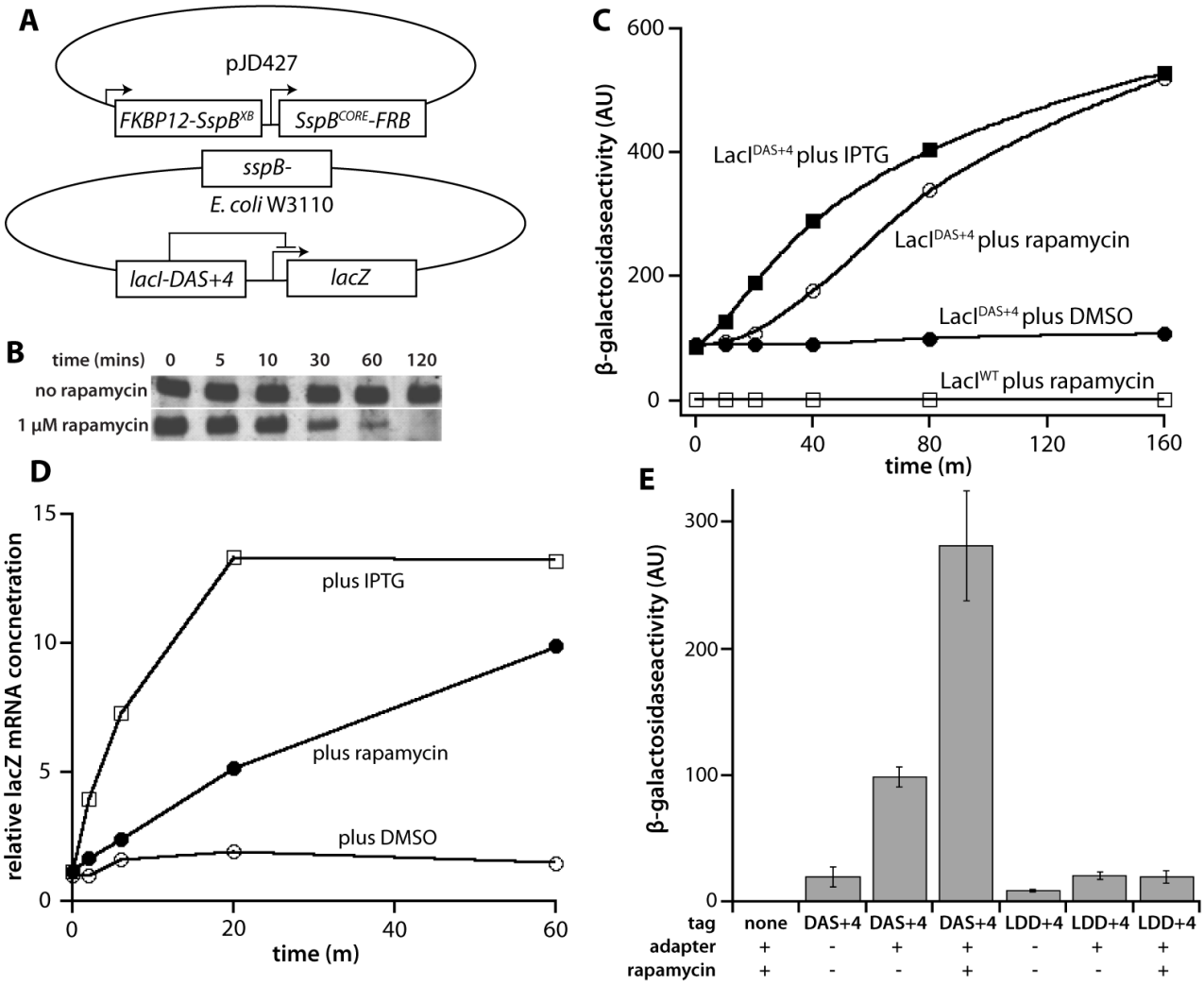

Figure 5.4. Degradation of LacI-DAS+4. (A) Assay strains contained an *ssp*B deletion and a DAS+4 tag at the C-terminus of LacI, which represses transcription of *lac*Z. Production of the split-adaptor components is mediated by plasmid-borne constitutive promoters. **(B)** Time course of degradation of LacI-DAS+4 (5 μM) by FKBP12-$(ClpX^{\Delta N})_3$ (0.3 μM), ClpP (0.9 μM), $SspB^{CORE}$-FRB (0.6 μM), and 1 μM rapamycin. **(C)** β-galactosidase activities were measured in strains containing LacI-DAS+4 following addition of rapamycin (10 μM), IPTG (5 mM), or DMSO. An isogenic LacI strain showed no response to rapamycin. **(D)** Concentrations of *lac*Z mRNA were measured by RT-qPCR and normalized to total RNA following addition of rapamycin (10 μM), IPTG (5 mM), or DMSO. **(E)** β-galactosidase activities were assayed 1 h after rapamycin or DMSO addition in strains containing LacI, LacI-DAS+4, or LacI-LDD+4 with or without the plasmid encoding the split-adaptor components.

Identification of tags that show minimal degradation without rapamycin

In an attempt to identify ssrA-tag variants that showed minimal degradation in the absence of rapamcyin, randomization and recombination methods were used to alter the three C-terminal residues of the tagged LacI protein. This procedure inserted a kanamycin-resistance marker downstream of the *lac*I gene in the chromosome. For comparison, otherwise isogenic strains bearing DAS+4 and LDD+4 tags were also constructed. We screened 92 clones from the randomized library in the presence of the split-adaptor components (see Fig. 5.S3 for selected results, *Supplementary Materials*). Based on these results, we focused on two clones, with C-terminal sequences RCN and KHG, that exhibited β-galactosidase activity equivalent to the LDD+4 strain without rapamcyin but increased activity upon addition of rapamycin (Figure 5.5). Interestingly, neither of these tag sequences resembles those found in good ClpXP substrates. Moreover, in both cases, addition of rapamycin resulted in lower β-galactosidase levels than observed in the isogenic DAS+4 strain (Figure 5.5). These tags may be useful in cases where the limited rapamycin-independent degradation observed for the DAS+4 tag inhibits function.

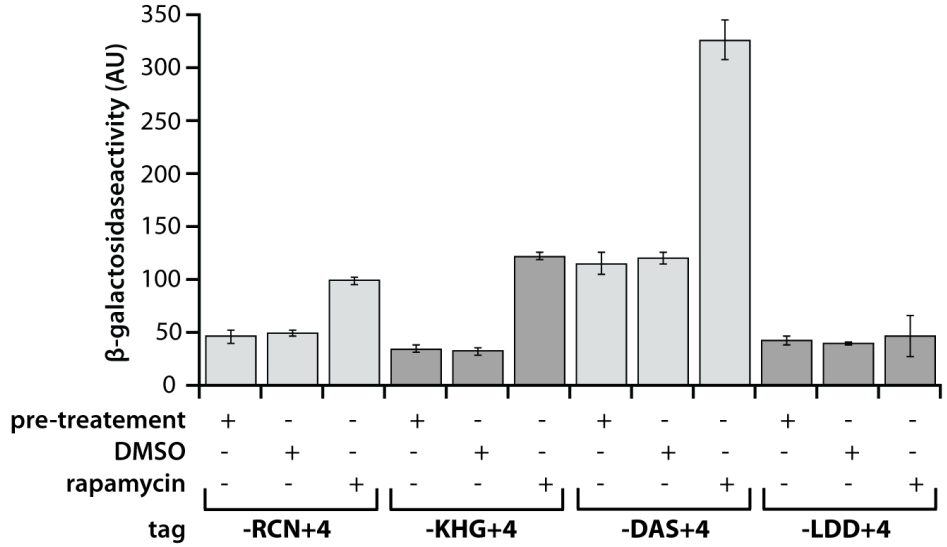

Figure 5.5. New degradation tags. β-galactosidase activities normalized to cell density were assayed in strains containing LacI variants with RCN, KHG, DAS, or LDD at the C-terminus before or 90 min after addition of DMSO or rapamycin.

Use of an orthogonal adaptor

The experiments described above were performed in *ssp*B-null strains to prevent uncontrolled delivery of DAS+4 substrates to ClpXP. To extend the potential applicability of the split-adaptor system, we investigated whether altering the adaptor and degradation tag would allow rapamycin-dependent degradation without needing to remove *E. coli* SspB. Previous studies revealed that the *C. crescentus* ssrA tag was bound well by the cognate *C. crescentus* SspB but contained a sequence change known to dramatically weaken binding by *E. coli* SspB (Flynn *et al.*, 2001; Chien *et al.*, 2007a; Griffith and Grossman, 2008). Moreover, substrates fused to a *C. crescentus* ssrA tag were degraded *E. coli* ClpXP. In initial experiments, we found that a protein consisting of residues 1-125 of *C. crescentus* SspB fused to FRB was actually a tethering-dependent substrate for ClpXP degradation but removal of the first 10 residues eliminated this proteolytic susceptibility (Figure 5.6A).

Using the stabilized *C. crescentus* fusion protein (*cc*SspBΔ10-FRB) and FKBP12-SspB^{XB}, ClpXP degraded GFP bearing a *cc*DAS+4 tag (ADNDNFAEESENYADAS) in a rapamcyin-dependent fashion (Fig 5.6B). Importantly, no degradation of this substrate was observed when *E. coli* SspB was substituted for the split-adaptor components (Fig 5.6B). These results suggest that use of the *cc*DAS+4 tag and appropriate split-adaptor components should allow rapamycin-dependent degradation in *ssp*B^{+} *E. coli* strains.

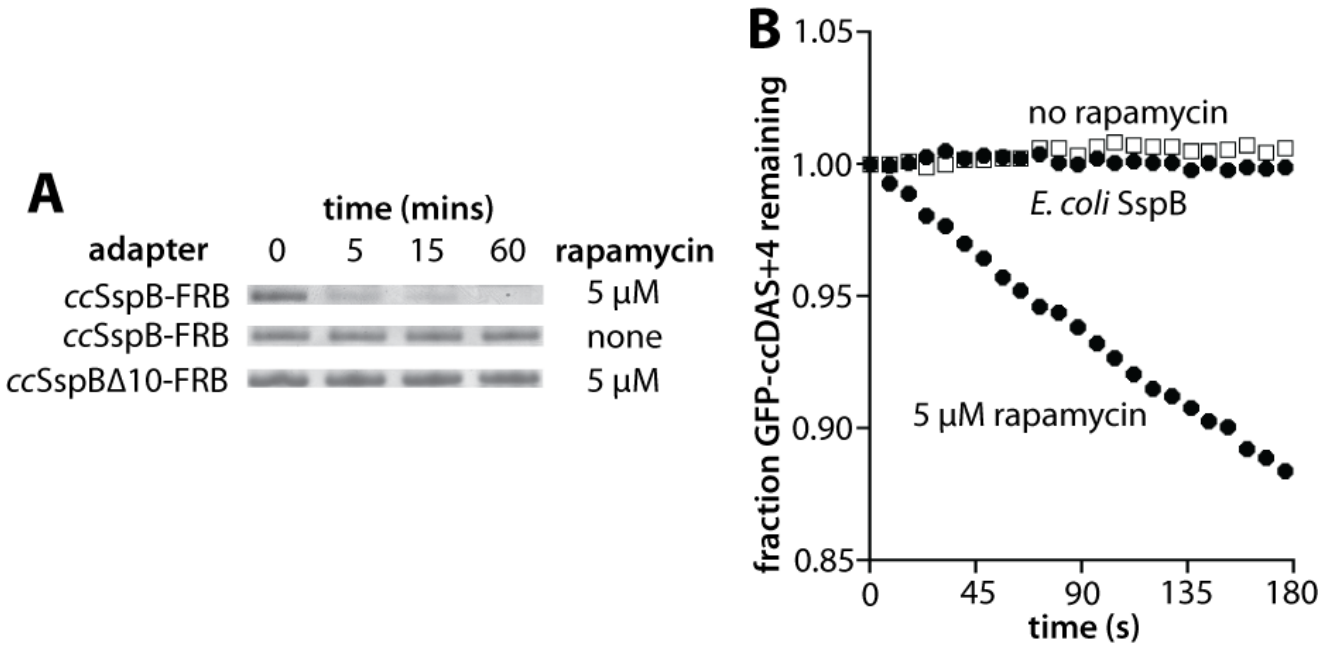

Figure 5.6. Degradation assays using *C. crescentus* SspB-FRB fusions. (**A**) A fusion protein consisting of residues 1-125 of *C. crescentus* SspB and FRB (4 μM) was degraded in a rapamycin-dependent fashion by FKBP12-(ClpX$^{\Delta N}$)$_3$ (0.3 μM) and ClpP (0.9 μM) as assayed by SDS-PAGE. Removing the 10 N-terminal residues of this fusion protein eliminated detectable degradation. **(B)** GFP-*cc*DAS+4 (2 μM) was degraded in a rapamycin-dependent reaction by ClpX (0.3 μM), ClpP (0.9 μM), *cc*SspBΔ10-FRB (5 μM), and FKBP12-SspB^{XB} (5 μM). No degradation of GFP-*cc*DAS+4 was observed when *E. coli* SspB (5 μM) was substituted for the split-adaptor components.

Discussion

Adaptor proteins must be capable of binding substrates and also tethering the bound substrate to a protease (Fig. 5.7A). In the wild-type SspB adaptor, these functions reside in a core substrate-binding domain and an unstructured C-terminal tail, respectively (Wah *et al.*, 2003; Bolon *et al.*, 2004a; Bolon *et al.*, 2004b; McGinness *et al.*, 2007). The results presented here show that these adaptor functions can reside in separate fusion proteins, with adaptor assembly and function depending on the presence of rapamycin, which mediates dimerization of the FRB and FKBP12 domains of the split-adaptor components (Fig. 5.7B). This split-adaptor system allows rapamycin-dependent degradation of appropriately tagged substrates by the wild-type ClpXP protease both *in vitro* and *in vivo*.

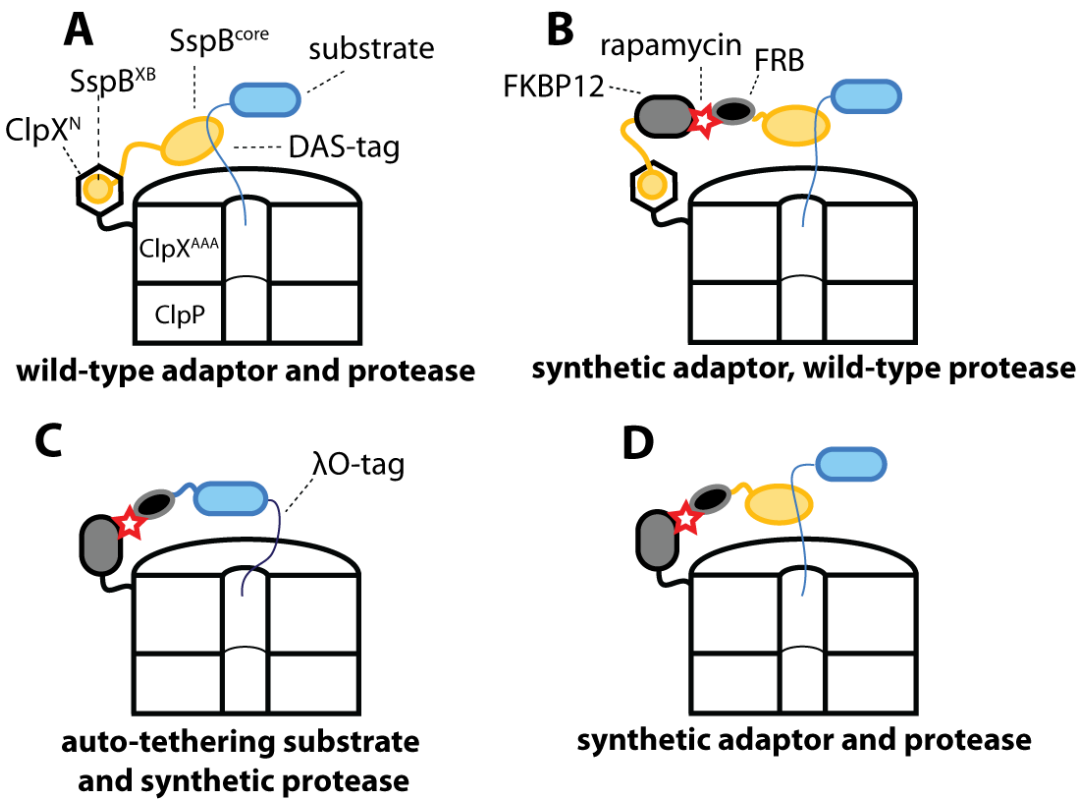

Figure 5.7. Substrate-delivery strategies. (A) SspB-mediated delivery of a DAS+4 tagged substrate to wild-type ClpXP. **(B)** Rapamycin-dependent delivery of a DAS+4 substrate to wild-type ClpXP using the split-adaptors FKBP12-SspB^{XB} and SspB^{CORE}-FRB. **(C)** Rapamycin-dependent delivery of a λO-substrate-FRB molecule to FKBP12-(ClpX$^{\Delta N}$)$_3$/ClpP. **(D)** Rapamycin-dependent SspB^{CORE}-FRB delivery of a substrate to FKBP12-(ClpX$^{\Delta N}$)$_3$/ClpP. Tagged substrates are colored blue. Rapamycin is depicted as a red star. The substrate-binding domain of SspB is shown as an orange oval. The ClpX-binding tail of SspB is shown as an orange circle. FRB is shown as a black oval. FKBP12 is shown as a larger gray oval. The N-domain of ClpX is shown as an open hexagon.

In previous studies, we investigated different strategies for rapamycin-controlled degradation (Davis *et al.*, 2009). In one system (Figure 5.7C), substrates bearing a weak N-terminal degron were fused to FRB, and FKBP12 was fused to three covalently linked subunits of ClpX lacking the N-domain (FKBP12-(ClpX$^{\Delta N}$)$_3$). This system requires modifications of both termini of the substrate and also relies on a highly engineered variant of ClpXP. A second system (Fig. 5.7D), used the same FKBP12-(ClpX$^{\Delta N}$)$_3$ enzyme and depended on a SspB^{CORE}-FRB fusion protein to deliver DAS+4 tagged substrates in a rapamycin-dependent fashion. Use of FKBP12-(ClpX$^{\Delta N}$)$_3$

systems *in vivo* was complicated by the requirement to use *clp*X^{-} strains to avoid subunit mixing and difficulties in maintaining functional FKBP12-(ClpX$^{\Delta N}$)$_3$ genes in *rec*A^{+} cells (Figure 5.S1, *Supplemental Material*). Moreover, use of monomeric FKPB12-ClpX$^{\Delta N}$, which can be maintained in *rec*A^{+} cells, required nearly perfect stoichiometry between the protease and the SspB^{CORE}-FRB adaptor for efficient substrate degradation (Figure 5.S2, *Supplemental Material*).

Our split-adaptor system retains the advantages of small-molecule control but utilizes the wild-type endogenous ClpXP protease. This feature eliminates problems associated with using the FKBP12-(ClpX$^{\Delta N}$)$_3$ enzyme for targeted intracellular degradation and also allows normal ClpXP degradation of natural substrates that require the ClpX N-domain for recognition, degradation, and homeostasis. ClpXP is present is many bacteria, and *E. coli* SspB has been shown to deliver substrates to ClpX homologs from *C. crescentus* and *B. subtilis* (Chien *et al.*, 2007a; Griffith and Grossman, 2008). Porting the split-adaptor system to these or other bacterial species may be as simple as introducing a plasmid driving roughly comparable production of the two adaptor components and adding an appropriate degradation tag to the protein target of interest. We found that a split-adaptor system using some parts of *C. crescentus* SspB and modified *C. crescentus* degradation tags worked well *in vitro*, in a manner that was not affected by the presence of *E. coli* SspB. This ability to interchange modular components should prove useful in optimizing split-adaptor systems with altered degradation-tag specificity.

We have shown that modification of the wild-type LacI repressor by addition of a C-terminal DAS+4 tag allows rapamcyin-dependent degradation and induction of β-galactosidase expression. The kinetics of β-galactosidase induction using rapamycin were slower than those

observed using IPTG, a small molecule that binds directly to LacI repressor and reduces its affinity for operator DNA ~1000-fold (Lewis *et al.*, 1996; Falcon and Matthews, 2000; Wilson *et al.*, 2007). Rapamycin may enter cells more slowly than IPTG, or the rate of rapamycin-mediated induction may be limited by other factors, such as the rate of repressor dissociation from the operator or the rate of ClpXP degradation. Importantly, however, similar final levels of β-galactosidase expression were observed for both small-molecule inducers. Other transcription factors may also be potential targets for rapamycin-dependent induction. For example, we found that modification of the tetracycline repressor (derived from transposon Tn10) with a C-terminal DAS+4 tag resulted in rapamycin-dependent ClpXP degradation *in vitro* (unpublished observations).

The addition of any sequence, whether or not it mediates conditional degradation, has the potential to perturb normal protein function. For example, compared to wild-type *lac*I strains, we observed higher levels of basal β-galactasidase synthesis in strains in which the DAS+4 tag was appended to the C-terminus of LacI. Control experiments showed that some of this effect was caused by adaptor-independent degradation, some by degradation-independent binding of one of the split-adaptor components to the tag, and some by simple addition of a C-terminal extension. By screening a library of tag mutants, we identified sequences that exhibited reduced basal β-galactasidase expression but still allowed rapamycin-dependent degradation, albeit to a lesser extent than observed for the DAS+4 tagged protein.

Targeted degradation that relies on rapamycin-dependent assembly of split adaptors has many potential applications. Because this system does not require synthesis of new adaptor proteins

(McGinness *et al.*, 2006; Griffith and Grossman, 2008), degradation could be initiated at the same time as inhibition of transcription or translation, allowing for faster clearance of target proteins from the cell. Our system could also be used to target ribosomal proteins for degradation. Indeed, Moore *et al.* (2008) demonstrated that ClpXP could forcefully extract a ribosomal protein from an intact 50S particle *in vitro*. Split-adaptor degradation systems might allow similar experiments *in vivo*, helping to elucidate the physiological function of essential genes required for translation.

Materials and Methods

Plasmids and strains

GFP variants used in this study contained the sequence H_6-IDDLG at the N-terminus in addition to the mutations S2R, S65G, S72A, and M78R. The sequences of the DAS+4, LDD+4, and *cc*DAS+4 tags were AANDENYSENYADAS, AANDENYSENYALDD, and ADNDNFAEESENYADAS, respectively.

LacI degradation experiments were performed in *E. coli* strain W3110 *ssp*B$^-$ with the degradation tag (DAS+4 or LDD+4) introduced at the C-terminus of *lac*I as described below. Strains also contained pJD427, a pSB3C5-derived (www.partsregistry.org) plasmid, which drives constitutive production of the split adaptor components, SspB^{CORE}-FRB and FKBP12-SspB^{XB} from the proB and proC promoters, respectively (Chapter 4 describes these promoters). For each split-adaptor gene, transcription was terminated using the Bba_B0011 element and translation was initiated using the Bba_B0032 ribosome-binding site (for descriptions of these sequences, see www.partsregistry.org).

A library of degradation-tag variants fused to *lac*I was generated by amplifying *lac*I-DAS+4 from a plasmid (pJD263) using degenerate primers to randomize the three C-terminal amino acids (see Table 5.1). This product was then cut with EcoRI and SpeI restriction enzymes and ligated in a 3-part reaction to a kanamycin marker, Bba_P1003 (www.partsregistry.org) and the vector pSB3C5. Transformation of *E. coli* ER2566 (New England Biolabs) with the ligation product resulted in ~6,000 clones. Sequencing of 10 clones from the library revealed a unique in-frame C-terminal tag. Mini-prepped plasmid from the pooled library (~6,000 elements) was used

as a template in PCR to amplify *lac*I-XXX+4 fused to the kanamycin-resistance marker. The primers used in this reaction encoded 40 nucleotides of homology to sequences immediately 3´ of the *lac*I stop codon (see Table 5.1). Purified dsDNA from the PCR reaction was electroporated into W3110 *ssp*B^{-}, which had been prepared for λ-red-mediated recombineering as described previously (Datta *et al.*, 2006). After allowing cells to recover for 12 h at 30 °C, successful recombinants were selected by addition of kanamycin. Recombinants were pooled, grown at 37 °C to cure the pSIM5 recombineering plasmid, and transformed with pJD427, the split-adaptor plasmid. For comparisons, *lac*I-DAS+4-kanamycin^{R} and *lac*I-LDD+4-kanamycin^{R} were also recombined onto the chromosome and were used for the experiments described in Figure 5.4E.

*lac*Ifwd-BioBrick	gaattcgcggccgcttctaggtgaaaccagtaacgttatacg
*lac*IDASrev-BioBrick	ctgcagcggccgctactagtattagctagcgtcagcatagtt ttcgctgtagttttcatc
*lac*IXXXrev-BioBrick	ctgcagcggccgctactagtattamnnmnnmnnagcatagtt ttcgctgtagttttcatc
*lac*I_chromosome_insertion_fwd	TGTTATATCCCGCCGTTAACCACC
*lac*I_chromosome_insertion_rev	TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCct gcagcggccgctactagt
*ssp*B-a	AAGCAGAACGTGAAATGCGTCTGGGCCGGAGTTAATCTGTga attcgcggccgcttctag
*ssp*B-b	CATTAAAAAGACAAAACAGGCCGCCTGGGCCTGTTTTGTAct gcagcggccgctactagt
*ssp*B-ko-1	AAGCAGAACGTGAAATGCGTCTGGGCCGGAGTTAATCTGTac tgatttgttgtgaagtaa
*ssp*B-ko-2	CATTAAAAAGACAAAACAGGCCGCCTGGGCCTGTTTTGTAtt acttcacaacaaatcagt
*lac*I-a	GCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGga attcgcggccgcttctag
*lac*I-a	TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCct gcagcggccgctactagt
*lac*I-DAS-1	GCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGgc agctaacgatgaaaacta
*lac*I-DAS-2	TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCtc agctagcgtcagcatagt

Table 5.1. Recombineering primers. Residues underlined are used to form duplexes with the target sequence. Residues in upper-case font are homologous to a locus on the *E. coli* chromosome. Residues **m** were synthesized with a mixture of A and C at that position; positions with **n** were synthesized with a mixture of all bases.

Scarless λ-red mediated chromosomal manipulation

The λ-red-mediated recombineering machinery, which is encoded on the plasmid vector pSIM5, was used to integrate degradation tags onto the chromosome and to knock-out *ssp*B (Datta *et al.*, 2006). As described in Chapter 1, scarless genomic manipulation was achieved by first introducing a cassette encoding both a selectable and counter-selectable marker. Successful recombinants were identified using the selectable marker, the entire cassette was then targeted for replacement, and recombinants were identified using the second counter-selectable marker.

Unlike previous methods, which relied on SacB or I-SceI for counter selection, we utilized a mutant variant of *E. coli* phenylalanine tRNA synthatase (mPheS), which incorporates ρ-chlorophenylalanine (ρ-Cl-Phe) into cellular tRNA and proteins (Kast, 1994; Lalioti and Heath, 2001; Cox *et al.*, 2007; Griffith *et al.*, 2009). Importantly, expression of mPheS results in cell death only in the presence of ρ-Cl-Phe.

Our targeting cassette contained a kanamycin-resistance marker (Bba_P1003), with production of mPheS driven by a constitutive promoter, (Bba_J23116), and ribosome-binding site, (Bba_B0032; www.partsregistry.org). This cassette was PCR amplified from plasmid pJD141, using primers with 20 bp of homology to the cassette and 40 bp of homology to target either the *ssp*B ORF (primers *ssp*B-a, *ssp*B-b, Table 5.1) or the 3´ terminus of *lac*I (primers *lac*I-a, *lac*I-b, Table 5.1). Plasmid DNA was removed by restriction digestion with DpnI, followed by gel purification of the PCR product.

After induction of the λ-red recombination proteins by heat shock for 15 min at 42 °C, PCR products (100 ng) were electroporated into cells, which were then allowed to recover for 6 h at 30 °C, before plating on LB/kanamycin (20 μg/mL) at 30 °C. Successful recombinants exhibited resistance to 20 μg/mL kanamycin and sensitivity to 16 mM ρ-Cl-Phe and were verified by colony-PCR. dsDNA cassettes bearing the desired insertion sequence (flanked by the same 40 bp overhangs described above) were prepared by PCR using primers *ssp*B-ko-1, *ssp*B-ko-2 (generating *ssp*B^{-}) or *lac*I-DAS-1, *lac*I-DAS-2 (resulting in *lac*I-DAS+4). Cassettes were electroporated into cells prepared for recombination, and were allowed to recover at 30 °C for 6 h before plating on YEG-agar/16 mM ρ-Cl-Phe at 30 °C (YEG-agar consists of 0.5% yeast

extract, 1% NaCl, 0.4% glucose, 1.5 % agar). Successful replacement of the mPheS-kan^{R} cassette was verified by PCR amplification and sequencing of the region of interest. pSIM5 was cured from the cells via serial dilution and growth at 30 °C under non-selective conditions.

Most ρ-Cl-Phe resistant colonies contained the insertion of interest. Often, sequencing revealed that the encoded mPheS no longer carried the mutation required for incorporation of ρ-Cl-Phe. These revertants were only observed after induction of the λ-red recombination system and probably result from recombination between the endogenous wild-type copy of Phe and the mPheS mutant.

Protein purification

ClpX, ClpP, SspB, GFP-DAS+4, and GFP-*cc*DAS+4 were expressed and purified as described (McGinness *et al.*, 2006; Davis *et al.*, 2009; Shin *et al.*, 2009). FKBP12-SspB^{XB} (which contains a N-terminal thrombin-cleavable His_6 tag) was expressed from a pET28 vector in *E. coli* BLR (F^{-}, *omp*T, *gal*, *dcm*, *lon*, *hsd*S_B(r_B^- m_B^-), λ(DE3), *rec*A^{-}). Cells were grown at room temperature grown in 1.5xYT broth (1.3 % tryptone, 0.75 % yeast extract, and 0.75 % NaCl, [pH 7.0]) to OD_{600} 0.7, induced with 1 mM IPTG, and harvested by centrifugation 4 h after induction. Cell pellets were resuspended in LB1 buffer (20 mM Hepes [pH 8.0], 400 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, and 10 mM 2-mercaptoethanol) and lysed by addition of 1 mg/mL lysozyme followed by sonication. Benzonase was added to lysates for 30 min prior to centrifugation at 8000 rpm in a Sorvall SA800 rotor. The supernatant was applied to a Ni^{2+}-NTA affinity column, washed with 50 mL of LB1 buffer, and eluted with LB1 buffer supplemented with 190 mM imidazole. Fractions containing FKBP12-SspB^{XB} were pooled and

chromatographed on a Sephacryl S-100 gel filtration column in GF-1 buffer (50 mM Tris-HCl [pH 7.6], 1 mM dithiothreitol, 300 mM NaCl, 0.1 mM EDTA, and 10% glycerol). Fractions were analyzed by SDS-PAGE, concentrated to 1 mL, and incubated with thrombin overnight at room temperature. Cleaved FKBP12-SspB^{XB} was purified away from thrombin using a S100 gel-filtration column. Analysis by SDS-PAGE confirmed complete cleavage. FKBP12-SspB^{XB} was concentrated and stored at -80 °C.

SspB^{CORE}-FRB, *cc*SspB-FRB, and truncated variants contained internal His_6 tags separating the two domains and were expressed from a pACYC-derived vector in *E. coli* BLR as described for FKBP12-SspB^{XB}. Harvested cells were resuspended in LB2 buffer (100 mM NaH_2PO_4 [pH 8.0], 10 mM Tris-HCl, 6 M GuHCl, 300 mM NaCl and 10 mM imidazole) and stored at -80 °C prior to purification. Cells were lysed by rapidly thawing the cell pellet, followed by centrifugation as described above. The supernatant was applied to a Ni^{2+}-NTA affinity column, washed with 50 mL LB2 buffer, and eluted with LB2 buffer supplemented with 240 mM imidazole. After overnight dialysis against GF2 buffer (20 mM Tris-HCl [pH 8.0], 10% glycerol, 25 mM NaCl, 25 mM KCl, 1 mM dithiothreitol), soluble protein was chromatographed on a Sephacryl S100 gel-filtration column. Fractions were analyzed by SDS-PAGE, concentrated, and stored at -80 °C. LacI-DAS+4 containing a cleavable His_6 tag was purified by Ni^{2+}-NTA affinity chromatography, tag cleavage, and gel-filtration chromatography as described above.

β-galactosidase activity assays

To measure β-galactosidase activity, strains were grown in 1 mL of supplemented M9 media (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$,

0.4% glycerol, and 35 µg/mL chloramphenicol when appropriate) at 37 °C in aerated culture vials. At mid-log phase, the OD_{600} was measured, an aliquot (20 µL) was taken for a Miller assay, and the cells were treated either with rapamycin (10 µM), IPTG (5 mM), or an equal volume of DMSO (the solvent for both IPTG and rapamycin). The final DMSO concentration in each sample was 0.5%. Growth of this strain was unaffected by DMSO concentrations up to 4% (data not shown). At different times after treatment, OD_{600} was measured (150 µL) in a SpectraMax plate reader (Molecular Devices) and samples (20 µL) were quenched by adding 80 µL Z-lysis buffer (B-PERII (Pierce) supplemented with 200 µg/mL spectinomycin and 1 mM PMSF). A 1 mL volume of Z-assay buffer (66 mM Na_2PO_4 [pH 7.4], 6 mM KCl, 700 µM $MgCl_2$, 1 mM DTT, and 0.67 mg/mL *o*-nitrophenyl β-d-galactoside) was then added to each sample, before aliquoting samples (150 µL) into a 96-well plate and measuring absorbance at 420 nm as a function of time using a SpectraMax plate reader (Molecular Devices). Sample OD_{600} was corrected for path-length using the equation,

$$OD_{600} = (OD_{600\,measured} - 0.04) * 3.39.$$

For each sample and each timepoint, β-galactosidase activity was reported as

$$\beta\text{-gal}_{\text{activity}} = 10{,}000 * \frac{\frac{\Delta A420}{\Delta time\ \ (mins\ )}}{OD_{600}}.$$

To facilitate screening the library of C-terminal degradation tags, a high-throughput form of this assay was developed. Cultures were grown overnight in LB/chloramphenicol (35 µg/mL) in 96-deep-well plates (Greiner Bio-One), diluted 100-fold into M9 media (600 uL), and grown at 37 °C with aeration for 4 h in deep-well plates. 150 µL of each sample was used to measure OD_{600} in a 96-well plate, and 10 µL of culture was added to 40 µL of Z-lysis buffer. 200 µL Z-assay

buffer was added to each sample and the change in absorbance as a function of time was measured using a plate reader as described above.

qPCR

Strains were grown at 37 °C in culture tubes in 5 mL of M9 broth to an OD_{600} of 0.2. At this time, 750 µL was removed, cells were harvested by centrifugation, the broth was aspirated, and the pellet was flash-frozen in liquid nitrogen. The remaining culture was treated with rapamycin (10 µM), IPTG (5 mM) or DMSO and, at each timepoint, 750 µL was removed and treated as described above. Cell pellets were thawed at room temperature for 10 min in 100 uL TE/readylyse (Qiagen). The cell lysate was kept on ice as 350 µL buffer RLT (Qiagen) was added to each sample. After vortexing, 250 µL of 95% ethanol was added and mixed, before the entire sample was applied to an RNAeasy purification column (Qiagen). A standard purification protocol was followed, and RNA was eluted in 30 µL DEPC-treated H_2O. Reverse transcription was performed at 42 °C for 1 h using random hexameric primers (Invitrogen) and Superscript II reverse transcriptase (Invitrogen). After quenching the reaction at 85 °C, RNA was removed by incubation RNaseH (NEB) at 37 °C for 1 h. The reaction was treated using a PCR Cleanup Kit (Qiagen) and total cDNA was determined spectrophotometrically. qPCR was performed on a Roche LightCycler 480 using sybrGreen reaction mix (Roche). For each sample, primers specific to either *lac*Z or FRB were used. Data were normalized to total cDNA for each sample. In each instance, measured FRB mRNA levels were constant as expected.

Degradation assays

All degradation assays *in vitro* were performed in PD buffer (25 mM Hepes KOH [pH 7.6], 5 mM $MgCl_2$, 10% glycerol, and 200 mM KCl) at 30 °C. GFP fluorescence was monitored by exciting with 467 nm light and measuring emission at 511 nm using either a SpectraMax M5 96-well fluorescence plate reader (Molecular Devices) or spectrofluorometer (Photon Technology International). Each degradation reaction contained an ATP-regeneration mix, consisting of 4 mM ATP, 16 mM creatine phosphate, and 0.32 mg/mL creatine kinase (Shin *et al.*, 2009). Degradation reactions of LacI-DAS+4 *in vitro* were quenched by boiling in GB buffer (50 mM Tris-HCl [pH 6.8], 75 mM dithiothreitol, 35 mM 2-mercaptoethanol, 0.003% bromphenol blue, 0.015% sodium dodecyl sulfate, 8 % glycerol) and resolved by SDS-PAGE.

When present, error bars represent the standard deviation of triplicate measurements.

Supplementary Material

FKBP12-(ClpX$^{\Delta N}$)$_3$ is unstable in *rec*A$^+$ cells

Using our FKBP12-(ClpX$^{\Delta N}$)$_3$–based degradation system (Figure 5.7D) in *E. coli* proved difficult. First, a plasmid (pJD151) bearing the chimeric protease, FKBP12-(ClpX$^{\Delta N}$)$_3$ was introduced into the cells. As shown in Figure 5.S1, this construct readily recombined in a *rec*A-dependent fashion even in the absence of selection for the recombination event. The plasmid instability likely results from homologous recombination between the three 1.4 kilobase ClpX$^{\Delta N}$ direct repeats that make up the FKPB12-ClpX$^{\Delta N}{}_3$ trimer (Morag *et al.*, 1999; Martin *et al.*, 2005). Such recombination would likely decrease the time-scale over which the degradation system was functional and could be limiting for many applications.

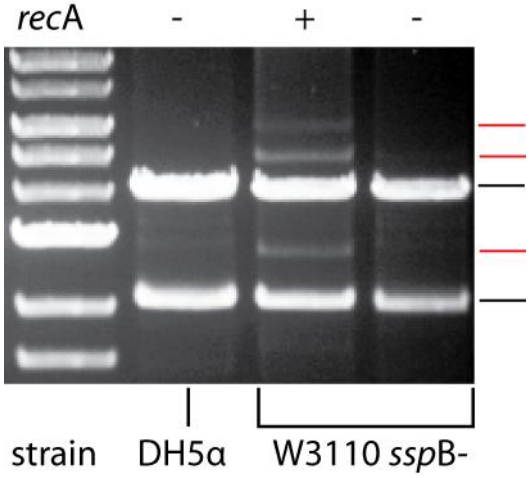

Figure 5.S1. Plasmid recombination of FKBP12-ClpX$^{\Delta N}{}_3$. A plasmid bearing FKBP12-(ClpX$^{\Delta N}$)$_3$ was transformed into *E. coli* DH5α. A single colony was picked and used to inoculate a culture that was grown to saturation. From this culture, the plasmid was miniprepped and transformed either into *ssp*B$^-$ or *ssp*B$^-$ *rec*A$^-$ *E. coli* W3110. Single colonies were picked, cultures were grown to saturation, and plasmids were purified. Each plasmid (45 ng) was digested with XbaI, PstI to remove the FKBP12-(ClpX$^{\Delta N}$)$_3$ insert, and the products of this reaction were resolved by electrophoresis on a 0.9% agarose gel. The bands marked in black correspond to the expected migration of the vector and insert. The bands marked in red are probably recombination products, as they are greatly enriched in cells containing functional *rec*A (lane 2: W3110 *ssp*B$^-$).

FKBP12-ClpX$^{\Delta N}$ is inhibited by high concentrations of the adaptor, SspB^{CORE}-FRB

To mitigate the effects of these direct repeats, we attempted to further characterize a monomeric version of the chimeric protease, FKBP12-ClpX$^{\Delta N}$ (Figure 5.7D). Previously, we had observed defects in protease hexamerization however, these defects could be overcome by using sufficiently high enzyme concentrations (Davis *et al.*, 2009). Substrates bearing the DAS+4 tag were degraded by 0.5 μM FKBP12-ClpX$^{\Delta N}$, 1.5 μM ClpP, and 3 μM SspB^{CORE}-FRB in a rapamycin-dependent fashion (data not shown). To assay the effect SspB^{CORE}-FRB on degradation, we fixed the concentration of FKBP12-ClpX$^{\Delta N}$ at 0.5 μM, ClpP at 1.5 μM, GFP-DAS+4 at 2 μM, rapamycin at 10 μM, and titrated SspB^{CORE}-FRB from 0 μM to 8 μM, measuring the rate of GFP-DAS+4 degradation at each concentration. We observed limited substrate degradation at low concentrations of adapter and pronounced inhibition of degradation at high concentrations of adaptor (Figure 5.S2). Inhibition by excess adaptor may result from the inability of the protease to exchange substrate-free adaptors for those bound to substrate due to the extreme stability of the rapamycin-induced complex.

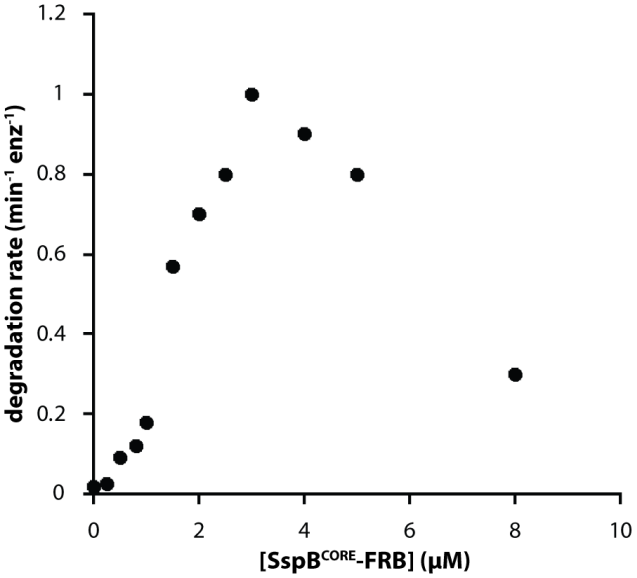

Figure 5.S2. GFP-DAS+4 degradation rate as a function of SspB^{CORE}-FRB concentration. Degradation of 2 μM GFP-DAS+4 by 0.5 μM FKBP12-ClpX$^{\Delta N}$ and 1.5 μM ClpP is strongly dependent on the relative concentration of the adaptor to the protease. Maximal degradation is observed at a 1:1 ratio of SspB^{CORE}-FRB to FKBP12-ClpX$^{\Delta N}$ (6 adaptors per hexameric unfoldase).

Characterization of a library of mutated ssrA degradation tags

We screened 92 mutants from our library of *lac*I C-terminal degradation tags. Clones were picked, grown overnight in LB/chloramphenicol/kanamycin, diluted into M9 broth, and grown for 4 h before measuring β-galactosidase activity. After addition of rapamycin (10 μM) and growth for 1 h, β-galactosidase activity was measured again. For a subset of clones, we PCR amplified the C-terminus of *lac*I and sequenced this product.

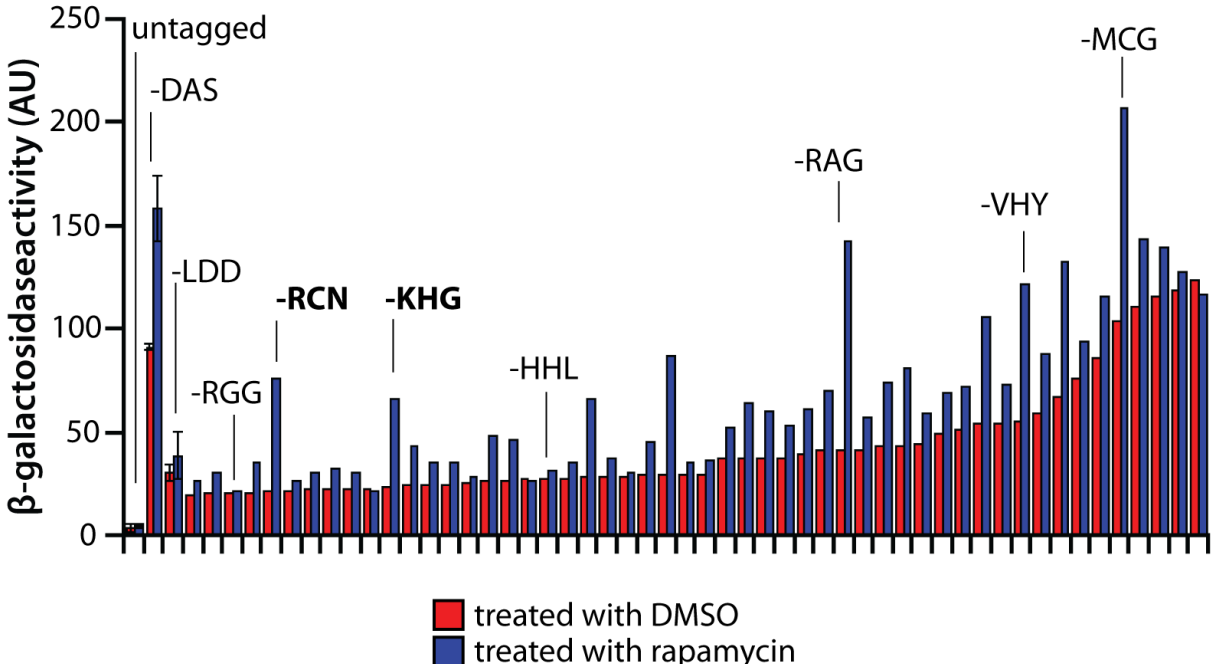

Figure 5.S3. A library of C-terminal degradation tags. β-galactosidase activity before and after addition of rapamcyin is plotted for 50/92 clones assayed. With the exception of the untagged control, DAS, and LDD, clones are sorted based on their level of activity in the absence of rapamycin. For the subset of clones sequenced (all had in-frame, error-free ssrA+4 tags), the three C- terminal amino acids are listed. Some clones show constitutive β-galactosidase activity and may result from inactivation of the repressor or introduction of a degradation tag recognized in the absence of SspB (either by ClpXP or another cellular protease).

We observed a distribution of uninduced β-galactosidase activity indicating either differential repression or degradation of these clones. Interestingly, some variants (e.g., RGG, RCN, KHG and HHL) displayed β-galactosidase activity similar that of a LDD tagged clone. Two strains

(RCN and KHG) could be induced by rapamcyin and were characterized further as described in *Results*. Although we have focused on the identification of tethering-dependent tags for ClpXP, this assay allows for a large number of clones to be screened in parallel and could prove valuable for the identification of C-terminal degradation tags for any cellular protease. If a protease were placed under the control of an inducible promoter, β-galactosidase activity could be used to identify protease-specific degrons.

Supplementary Methods

Plasmid pJD151 encodes FKBP12-$(ClpX^{\Delta N})_3$ under the control of the constitutive Bba_J23101 promoter (www.partsregistry.org) and was maintained in strain *E. coli* DH5α (F^-, λ^-, φ80*lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *deo*R, *rec*A1, *end*A1, *hsd*R17(rk^-, mk^+), *pho*A, *sup*E44, *thi*-1, *gyr*A96, *rel*A1). Assays for intergenic recombination of this construct were performed in *E. coli* W3110 (F^-, λ^-, IN(*rrn*D-*rrn*E), *rph*-1, *ssp*B^-) with *ssp*B removed from the chromosome as described in *Methods*. In one strain, *rec*A was also knocked out by P1-phage-mediated transduction of a *rec*A*::kan* cassette from an engineered *rec*A+ donor strain (a gift of Dr. Sean Moore, University of Central Florida). The kanamycin resistance marker was excised using Flp recombinase encoded on the plasmid, pCP-20, which was later cured. FKBP12-$(ClpX^{\Delta N})_3$ and FKBP12-$ClpX^{\Delta N}$ were expressed and purified as described (Davis *et al.*, 2009).

Acknowledgements

We thank the Peter Chien, Kevin Griffith, Kathleen McGinness and the entire Sauer and Baker labs for helpful discussions, equipment and reagents.

References

1. Baker, T. A., and Sauer, R. T. 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**(12): 647-653.

2. Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G., and Wandless, T. J. 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**(5): 995-1004.

3. Barkow, S. R., Levchenko, I., Baker, T. A., and Sauer, R. T. 2009. Polypeptide translocation by the AAA+ ClpXP protease machine. *Chem Biol* **16**(6): 605-612.

4. Bolon, D. N., Grant, R. A., Baker, T. A., and Sauer, R. T. 2004a. Nucleotide-dependent substrate handoff from the SspB adaptor to the AAA+ ClpXP protease. *Mol Cell* **16**(3): 343-350.

5. Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A., and Sauer, R. T. 2004b. Bivalent tethering of SspB to ClpXP is required for efficient substrate delivery: a protein-design study. *Mol Cell* **13**(3): 443-449.

6. Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., and Young, R. A. 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12**(2): 323-337.

7. Chen, J., Zheng, X. F., Brown, E. J., and Schreiber, S. L. 1995. Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. *Proc Natl Acad Sci U S A* **92**(11): 4947-4951.

8. Chien, P., Grant, R. A., Sauer, R. T., and Baker, T. A. 2007a. Structure and substrate specificity of an SspB ortholog: design implications for AAA+ adaptors. *Structure* **15**(10): 1296-1305.

9. Chien, P., Perchuk, B. S., Laub, M. T., Sauer, R. T., and Baker, T. A. 2007b. Direct and adaptor-mediated substrate recognition by an essential AAA+ protease. *Proc Natl Acad Sci U S A* **104**(16): 6590-6595.

10. Clarkson, B. K., Gilbert, W. V., and Doudna, J. A. 2010. Functional overlap between eIF4G isoforms in Saccharomyces cerevisiae. *PLoS One* **5**(2): e9114.

11. Cox, M. M., Layton, S. L., Jiang, T., Cole, K., Hargis, B. M., Berghman, L. R., Bottje, W. G., and Kwon, Y. M. 2007. Scarless and site-directed mutagenesis in Salmonella enteritidis chromosome. *BMC Biotechnol* **7**59.

12. Datta, S., Costantino, N., and Court, D. L. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**109-115.

13. Davis, J. H., Baker, T. A., and Sauer, R. T. 2009. Engineering synthetic adaptors and substrates for controlled ClpXP degradation. *J Biol Chem* **284**(33): 21848-21855.

14. Elowitz, M. B., and Leibler, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767): 335-338.

15. Falcon, C. M., and Matthews, K. S. 2000. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry* **39**(36): 11074-11083.

16. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

17. Flynn, J. M., Levchenko, I., Seidel, M., Wickner, S. H., Sauer, R. T., and Baker, T. A. 2001. Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc Natl Acad Sci U S A* **98**(19): 10584-10589.

18. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. 2003. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**(3): 671-683.

19. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A., and Sauer, R. T. 2009. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**(4): 744-756.

20. Griffith, K. L., and Grossman, A. D. 2008. Inducible protein degradation in Bacillus subtilis using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Mol Microbiol* **70**(4): 1012-1025.

21. Griffith, K. L., Fitzpatrick, M. M., Keen, E. F., 3rd, and Wolf, R. E., Jr. 2009. Two functions of the C-terminal domain of Escherichia coli Rob: mediating "sequestration-dispersal" as a novel off-on switch for regulating Rob's activity as a transcription activator and preventing degradation of Rob by Lon protease. *J Mol Biol* **388**(3): 415-430.

22. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**(14): 4121-4130.

23. Ji, Y., Zhang, B., Van, S. F., Horn, Warren, P., Woodnutt, G., Burnham, M. K., and Rosenberg, M. 2001. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**(5538): 2266-2269.

24. Joshi, S. A., Hersch, G. L., Baker, T. A., and Sauer, R. T. 2004. Communication between ClpX and ClpP during substrate processing and degradation. *Nat Struct Mol Biol* **11**(5): 404-411.

25. Kast, P. 1994. pKSS--a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138**(1-2): 109-114.

26. Kenniston, J. A., Baker, T. A., Fernandez, J. M., and Sauer, R. T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114**(4): 511-520.

27. Kenniston, J. A., Burton, R. E., Siddiqui, S. M., Baker, T. A., and Sauer, R. T. 2004. Effects of local protein stability and the geometric position of the substrate degradation tag on the efficiency of ClpXP denaturation and degradation. *J Struct Biol* **146**(1-2): 130-140.

28. Kenniston, J. A., Baker, T. A., and Sauer, R. T. 2005. Partitioning between unfolding and release of native domains during ClpXP degradation determines substrate selectivity and partial processing. *Proc Natl Acad Sci U S A* **102**(5): 1390-1395.

29. Knight, Z. A., and Shokat, K. M. 2007. Chemical genetics: where genetics and pharmacology meet. *Cell* **128**(3): 425-430.

30. Lalioti, M., and Heath, J. 2001. A new method for generating point mutations in bacterial artificial chromosomes by homologous recombination in Escherichia coli. *Nucleic Acids Res* **29**(3): E14.

31. Levchenko, I., Seidel, M., Sauer, R. T., and Baker, T. A. 2000. A specificity-enhancing factor for the ClpXP degradation machine. *Science* **289**(5488): 2354-2356.

32. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**(5253): 1247-1254.

33. Martin, A., Baker, T. A., and Sauer, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**(7062): 1115-1120.

34. McGinness, K. E., Baker, T. A., and Sauer, R. T. 2006. Engineering controllable protein degradation. *Mol Cell* **22**(5): 701-707.

35. McGinness, K. E., Bolon, D. N., Kaganovich, M., Baker, T. A., and Sauer, R. T. 2007. Altered tethering of the SspB adaptor to the ClpXP protease causes changes in substrate delivery. *J Biol Chem* **282**(15): 11465-11473.

36. Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, USA

37. Moore, S. D., Baker, T. A., and Sauer, R. T. 2008. Forced extraction of targeted components from complex macromolecular assemblies. *Proc Natl Acad Sci U S A* **105**(33): 11685-11690.

38. Morag, A. S., Saveson, C. J., and Lovett, S. T. 1999. Expansion of DNA repeats in Escherichia coli: effects of recombination and replication functions. *J Mol Biol* **289**(1): 21-27.

39. Rappleye, C. A., and Roth, J. R. 1997. A Tn10 derivative (T-POP) for isolation of insertions with conditional (tetracycline-dependent) phenotypes. *J Bacteriol* **179**(18): 5827-5834.

40. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

41. Shin, Y., Davis, J. H., Brau, R. R., Martin, A., Kenniston, J. A., Baker, T. A., Sauer, R. T., and Lang, M. J. 2009. Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc Natl Acad Sci U S A*

42. Sousa, M. C., Trame, C. B., Tsuruta, H., Wilbanks, S. M., Reddy, V. S., and McKay, D. B. 2000. Crystal and solution structures of an HslUV protease-chaperone complex. *Cell* **103**(4): 633-643.

43. Stockwell, B. R. 2000. Chemical genetics: ligand-based discovery of gene function. *Nat Rev Genet* **1**(2): 116-125.

44. Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**(7221): 516-519.

45. Taxis, C., Stier, G., Spadaccini, R., and Knop, M. 2009. Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol Syst Biol* **5**267.

46. VanGuilder, H. D., Vrana, K. E., and Freeman, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44**(5): 619-626.

47. Wah, D. A., Levchenko, I., Rieckhof, G. E., Bolon, D. N., Baker, T. A., and Sauer, R. T. 2003. Flexible linkers leash the substrate binding domain of SspB to a peptide module that stabilizes delivery complexes with the AAA+ ClpXP protease. *Mol Cell* **12**(2): 355-363.

48. Wilson, C. J., Zhan, H., Swint-Kruse, L., and Matthews, K. S. 2007. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64**(1): 3-16.

49. Wong, W. W., Tsai, T. Y., and Liao, J. C. 2007. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol Syst Biol* **3**130.

Appendix A

Supplementary Information for "Single-molecule Denaturation and Degradation of Proteins by the AAA+ ClpXP Protease"

This work was published as Yongdae Shin[†], Joseph H. Davis[†], Ricardo R. Brau[†], Andreas Martin[†], Jon A. Kenniston, Tania A. Baker, Robert T. Sauer, and Matthew J. Lang. 2009 *PNAS* 106:19340-19345.

R.R.B and J.A.K initiated this work by constructing the TIRF microscope and investigating the on-surface stability of full-length ClpX respectively. A.M. aided in experimental design as well as the construction of substrates and single-chain ClpX. Y.S. collected and analyzed the majority of the single-molecule datasets and composed the initial manuscript.

[†]These authors contributed equally to this work

Supplementary Experiments

The experiments shown in Figure A.S1 demonstrate that the biotinylated single-chain ClpX enzyme ($ClpX^{SC}$) used for single-molecule studies has essentially the same activity as wild-type ClpX in directing ClpP degradation of a GFP-ssrA substrate.

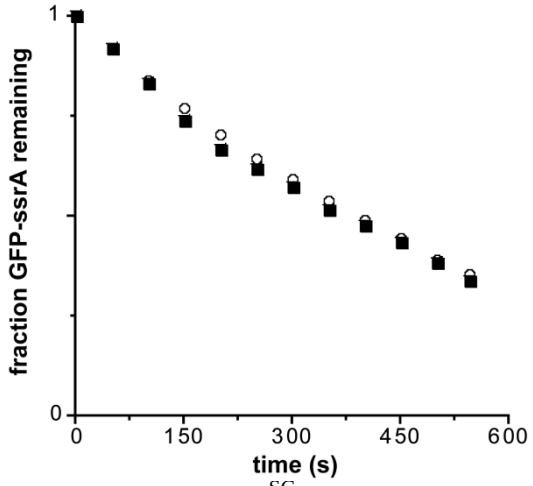

Figure A.S1. Activities of biotinylated $ClpX^{SC}$ (0.3 μM hexamer; open circles) and wild-type ClpX (0.3 μM hexamer; closed squares) in supporting ClpP (0.9 μM) degradation of GFP-ssrA (2.5 μM) as monitored by loss of GFP fluorescence. Reactions were performed in PD-1 buffer at 30 °C.

$ClpX^{SC}$/ClpP stalls after degrading most of the $titin^{V15P}$-ssrA portion of the Cy3-GFP-$titin^{V15P}$-ssrA substrate in the presence of ATPγS/Mg^{2+}, but immobilized complexes of the enzyme and partially degraded substrate remain stably associated after chelation of Mg^{2+} by EDTA (Figure A.S2, left) and after replacing ATPγS with ATP in the presence of EDTA (Figure A.S2, right).

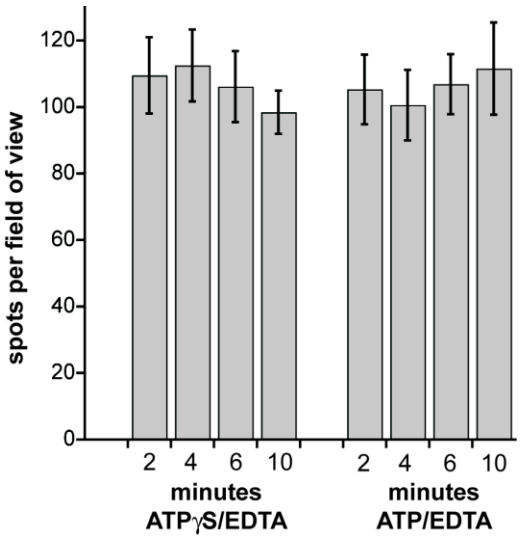

Figure A.S2. Stability of the stalled complex of Cy3-GFP-titin^{V15P}-ssrA with immobilized ClpX^{SC}/ClpP under pre-assembly conditions. (*Left*) Spots per field of view in the presence of 2 mM ATPγS and 50 mM EDTA. (*Right*) Spots per field of view after exchange into 1 mM ATP and 6 mM EDTA.

In the continual presence of ATPγS/Mg^{2+}, the fluorescence of immobilized stalled complexes is lost by photobleaching with a time constant of ≈330 s (Figure A.S3A). If immobilized complexes are washed with buffer with Mg^{2+} but no ATPγS, most spots disappear in an exponential process with a time constant of ≈12 s (Figure A.S3B), suggesting that the enzyme-substrate complex is unstable without nucleoside triphosphate.

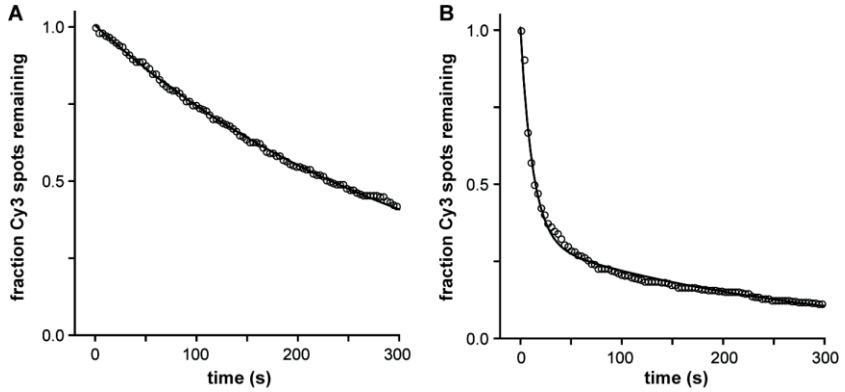

Figure A.S3. Fraction spots remaining in control single-molecule experiments. (A) Photobleaching of enzyme-bound Cy3-GFP-titin^{V15P}-ssrA was assayed in the presence of ATPγS/Mg^{2+} (circles). The line is a single-exponential fit with a time constant of 330 s. (B) Fraction of Cy3-labeled substrate remaining after washing with buffer with Mg^{2+} but no ATPγS (circles). The line is a double-exponential fit with time constants of 13.7 s for the fast phase (amplitude 71%) and 330 s (amplitude 29%) for the slow phase.

In single-molecule assays of Cy3-GFP-titinV15P-ssrA degradation by immobilized $ClpX^{SC}$/ClpP, reducing the ATP concentration from 1 to 0.1 mM slowed the reaction (Figure A.S4A). The reaction was also slowed when degradation was performed in the presence of a mixture of 1 mM ATP and 0.25 mM ATPγS (Figure A.S4B).

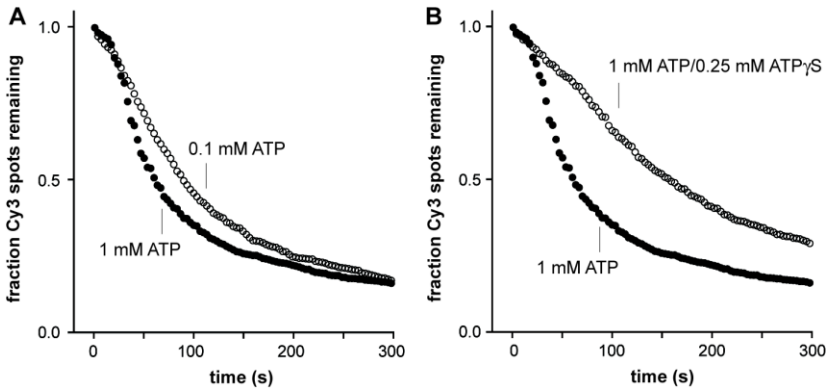

Figure S4. Nucleotide dependence of single-molecule degradation of Cy3-GFP-titinV15P-ssrA. (**A**) Open circles show assays performed with 0.1 mM ATP. Filed circles show assays performed with 1 mM ATP (taken from Figure A.3B). (**B**) Open circles show assays performed with 1 mM ATP plus 0.25 mM ATPγS. Filed circles show assays performed with 1 mM ATP (taken from Figure A.3B).

The substrates used in our single-molecule experiments were modified by addition of a cysteine, at the second amino acid, to allow N-terminal labeling with Cy3 dye. The experiments shown below (Figure A.S5) show that a substrate lacking cysteine 2 is labeled at only 6% of the efficiency of an otherwise identical substrate with cysteine 2. The low degree of fluorescent labeling of the substrate without cysteine 2 is probably caused by labeling of two cysteines in the $titin^{V15P}$ portion of the substrate, which are buried in the native molecule but become accessible upon transient unfolding. This portion of the substrate is removed by degradation by $ClpX^{SC}$/ClpP in the presence of ATPγS and would not be present in the stalled complexes that represent the starting material for our experiments.

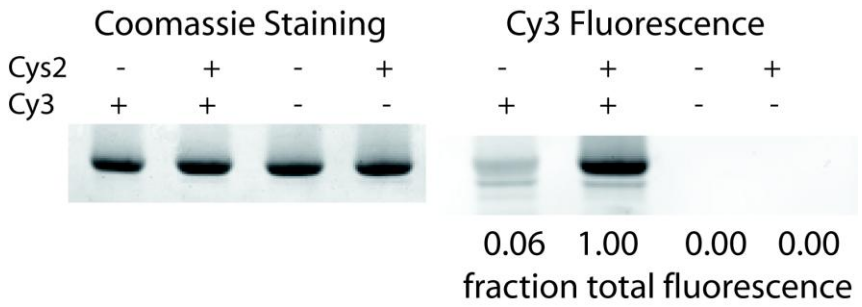

Figure S5. CFP-GFP-titin^{V15P}-ssrA substrates with (+) or without (-) cysteine at position 2 (Cys2) were labeled with Cy3 (+) or in a mock reaction (-) and were analyzed by SDS-PAGE and staining with Coomassie Blue (left panel) or by fluorography after excitation of Cy3 (right panel). The substrate without Cys2 was labeled at 6% of the level of the substrate with Cys2.

As shown in the kinetic traces below (Figure A.S6), Cy3-labeled substrate-enzyme complexes disappeared in a single kinetic step as a consequence either of completion of degradation or of photobleaching.

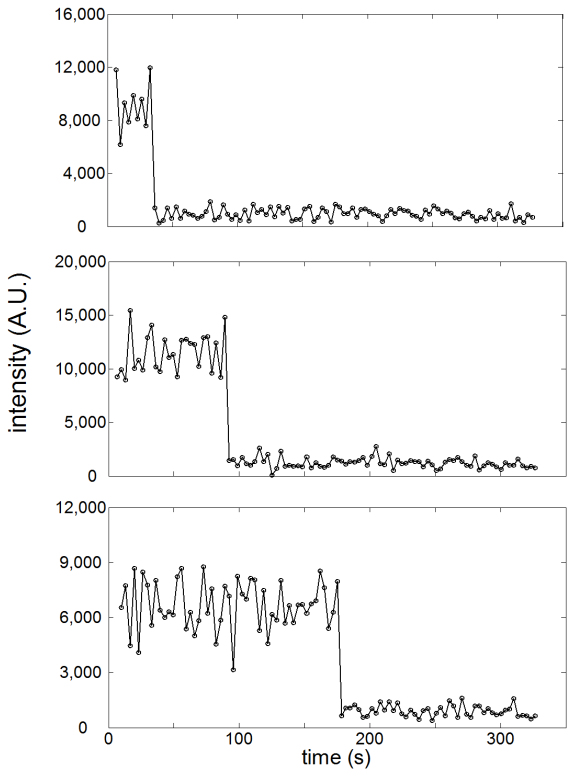

Figure A.S6. Kinetic traces of intensities of single Cy3-labeled substrate-enzyme complexes spots show that fluorescence is lost permanently in a single step.

In experiments monitoring GFP fluorescence, we reduced excitation power to extend the fluorophore lifetime and signal-to-noise was somewhat poorer than for the Cy3 experiments (Figure A.S6 and A.S7). Moreover, instances of GFP blinking were observed (marked by arrows in Figure A.S7A). The MATLAB code used for data analysis only counted degradation or photobleaching events that lead to permanent spot disappearance (Figure A.S7B).

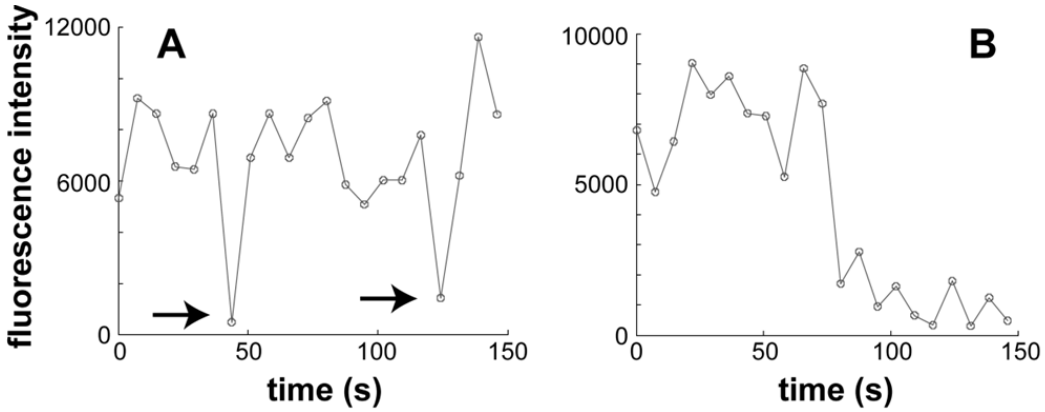

Figure S7. Kinetic traces of intensities of single GFP-substrate•enzyme complexes. (**A**) The transient decreases marked by arrows were considered to be blinking and were not counted as degradation or photobleaching events. (**B**) The permanent loss of GFP fluorescence at ≈73 s was counted as a degradation or photobleaching event.

Table A.S1 summarizes the time constants for different steps in the degradation of individual substrates. These values provide good fits of the experimental data and are generally consistent for single-molecule and solution experiments.

Table A.S1. Time constants for individual steps in degradation reactions.

		Time constants (s)					
Substrate	Experiment	Den1	Trans1	Den2	Trans2	Term	Overall
Cy3-GFP-titin^{V15P}-ssrA	Preengaged single-molecule	19	6.2			29.6	55
Cy3-CFP-GFP-titin^{V15P}-ssrA	Preengaged single-molecule	19	6.2	19	6.2	29.6	80
Cy3-titin-GFP-titin^{V15P}-ssrA	Preengaged single-molecule	19	6.2	347	3.1	29.6	405
GFP-titin^{V15P}-ssrA	Preengaged solution	25					
CFP-GFP-titin^{CM}-ssrA	Single turnover solution	20	6	20			

Supplementary Methods

Protein expression and purification

The single-chain $ClpX^{SC}$ variant contained an N-terminal FLAG tag, six repeats of *E. coli* ClpX-ΔN (residues 61–423) connected by flexible linkers (Martin *et al.*, 2005), a BirA-acceptor peptide (HAAGGLNDIFEAQKIEWHEDT; ref. Chen *et al.*, 2005), and a C-terminal H_6 tag. $ClpX^{SC}$ was expressed from a pACYC-derived vector in *E. coli* strain ERL (ER2566 with a chromosomal λ-lysozyme gene; a gift from S. Moore, University of Central Florida, Orlando, FL). To allow incorporation of an N-terminal fluorophore, the genes for all substrates (GFP-titin^{V15P}-H_6-ssrA; CFP-GFP-titin^{V15P}-H_6-ssrA; and FLAG-I27-GFP-titin^{V15P}-H_6-ssrA) were modified to encode an N-terminal Met-Cys sequence (the methionine is removed post translationally). These substrates were expressed from pACYC vectors in *E. coli* strain BLR (BL21 *recA*$^-$ λ(DE3)). The BirA enzyme contained a C-terminal H_6 tag and was expressed in strain BLR from a pET22 vector (a gift from J. Damon, M.I.T.). *E. coli* ClpP-H_6 was expressed in strain BL21 *clpP*$^-$ λ(DE3).

Enzymes and substrates were expressed from IPTG-inducible T7 promoters. Cells were grown to OD_{600} 0.7 at 37 °C in 1-2 L of 1.5xYT broth, chilled to 18 °C, and expression was induced with 1 mM IPTG. Cells were harvested 4 h later, resuspended in 15 mL of LB1 buffer [20 mM Hepes (pH 7.6), 400 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, and 10 mM 2-mercaptoethanol] per liter of culture, and frozen at -80 °C until purification. For proteins expressed in strain BLR, cells were subjected to two freeze/thaw cycles before addition of 1 mM PMSF, 0.1 mg/mL lysozyme, and 250 units of benzonase nuclease. After 30 min at 4 °C, lysates were centrifuged at 8,000 rpm in a Sorvall SA600 rotor for 40 min, and the supernatant was

decanted and saved. For proteins expressed in strain ERL, frozen cells were thawed, 10 μL/mL chloroform was added, and the mixture was vortexed briefly prior to PMSF/benzonase addition and centrifugation. BirA was purified by applying the lysate supernatant to a 1 mL Ni^{2+}-NTA affinity column, washing with 20 vol of LB1, and eluting with 8 aliquots of 0.5 mL EB1 buffer [20 mM Hepes (pH 7.6), 400 mM NaCl, 100 mM KCl, 200 mM imidazole, 10% glycerol, 10 mM 2-mercaptoethanol]. BirA fractions were applied to a Superdex S200 gel-filtration column (Amersham-Pharmacia) equilibrated with GF1 buffer [50 mM Tris-HCl (pH 7.6), 1 mM DTT, 300 mM NaCl, 0.1 mM EDTA, and 10% glycerol]. Fractions containing BirA were identified by SDS-PAGE, concentrated, and stored frozen at -80 °C. $ClpX^{SC}$ and substrates were purified by Ni^{2+}-NTA affinity and size-exclusion chromatography (Martin *et al.*, 2008; Davis *et al.*, 2009). After the affinity-chromatography step, $ClpX^{SC}$ was biotinylated as described (Chen *et al.*, 2005). ClpP-H_6 was purified as described (Kim *et al.*, 2000). To unfold the titin domains of GFP-titin^{V15P}-ssrA and CFP-GFP-titin^{V15P}-ssrA, proteins were carboxymethylated for 1 h at room temperature in 3 M GuHCl by using a 10-fold molar excess of iodoacetic acid and exchanged into fresh buffer using Ni^{2+}-NTA chromatography.

Equations used for fitting

ClpXP degradation of the preengaged Cy3-GFP-titin^{V15P}-ssrA substrate was modeled to include an unfolding step (k_1), a translocation step (k_2), and a termination step (k_3). In the model shown below, the starting enzyme-substrate complex (A_{1a}) and the intermediate species (A_{2a} and A_{3a}) are fluorescent, but the final substrate-free species A_{4a} is not. We assumed that fluorescence could also be lost by conversion of species A_{1a}, A_{2a}, or A_{3a} to a photobleached state (B) with rate constant k (i.e., we assume that each subpopulation photobleaches at the same rate).

$$\begin{array}{ccccccc} A_{1a} & \xrightarrow{k_1} & A_{2a} & \xrightarrow{k_2} & A_{3a} & \xrightarrow{k3} & A_{4a} \\ \downarrow & k & \downarrow & k & \downarrow & k & \\ & & B & & & & \end{array}$$

The relative populations of species A_{1a}, A_{2a}, and A_{3a} are given by the equations:

$$P_{A1a} = \exp(-(k + k_1)t)$$

$$P_{A2a} = \frac{k_1}{k_2 - k_1}(\exp(-(k + k_1)t) - \exp(-(k + k_2)t))$$

$$P_{A3a} = \frac{k_1 k_2}{(k_2 - k_1)(k_3 - k_1)}(\exp(-(k + k_1)t) - \exp(-(k + k_3)t)) - \frac{k_1 k_2}{(k_2 - k_1)(k_3 - k_2)}(\exp(-(k + k_2)t) - \exp(-(k + k_3)t))$$

The equation used for fitting was:

$$\text{Total number of spots} = D_1 \times (P_{A1a} + P_{A2a} + P_{A3a}) + D_2 \times \exp(-kt)$$

where D_1 is the amplitude of the degradation/photobleaching pathway and D_2 is the amplitude of a subset of spots that are lost only by photobleaching.

ClpXP degradation of the pre-engaged Cy3-CFP-GFP-titin^{V15P}-ssrA substrate was modeled to include a GFP unfolding step (k_1), a GFP translocation step (k_2), a CFP unfolding step (k_1), a CFP translocation step (k_2), a termination step (k_3), and photobleaching steps as described above.

$$\begin{array}{ccccccccccc} A_{1b} & \xrightarrow{k_1} & A_{2b} & \xrightarrow{k_2} & A_{3b} & \xrightarrow{k_1} & A_{4b} & \xrightarrow{k_2} & A_{5b} & \xrightarrow{k_3} & A_{6b} \\ \downarrow & k & \downarrow & k & \downarrow & k & \downarrow & k & \downarrow & k & \\ & & & & \text{B} & & & & & & \end{array}$$

The relative populations of species A_{1b}, A_{2b}, A_{3b}, A_{4b}, and A_{5b} are given by the equations:

$$P_{A1b} = \exp(-(k + k_1)t)$$

$$P_{A2b} = \frac{k_1}{k_2 - k_1}(\exp(-(k + k_1)t) - \exp(-(k + k_2)t))$$

$$P_{A3b} = \frac{k_1 k_2}{k_2 - k_1} t \exp(-(k + k_1)t) + \frac{k_1 k_2}{(k_1 - k_2)^2}(\exp(-(k + k_2)t) - \exp(-(k + k_1)t))$$

$$P_{A4b} = \frac{{k_1}^2 k_2}{(k_1 - k_2)^3}(((k_1 - k_2)t - 2)\exp(-(k + k_2)t) + ((k_1 - k_2)t + 2)\exp(-(k + k_1)t))$$

$$P_{A5b} = {k_1}^2 {k_2}^2 (\frac{-(k_2 - k_3)(k_1 - k_2)t - k_1 + 3k_2 - 2k_3}{(k_2 - k_3)^2 (k_1 - k_2)^3}\exp(-(k + k_2)t) + \frac{1}{(k_2 - k_3)^2 (k_1 - k_3)^2}\exp(-(k + k_3)t) + \frac{-(k_1 - k_3)(k_1 - k_2)t + k_2 - 3k_1 + 2k_3}{(k_1 - k_3)^2 (k_1 - k_2)^3}\exp(-(k + k_1)t))$$

The equation used for fitting was:

$$\text{Total number of spots} = D_3 \times (P_{A1b} + P_{A2b} + P_{A3b} + P_{A4b} + P_{A5b}) + D_4 \times \exp(-kt)$$

where D_3 is the amplitude of the degradation/photobleaching pathway and D_4 is the amplitude of a subset of spots lost only by photobleaching.

References

1. Chen, I., Howarth, M., Lin, W., and Ting, A. Y. 2005. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**(2): 99-104.

2. Davis, J. H., Baker, T. A., and Sauer, R. T. 2009. Engineering synthetic adaptors and substrates for controlled ClpXP degradation. *J Biol Chem* **284**(33): 21848-21855.

3. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. 2000. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**(4): 639-648.

4. Martin, A., Baker, T. A., and Sauer, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**(7062): 1115-1120.

5. Martin, A., Baker, T. A., and Sauer, R. T. 2008. Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat Struct Mol Biol* **15**(2): 139-145.

Appendix B

Single-Molecule FRET Assays for the Study of the AAA^+ Protease ClpXP

This work was initiated by Dr. Andreas Martin who built the first (non-sortase linked) versions of the $ClpX^{\Delta N}$ trimers and showed nucleotide-dependent florescence resonance energy transfer in bulk assays. The single molecule assays have been carried out in collaboration with Matt Lang's lab. Yongdae Shin developed the dual-view TIRF microscope assay as well as the computational tools to analyze the data.

Introduction

AAA^+ protein machines are present in all kingdoms of life and are vital for a range of cellular functions, including vesicular transport, protein secretion, microtubule motor function, and the maintenance of protein homeostasis via chaperone and proteolytic activities (Baker and Sauer, 2006; White and Lauring, 2007). These seemingly disparate actions are all dependent the ability of AAA^+ enzymes to convert the energy of ATP binding and hydrolysis into mechanical work. Pronounced sequence conservation and structural homology between members of this large family suggest the possibility of a common, underlying mechanism for the entire class (Erzberger and Berger, 2006).

E. coli ClpXP is a AAA^+ unfoldase, disassembly machine, and protease consisting of either one or two ring-shaped ClpX homohexamers that cap the ends of the tetradecameric peptidase, ClpP (Figure 1.2). The unfoldase, ClpX, oligomerizes into a ring containing a narrow axial pore that serves as a binding site for the degradation tags of substrates. Biochemical studies revealed that successive rounds of ATP hydrolysis drive conformational changes in ClpX that result in the application of a pulling force to bound proteins and eventually lead to global denaturation of the substrate. In an ATP-hydrolysis-dependent reaction, the unfolded polypeptide is translocated through the axial pore of ClpX to sequestered degradation chamber of ClpP, where it is cleaved and the released as short peptides (Sauer *et al.*, 2004). Because the ClpP active sites can only be accessed via translocation by ClpX, the system achieves a high level of substrate specificity (Joshi *et al.*, 2004).

Biochemical and structural studies indicate that ClpX is a highly asymmetric machine that undergoes large conformational changes. Indeed, Hersch *et al.* (2005) found that ClpX hexamers bind a maximum of four nucleotides in solution, implying that two or more distinct classes of subunits must exist at any given time (Hersch *et al.*, 2005). Moreover, recent crystallographic studies show that even in the absence of ATP, ClpX is asymmetric with at least two classes of subunits (Glynn *et al.*, 2009). Interestingly, Glynn *et al*. (2009) observed large conformational changes upon addition of nucleotide, hinting at a mechanism to couple the energy of ATP binding to mechanical work.

We are interested in developing techniques to directly observe nucleotide-dependent conformational changes in ClpX with the ultimate goal of answering the following questions. How many distinct states exist in the ATP-hydrolysis cycle? Which transitions are rate-limiting? Does the rate-limiting step for the cycle change in the presence of substrate? Does the protease operate in different modes during substrate unfolding and translocation? Do individual ClpX subunits act independently, or is ATP hydrolysis in one subunit sufficient to drive conformational changes in distal subunits? Is there a preferred order of subunit firing?

In collaboration with the Lang Lab (MIT, Biological Engineering), we have begun developing a fluorescence resonance energy transfer (FRET) based assay with single-molecule resolution. In bulk studies, the conformational state observed is averaged over the entire population, thus obscuring unsynchronized molecular motions. By contrast, these single-molecule experiments will allow us to inspect an isolated engine as it progresses through its reaction cycle.

Results and Discussion

The subunits in a ClpX$^{\Delta N}$ ring can be connected with flexible linkers by genetically encoding linked dimers, trimers, or the entire hexamer as a single polypeptide (Martin *et al.*, 2005). This technology allows for the generation of homogenous populations of molecules, each containing mutations in one or more specific subunits of the hexamer. These engineered enzymes recapitulate many activities of wild-type ClpX and have been used extensively to address ClpX structure-function questions (Martin *et al.*, 2005; Martin *et al.*, 2007; Martin *et al.*, 2008a; Martin *et al.*, 2008b). Popp *et al.* (2007) have described another method of linking proteins using the Sortase A enzyme from *Staphylococcus aureus* (Popp et al., 2007). This transpeptidase recognizes a short peptide sequence (LPXTG) and cleaves the peptide bond between the threonine and the glycine thereby forming a stable acyl-enzyme intermediate between the active site cysteine of sortase and the newly generated C-terminus of the substrate. This acyl-enzyme intermediate can be resolved by addition of a peptide bearing an N-terminal poly-glycine repeat, resulting in release of the enzyme and formation of a native peptide bond linking the substrate and the peptide (Antos *et al.*, 2009).

We utilized sortase-mediated protein ligation to fuse fluorescently-labeled ClpX$^{\Delta N}$ trimers (Figure B.1A). This linkage resulted in a ClpX$^{\Delta N}$ single-chain hexamer bearing two different fluorescent labels (Figure B.1B). First, we generated a single-chain trimer bearing a single-reactive cysteine at position 169 in the first subunit (*Materials and Methods*, Table B.1). At the C-terminus, we appended a linker fused to a sortase recognition site (Figure B.1A, trimer A). We then built a second single-chain trimer encoding a reactive cysteine at position 153 in the first

subunit, a biotinylation acceptor peptide at the C-terminus, and the sortase resolution sequence, GGGG, at the N-terminus (Figure B.1A, trimer B). We expressed and purified these trimers and used Alexa488-maleimide to label C169 of trimer A and Alexa647-maleimide to label C153 of trimer B. As shown in Figure B.1C, ~50% of the trimers could be covalently linked via the sortase reaction resulting in a species that chromatographs on a gel-filtration column similarly to single-chain ClpX. The reactants and products of the sortase reaction are roughly isoenergetic and thus incomplete conversion observed is expected (M. Popp, personal communication).

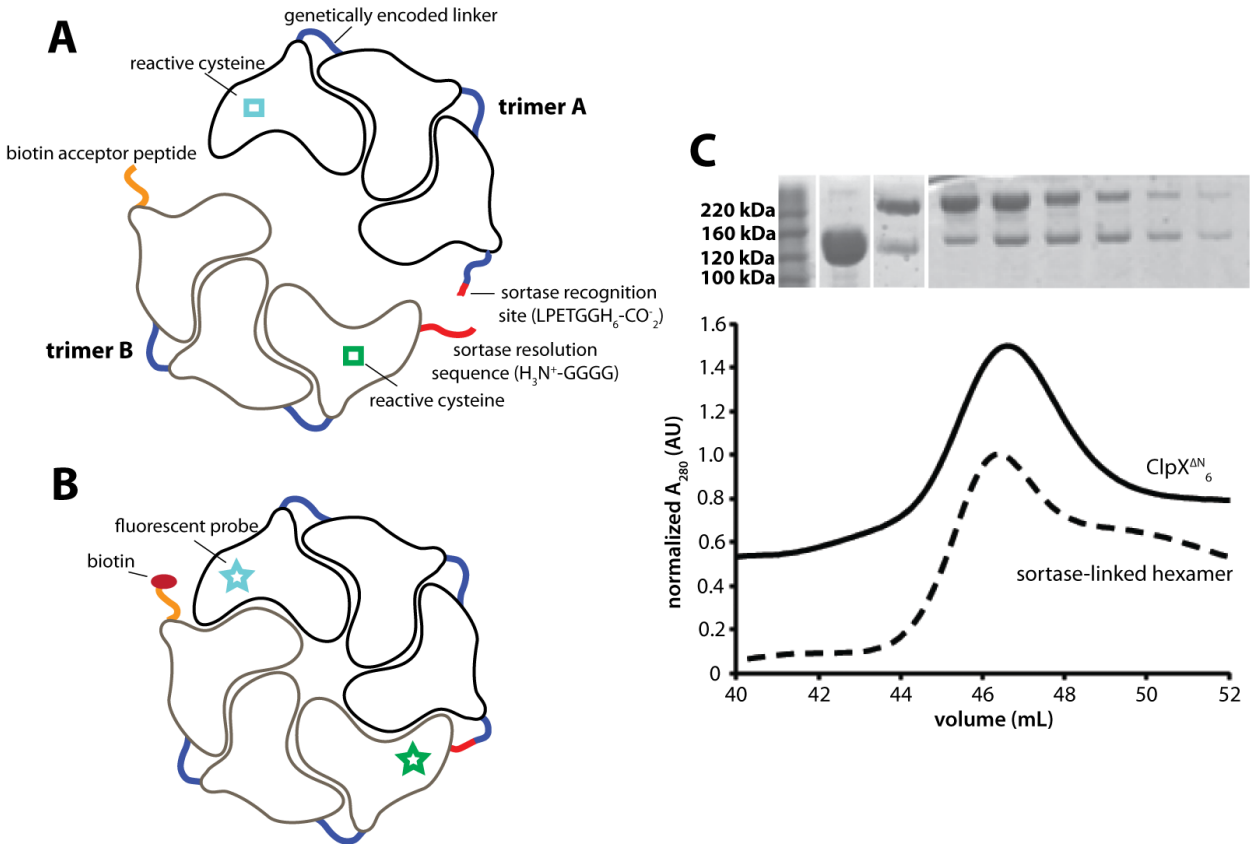

Figure B.1. Sortase linking of ClpX$^{\Delta N}$ trimers to form a single-chain hexamer. (A) Schematic of single-chain ClpX$^{\Delta N}$ trimers. The terminal subunit of **trimer A** contains a sortase recognition sequence. The N-terminal subunit of **trimer B** contains a sortase resolution sequence and the C-terminal subunit encodes a biotin acceptor peptide. Each trimer contains a single-reactive cysteine allowing modification using maleimide-conjugated dyes. **(B)** Linked hexamer after labeling, biotinylation, and sortase-mediated linkage. **(C)** Elution profile from a Superdex S200 gel-filtration column (bottom) of either single-chain ClpX$^{\Delta N}{}_6$ or sortase-linked ClpX$^{\Delta N}{}_6$. SDS-PAGE analysis (top) of trimer A mixed with trimer B (lane 2), trimer A and trimer B treated with sortase (lane 3) and fractions obtained after size-exclusion chromatography of the sortase-linked hexamer (lanes 4-9).

In solution studies, we monitored the changes in fluorescence-resonance-energy transfer (FRET) between the two dyes as a function of nucleotide state. In the presence of ATPγS, a slowly-hydrolyzed ATP analogue, ClpX adopted a "high-efficiency" transfer state, whereas it adopted a "low-efficiency" state in the absence of nucleotide or in the presence of ADP (Figure B.2A). When ATP was added, ClpX converted to the high-efficiency state and then converted back to

the low-efficiency state as a function of time (Figure B.2B). At the concentrations of ClpX utilized (1 μM of each trimer), most molecules should be hexameric even in absence of nucleotide, but it is difficult to distinguish if the observed FRET changes result from conformational movements within a hexamer or because ATP facilitates formation of hexamers from contaminating, unlinked, trimeric precursors.

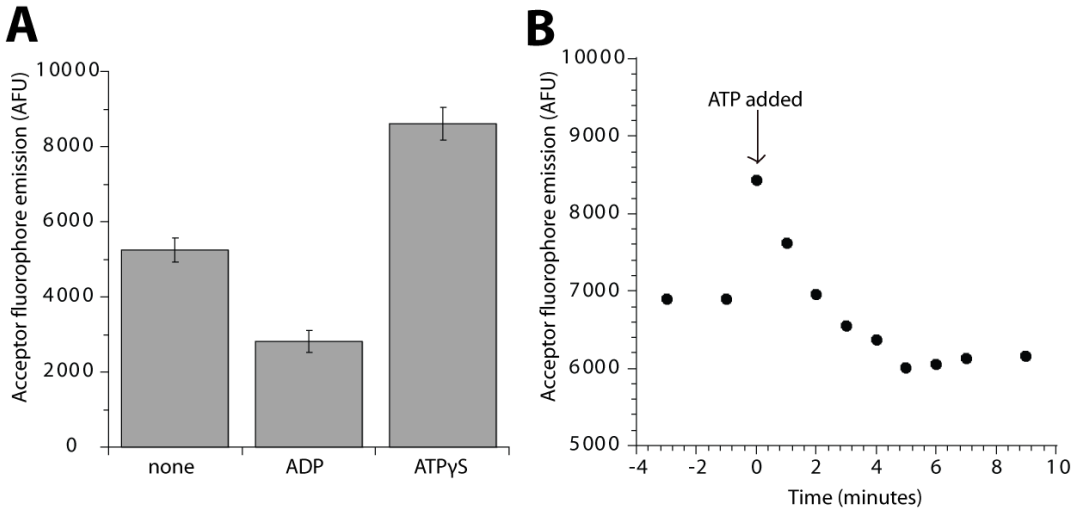

Figure B.2. Solution assays for nucleotide-dependent changes in FRET efficiency. (A) Sortase-linked $ClpX^{\Delta N}{}_6$ was mixed with ADP (4 mM) or ATPγS (4 mM). Fluorophores were excited with 495 nm light and emission was measured at 700 nm. **(B)** 2 mM ATP was added to sortase-linked $ClpX^{\Delta N}{}_6$ at the time indicated. At subsequent times, fluorescence emission was measured. Because these experiments were performed on different days with different enzyme stocks, the fluorescence units are not comparable.

In the assays described above, changes in energy-transfer efficiency are averaged over the entire population of molecules. To improve resolution of the dynamic conformational states adopted by individual molecules, we attached ClpX hexamers to a microscope slide via a biotin-streptavidin linkage as described previously (Shin *et al.*, 2009). Using a dual-view microscope developed by Yongdae Shin, we excited the donor florophore (Cy3) and simultaneously observed emission from both the donor fluorophore and the acceptor fluorophore (Alexa647). Very few FRET-paired $ClpX^{\Delta N}{}_6$ molecules were observed, presumably because of inefficient labeling at these

positions combined with incomplete sortase linkage (data not shown). For those rare molecules labeled with both fluorophores, however, we observe time-dependent changes in FRET efficiency in the presence of ADP, ATP, or ATPγS (representative traces shown in Figure B.3). Critically, the donor and acceptor emission intensity are anti-correlated and each dye photo-bleaches in a single event indicating that the observed fluorescence emission is a direct result of single-molecule FRET.

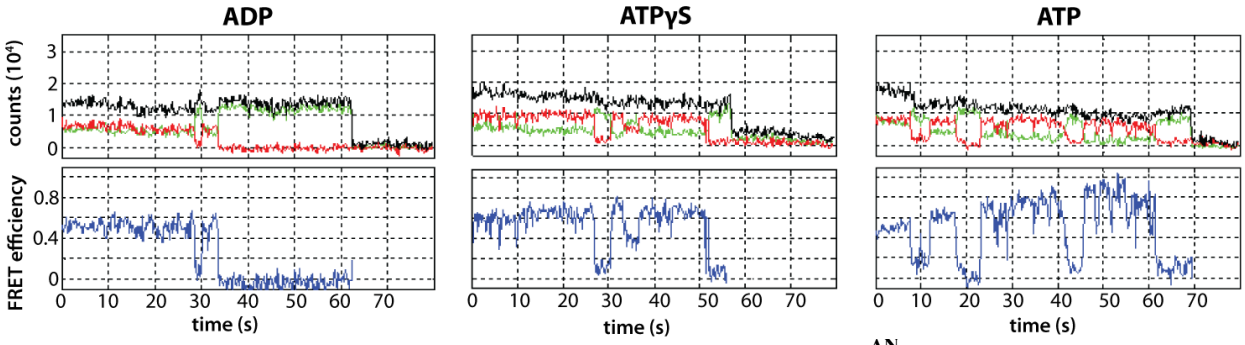

Figure B.3. Single-molecule FRET in sortase-linked ClpX$^{\Delta N}_{6}$. Donor fluorescence emission (green) is anti-correlated with acceptor fluorescence (red) resulting in a constant total emission (black). FRET efficiency (blue) is calculated using the ratio of acceptor and donor fluorescence.

It is difficult to distinguish if the FRET changes we observe result from nucleotide binding and hydrolysis or from nucleotide-independent interconversion of conformational states. Interestingly, the rate of fluctuation and the time-averaged occupancy of different FRET states appear to vary depending on the nucleotide present implying that some fraction of the changes we observe do depend on nucleotide. Because ClpX is an allosteric machine, ATP/ATPγS binding, hydrolysis or ADP dissociation in any subunit could theoretically give rise to the FRET changes we observe. In future studies, the use of mutant subunits defective in either nucleotide binding or hydrolysis might greatly simplify interpretation.

The single-molecule methods we have developed allow for the observation of rapid conformational changes at a level of resolution previously unattainable. By moving the fluorophores to different positions in ClpX, these techniques could be extended to the investigation of rigid-body motions in ClpX, the movements of various pore loops relative to a AAA+ domain, or the movements between identical amino acid positions in adjacent or non-adjacent subunits. Using a single label on ClpX, one might also monitor interaction with a labeled substrate or potentially with the peptidase, ClpP.

Methods

Plasmids and strains

$ClpX^{\Delta N}{}_3$ expression plasmids derived from pACYC were constructed with unique restriction sites between subunits, purification tags and sortase recognition sites allowing for facile manipulation of the individual components (Table B.1). Linked $ClpX^{\Delta N}$ trimers were expressed from a derivative of *E. coli* strain ER2566 (JD456) bearing a chromosomally-encoded phage lysozyme gene fused to a kanamycin selection marker (a gift of Sean Moore, University of Central Florida).

Protein expression, purification, biotinylation and labeling with fluorescent probes

Linked $ClpX^{\Delta N}$ trimer expression strains were grown at 37 °C in three liters of 1.5xYT broth (1.3 % tryptone, 0.75 % yeast extract, and 0.75 % NaCl, [pH 7.0]) to OD 0.5 and induced with 1 mM IPTG at room temperature. Four hours post-induction, cells were harvested by centrifugation, resuspended in 30 mL of degassed LB1 buffer (20 mM Hepes [pH 7.6], 400 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, and 10 mM 2-mercaptoethanol) and stored at -80 °C. After thawing at room temperature, 10 µL/mL chloroform was added to initiate lysis, and 1 mM PMSF was added to inhibit proteolysis. Cells were incubated with agitation at 4 °C for 30 min before adding 250 Units of benzonase and incubating an additional 30 min at 4 °C. The lysate was centrifuged at 8000 rpm in a Sorvall SA600 rotor for 40 min, and the supernatant was decanted and incubated for 5 min with 1 mL Ni^{2+}-NTA resin, which had been equilibrated with LB1 buffer. After washing with 15 column volumes (CV) of degassed LB1 buffer, resin was incubated with 5 CV of LB1 supplemented with 2 mM TCEP for 20 min. The resin was then

washed with 10 CV of LB2 buffer (20 mM Hepes [pH 7.6], 400 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol) before incubation with 2 CV of LB2 buffer supplemented with 100 µM maleimide-conjugated fluorophore (Cy3, Alexa-488, or Alexa-647), for 2 hours. The labeling reaction was quenched by washing the resin with 20 CV of LB1 buffer, before elution with 3 CV of LB1 buffer supplemented with 240 mM imidazole. Eluate was incubated for 6 hours at room temperature with 0.1 µM BirA, 1 mM biotin and 4 mM ATP in BB1 buffer (50 mM bicine [pH 8.3], 5 mM magnesium acetate) resulting in quantitative biotinylation (data not shown). The sample was then spin-concentrated and chromatographed on a Superdex S-200 gel filtration column equilibrated in GF1 buffer (50 mM Tris-HCl [pH 7.6], 1 mM dithiothreitol, 300 mM NaCl, 0.1 mM EDTA, and 10% glycerol). Fractions containing $ClpX^{\Delta N}{}_3$ were spin-concentrated and stored at -80 °C.

Sortase A from *Staphylococcus aureus* encoded on a plasmid bearing an IPTG-inducible T7 promoter (A generous gift of M. Popp, MIT) was overexpressed from strain JD530 (*E. coli* BL21 *rec*A-) and purified as described in Popp *et al*. (2007). BirA was expressed from strain JD490 (*E. coli* BL21 *rec*A$^-$) and purified as described in Shin *et al.* (2009).

Sortase linkage

Sortase A (1 µM) was incubated overnight at room temperature with stoichiometric amounts (15 µM) of GGGG-$ClpX^{\Delta N}{}_3$–bioAP-His_6 and $ClpX^{\Delta N}{}_3$–LPETGG-His_6 in SB buffer (100 mM Tris-HCl [pH 7.6], 1 mM dithiothreitol, 450 mM NaCl,10 mM $CaCl_2$, 0.1 mM EDTA, and 10% glycerol). Chromatography on a Superdex S-200 gel-filtration column resulted in clean separation of Sortase A from an un-resolved mixture of $ClpX^{\Delta N}{}_3$ and $ClpX^{\Delta N}{}_6$. Anion-exchange

chromatography on a monoQ 5/50 G column (Amersham Pharmacia) failed to resolve linked-hexamers from the trimeric precursor (data not shown).

Plasmid	*rs-1*	subunit 1	*rs-2*	subunit 2	*rs-3*	subunit 3	*rs-4*	subunit 4	*rs-5*	subunit 5	*rs-6*
pJD313	*NcoI*	**His_6**	*NdeI*	**$ClpX^{\Delta N}$ $C^{169}S$ $Y^{153}C$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SpeI*	-	*HinDIII*
pJD314	*NcoI*	**His_6**	*NdeI*	**$ClpX^{\Delta N}$ $C^{169}S$ $R^{200}C$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SpeI*	-	*HinDIII*
pJD315	*NcoI*	**FLAG**	*NdeI-SpeI*	**$ClpX^{\Delta N}$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SphI*	**bioAP-His_6-***	*HinDIII*
pJD356	*NdeI*	**$ClpX^{\Delta N}$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SpeI*	**LPETGG-His_6-***	*HinDIII*		
pJD338	*NdeI*	**$ClpX^{\Delta N}$ $C^{169}S$ $Y^{153}C$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SpeI*	**LPETGG-His_6-***	*HinDIII*		
pJD339	*NdeI*	**$ClpX^{\Delta N}$ $C^{169}S$ $R^{200}C$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SpeI*	**LPETGG-His_6-***	*HinDIII*		
pJD337	*NcoI*	**GGGG**	*NdeI*	**$ClpX^{\Delta N}$**	*KpnI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*BamHI*	**$ClpX^{\Delta N}$ $C^{169}S$**	*SphI*	**bioAP-His_6-***	*HinDIII*

Table B.1. Description of ClpX expression plasmids. Each $ClpX^{\Delta N}$ subunits encodes residues 61-424 of *E. coli* ClpX fused to a C-terminal linker (ASGAGGSEGGGSEGGTSGAT). Stop codons (*) are lacking form pJD313 and pJD314, allowing for restriction digestion (*NdeI* and *SpeI*) and ligation into pJD315 generating a single-chain hexamer. The protein product of pJD337 can be linked to that of pJD356, pJD338 or pJD339 *in vitro* using sortase resulting in the linker sequence (ASGAGGSEGGGSEGGTSGATTSLPETGGGGH) between the $ClpX^{\Delta N}$ trimers.

References

1. Antos, J. M., Popp, M. W., Ernst, R., Chew, G. L., Spooner, E., and Ploegh, H. L. 2009. A straight path to circular proteins. *J Biol Chem* **284**(23): 16028-16036.

2. Baker, T. A., and Sauer, R. T. 2006. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**(12): 647-653.

3. Erzberger, J. P., and Berger, J. M. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct* **35**93-114.

4. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A., and Sauer, R. T. 2009. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**(4): 744-756.

5. Hersch, G. L., Burton, R. E., Bolon, D. N., Baker, T. A., and Sauer, R. T. 2005. Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: allosteric control of a protein machine. *Cell* **121**(7): 1017-1027.

6. Joshi, S. A., Hersch, G. L., Baker, T. A., and Sauer, R. T. 2004. Communication between ClpX and ClpP during substrate processing and degradation. *Nat Struct Mol Biol* **11**(5): 404-411.

7. Martin, A., Baker, T. A., and Sauer, R. T. 2005. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature* **437**(7062): 1115-1120.

8. Martin, A., Baker, T. A., and Sauer, R. T. 2007. Distinct static and dynamic interactions control ATPase-peptidase communication in a AAA+ protease. *Mol Cell* **27**(1): 41-52.

9. Martin, A., Baker, T. A., and Sauer, R. T. 2008a. Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat Struct Mol Biol* **15**(2): 139-145.

10. Martin, A., Baker, T. A., and Sauer, R. T. 2008b. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat Struct Mol Biol* **15**(11): 1147-1151.

11. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E., and Ploegh, H. L. 2007. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**(11): 707-708.

12. Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., Hersch, G. L., Joshi, S. A., Kenniston, J. A., Levchenko, I., Neher, S. B., Oakes, E. S., Siddiqui, S. M., Wah, D. A., and Baker, T. A. 2004. Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* **119**(1): 9-18.

13. Shin, Y., Davis, J. H., Brau, R. R., Martin, A., Kenniston, J. A., Baker, T. A., Sauer, R. T., and Lang, M. J. 2009. Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc Natl Acad Sci U S A*

14. White, S. R., and Lauring, B. 2007. AAA+ ATPases: achieving diversity of function with conserved machinery. *Traffic* **8**(12): 1657-1667.

Appendix C

Measuring the Activity of BioBrick Promoters Using an *in vivo* Reference Standard

This work was published as Jason Kelly, Adam Rubin, Joseph Davis, Caroline Ajo-Franklin, John Cumbers, Michael Czar, Kim de Mora, Aaron Glieberman, Dileep Monie, and Drew Endy. 2009. *Journal of Biological Engineering* **3**(4).

JK, AR, and JD designed and conducted all experiments except those done at other laboratories. JK, KD, and CA designed the multi-institution experiments which were conducted by AR, CA, JC, MC, KD, AG, and DM. JK and DE conceived of the study and wrote the manuscript.

Abstract

Background

The engineering of many-component, synthetic biological systems is being made easier by the development of collections of reusable, standard biological parts. However, the complexity of biology makes it difficult to predict the extent to which such efforts will succeed. As a first practical example, the Registry of Standard Biological Parts started at MIT now maintains and distributes thousands of BioBrick™ standard biological parts. However, BioBrick parts are only standardized in terms of how individual parts are physically assembled into multi-component systems, and most parts remain uncharacterized. Standardized tools, techniques, and units of measurement are needed to facilitate the characterization and reuse of parts by independent researchers across many laboratories.

Results

We found that the absolute activity of BioBrick promoters varies across experimental conditions and measurement instruments. We choose one promoter (BBa_J23101) to serve as an *in vivo* reference standard for promoter activity. We demonstrated that, by measuring the activity of promoters relative to BBa_J23101, we could reduce variation in reported promoter activity due to differences in test conditions and measurement instruments by ~50%. We defined a Relative Promoter Unit (RPU) in order to report promoter characterization data in compatible units and developed a measurement kit so that researchers might more easily adopt RPU as a standard unit for reporting promoter activity. We distributed a set of test promoters to multiple labs and found good agreement in the reported relative activities of promoters so measured. We also

characterized the relative activities of a reference collection of BioBrick promoters in order to further support adoption of RPU-based measurement standards.

Conclusion

Relative activity measurements based on an *in vivo* reference standard enables improved measurement of promoter activity given variation in measurement conditions and instruments. These improvements are sufficient to begin to support the measurement of promoter activities across many laboratories. Additional *in vivo* reference standards for other types of biological functions would seem likely to have similar utility, and could thus improve research on the design, production, and reuse of standard biological parts.

Background

The engineering of many-component, synthetic biological systems is being made easier by the development of collections of reusable, standard biological parts (Arkin, 1999; Knight, 2002; Endy, 2003; Knight, 2003; Endy, 2005a; Voigt, 2006). Standardization of components has been instrumental in managing complexity in other engineering fields by helping engineers to reliably design and deploy systems comprised of combinations of parts (Texas-Instruments, 1988). However, it is an open question whether the overwhelming complexity of living systems will prevent biological engineers from fully achieving similar capabilities (below). To help answer this question, a Registry of Standard Biological Parts started at MIT now maintains and distributes thousands of BioBrick standard biological parts (www.partsregistry.org). BioBrick parts provide the first popular example of standard biological parts. However, BioBrick parts are currently only standardized in terms of how individual parts are assembled into multi-component systems (that is, "physical composition") (Knight, 2003; Canton *et al.*, 2008).

The utility of so-called standard biological parts would increase if the behavior of parts, both in isolation and in combination, were more predictable (that is, "functional composition") (Canton *et al.*, 2008). Prediction of behavior, in turn, depends on the initial designs and refinement of the parts themselves, the characterization of part functions, and the representation of part functions via abstract models (Guido *et al.*, 2006; Ajo-Franklin *et al.*, 2007; Rosenfeld *et al.*, 2007). Today, most BioBrick parts are directly derived from natural DNA sequences with only slight modifications to support at least one physical assembly standard, and many parts remain to be characterized. For example, fewer than 50 out of over 500 transcriptional promoters now available via the Registry have been characterized. Making matters worse, for the 50

characterized promoters, the methods of characterization are disparate and the resulting data incomparable. Shared and standardized approaches are needed in order to begin to address the challenge of characterizing promoters (and other types of standard biological parts) across a distributed community of biological engineers.

Making reliable and comparable *in vivo* measurements of biological parts has proven challenging. For example, five different efforts to measure the abundances of proteins in the yeast pheromone mating response system, one of the best characterized eukaryotic signaling systems, produced reports for the numbers per cell (abundances) of key system proteins that vary over a factor of ~12 (Ty Thompson, MIT, unpublished observation). Such examples suggest that measurement of the state or activity of biological systems, whether natural or engineered, may be unlike past engineering experiences, in that the minor differences in experimental conditions (relative to what can be readily controlled for, below) may cause large changes in the properties being measured. Even if conditions could be controlled for, it has proven challenging for researchers to develop and adopt standard approaches for characterizing biological parts. For example, an analysis of 80 published papers in which researchers used beta-galactosidase (β-gal) activity as a measure of gene expression found that at least six different protocols were used to measure enzyme activity (Serebriiskii and Golemis, 2000). In addition, nearly all activities were reported in "Miller units" even though in several cases there were differences in the substrates used to quantify enzymatic activity (CPRG or ONPG), the experimental conditions (pH and temperature for the assay), and even the absolute units of the Miller unit (nmol/min or μmol/min) (Miller, 1972). Differences in conditions such as using either CPRG or ONPG as a substrate for enzymatic assays lead to incompatible results (Eustice *et al.*, 1991), and thus Miller units should

generally not be considered comparable unless they have been calibrated against a common reference standard (Serebriiskii and Golemis, 2000).

The challenge of making reliable *in vivo* measurements of biological parts is further compounded by the need to measure many part properties indirectly via biological "measurement instruments" such as reporter proteins whose production can also be sensitive to experimental conditions. For example, β-gal activity can be used as an indirect measure of the behavior of a promoter, but the translation and activity of the β-gal protein is itself sensitive to experimental conditions such as temperature or choice of media. Since both the measurement instrument (β-gal) and the property being measured (promoter activity) are sensitive to measurement conditions (perhaps in differing ways) correcting for errors in measured promoter activity due to changes in conditions is more difficult. In theory such challenges could be addressed by strict adherence to standard measurement conditions. However, the adoption of standard measurement conditions in biological engineering is prevented by both practical constraints (as noted above) and also engineering constraints, such as culture or performance requirements that are specific to a particular biotechnology application. The overall situation is summed up nicely via the following quote: "There is no such thing as a standard (biological) component, because even a standard component works differently depending on the environment" (Pollack, 2006).

Although the characterization of standard biological parts is challenging, lessons from the measurement of other types of physical objects are worth considering. For example, one approach to controlling for variation in the measured property of an object in response to changing experimental conditions is to collect data from which to develop a model that describes

the relationship between the measured property and experimental conditions. As a specific example, models based on empirically determined coefficients of thermal expansion for common building materials (for example, Oak = $54*10^{-6}$/K at 20°C; Stainless Steel = $17.3*10^{-6}$/K at 20°C) are now sufficient to enable the reliable construction of structures across a range of environments. However, given the complexity of living matter, the relationships between the measured properties of biological parts and experimental conditions may be difficult to determine (at first). Thus, a second lesson worth considering is the measurement of relative (or ratio) properties rather than absolute characteristics. A relative measure is the ratio of the measurement of some aspect of the object being characterized in comparison to a standard reference object that is measured under the same conditions. For example, early methods for the diagnosis of osteoporosis made use of a measure of spinal cord deformity that was based on the ratio of various length measurements of vertebra within an individual patient (Barnett and Nordin, 1960). Doctors, by using a relative measurement for length, could account for variation in vertebra sizes between individuals of different body types or heights. As a second example, microarray experiments are frequently performed by co-hybridizing probes synthesized from both a reference and experimental RNA sample that have been labeled with different colors (Schena *et al.*, 1995); gene expression levels are then reported as the ratio of the experimental and reference intensities on each array spot. Thus, measurements made in relation to defined reference standards may provide an important first approach in characterizing the *in vivo* activity of biological parts and, over time, could enable the collection of empirical data sufficient to support the development of models that describe the effect of varying conditions on part properties.

Here, we characterized the *in vivo* activity of BioBrick promoters in order to evaluate if measuring relative activities might provide a useful initial framework for measuring the activity of standard biological parts across varying conditions. We chose to characterize promoters as a first example since they are ubiquitous in engineered biological systems, relatively well-understood, practically useful to biological engineers, and poorly characterized in the existing BioBrick collection (Weiss, 2001; Alper *et al.*, 2005). We developed a system that allows indirect measurement of the activity of promoters via observation of the synthesis rate of Green Fluorescent Protein (GFP) encoded by mRNA transcribed from each promoter. Our system requires the use of a quantitative model that allows promoter activity to be estimated from observed rates of GFP synthesis (below). Using this approach we demonstrate that normalizing the apparent absolute activity of a promoter to a defined reference standard promoter can help account for variation in conditions that would otherwise lead to significant differences in reported measurements.

Results

Definitions and models for absolute promoter activity

Our first step in characterizing standard biological promoters was to choose the property or properties whose measure would best support the reuse of such parts by biological engineers. Since the primary use of promoters is to initiate transcription, we chose the rate of transcription initiation as the property to be measured. We next chose the promoter clearance rate as the specific property that best describes transcription initiation; we refer to this property as "promoter activity" throughout. In turn, we defined promoter activity as the number of RNA polymerase molecules that pass by (or clear) the final base pair of the promoter and continue along DNA as an elongation complex. We report promoter activity using the generic unit of "Polymerases Per Second," or PoPS, in place of the more traditional "promoter clearance rate" because reporting activity in PoPS allows promoters to be directly compared to other genetic parts whose functioning impacts elongating polymerases, such as transcription terminators (Endy, 2005b). Other properties of promoters such as the binding constant of RNA polymerase to the DNA encoding the promoter, or the secondary structure of the DNA were not considered; while such properties may be relevant to researchers who are studying or engineering new promoters, our focus here was to support researchers who are characterizing or reusing existing promoters.

Directly measuring PoPS *in vivo* is challenging and, to our knowledge, has not yet been reported. However, by placing a promoter upstream of the coding sequence for green fluorescent protein (GFP) we could use the rate of GFP synthesis as an indirect measure of promoter activity. We

could then use a quantitative model to relate observed GFP synthesis rates to promoter activities reported as PoPS.

We adopted a previously described ordinary differential equation (ODE) model of GFP expression from a constitutive promoter to relate GFP synthesis rates per cell to promoter activities (Leveau and Lindow, 2001; Canton *et al.*, 2008). We evaluated this ODE model at steady-state (see supplemental information online, www.jbioleng.org/content/3/1/4/additional) in order to determine the rates of successful mRNA initiation events per DNA copy of each promoter ($PoPS^{SS}$) given observed GFP synthesis rates per cell (S_{cell}^{ss}):

$$\text{Absolute activity of promoter } (PoPS^{SS}) = \frac{\gamma_M (a + \gamma_I) S_{cell}^{SS}}{\rho a n} \quad (1)$$

where γ_M is the mRNA degradation rate, a is the GFP maturation rate, γ_I is the degradation rate of immature GFP, ρ is the translation rate of immature GFP from mRNA, and n is the number of copies of the promoter in the cell.

Variability due to equipment and conditions

We explored the sensitivity of our observable measure of promoter activity, GFP synthesis rates, to different measurement conditions and different measurement instruments. We estimated the per cell GFP synthesis rates of two promoters (BBa_J23101 and BBa_J23150) across seven different measurement conditions and instruments (Figure C.1A). We estimated GFP synthesis rate by reporting the change in arbitrary fluorescence units per absorbance over a 1-hour period in log phase growth (*Methods*). We varied the experimental conditions by changing the cell strain (TOP10 or W3110), carbon source (glucose or glycerol), and temperature (30C or 37C)

during growth. We found that the observed GFP synthesis rates were sensitive to the choice of strain (varying up to 2-fold) but insensitive to temperature and carbon source (within experimental error). We also varied the plasmid copy number and plasmid antibiotic resistance marker in order to explore how different genetic "measurement instruments" might impact the measured GFP synthesis rates. We found that GFP synthesis rates were sensitive to the plasmid copy number (varying up to 3-fold) and antibiotic resistance marker (varying up to 1.5-fold).

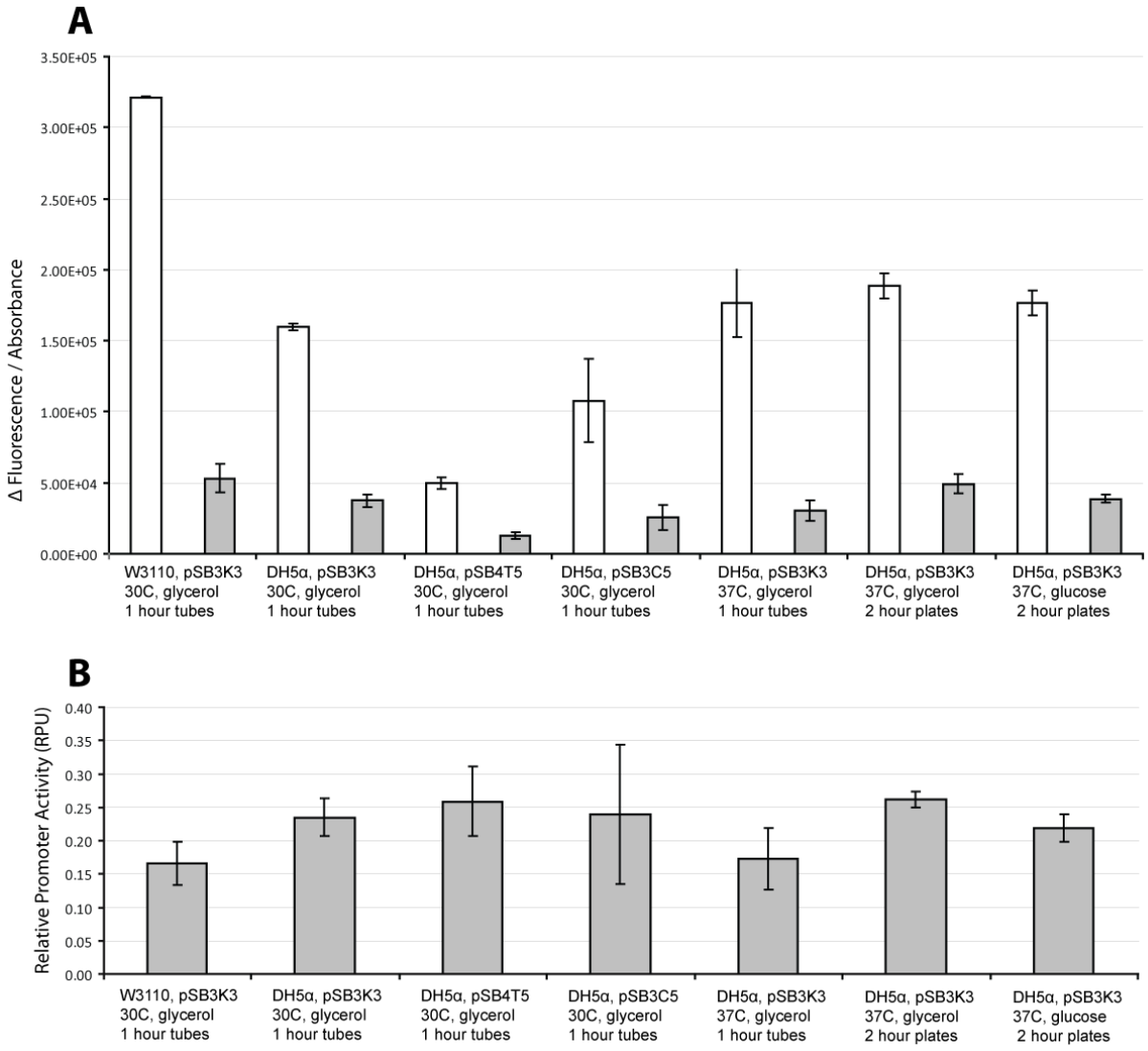

Figure C.1. Reference standards reduce variation in reported promoter activities under different measurement conditions. We measured the activity of 2 promoters, J23101 (white columns) and J23150 (grey columns) under seven different measurement conditions and measurement instruments. We varied the media, temperature, cell strain, and plasmid copy number of the promoter test construct. **(A)** To estimate the per cell GFP synthesis rate we reported the change in fluorescence over a 1 hour period in exponential phase divided by the average absorbance during this period. The coefficient of variation of the GFP synthesis rates across the seven measurement approaches was 49% for J23101 and 39% for J23150. **(B)** We used the same data and divided the GFP synthesis rate of J23150 (grey bars) by that of J23101 (white bars) in order to calculate the relative promoter activity of J23150 in RPUs. The coefficient of variation of the relative promoter activity across the seven measurement approaches was only 17% suggesting that the relative promoter activity is less sensitive to conditions then absolute activity measured by per cell GFP synthesis rate.

Definition and models for relative promoter activities

We noted that the activity of promoters (for example, J23101 and J23150) measured across different conditions or with different instruments was correlated (Figure C.1A). This correlation suggested that a measure of relative promoter activity might be less sensitive to varying experimental conditions or measurement instruments. To test this idea, we defined a new property – relative promoter activity – as the ratio of the absolute activity of a sample promoter, φ, relative to the absolute activity of a standard reference promoter, BBa_J23101, with both promoters measured under equivalent conditions and with the same measurement instrument. We reported relative promoter activities in a newly defined unit: Relative Promoter Units or RPUs. By our definition, a sample promoter with a relative activity of 1 RPU has activity equivalent to BBa_J23101.

An important consequence of considering a relative unit of measurement for reporting promoter activities is that many of the difficult-to-measure model parameters (Equation 1) that might change with changing environmental conditions can be cancelled when calculating relative promoter activities:

$$\text{Relative activity of promoter } \varphi \text{ (RPUs)} = \frac{PoPS_{\varphi}^{SS}}{\text{PoPS}_{\text{J23101}}^{\text{SS}}} \quad (2)$$

Thus, by substituting Eq. 1 into Eq. 2, we calculated promoter strengths in relative units of RPU and thereby eliminated many of the elementary parameters found in Eq. 1 via a cancellation of terms:

$$\text{Relative activity of promoter } \varphi \text{ (RPU)} = \frac{\dfrac{\gamma_{M,\varphi}(a_{\varphi}+\gamma_{I,\varphi})S^{SS}_{cell,\varphi}}{\rho_{\varphi}a_{\varphi}n_{\varphi}}}{\dfrac{\gamma_{M,J23101}(a_{J23101}+\gamma_{I,J23101})S^{SS}_{cell,J23101}}{\rho_{J23101}a_{J23101}n_{J23101}}} \quad (3)$$

We then made four additional assumptions that further simplified Eq. 3. First, we assumed that GFP expressed from either the test promoter φ or the reference standard promoter has an equivalent maturation rate ($a_{\varphi} = a_{J23101} = a$; given that the two promoters are measured under the same culture conditions). Second, since both promoters are carried on the same backbone plasmid, we assumed that each promoter is at the same average copy number ($n_{\varphi} = n_{J23101}$); while there are reported cases of promoter activity influencing the copy number of plasmids due to RNA polymerases transcribing through the plasmid origin of replication (Adams and Hatfield, 1984), a transcription terminator (BBa_B0015) downstream of our test construct's GFP coding sequence as well as the transcription terminators flanking the BioBrick cloning site (Shetty *et al.*, 2008) should largely prevent differences in promoter activity from impacting plasmid copy number. Third, since the promoters tested here have been standardized to have identical transcription initiation sites (predicted) and identical sequences downstream of the initiation site (see supplemental information online) we expected that each promoter produces the same mRNA sequence (Hawley and McClure, 1983). Since the transcribed mRNAs are expected to be identical we assumed that their mRNA degradation rates are equivalent ($\gamma_{M,\varphi} = \gamma_{M,J23101}$) and that the translation rates of immature GFP from mRNA are also equivalent ($\rho_{\varphi} = \rho_{J23101}$); while mRNA degradation is also a function of dilution due to cell growth, the dilution rate is negligible relative to typical rates of active mRNA degradation in *E. coli* (Bernstein *et al.*, 2002). Finally, we assumed that immature GFP is stable so that protein degradation is negligible compared to

dilution due to cell growth ($\gamma_{I,\varphi} = \mu_{\varphi}$ and $\gamma_{I,J23101} = \mu_{J23101}$, where μ is the cellular growth rate). Following the above assumptions, we simplified Eq. 3 to:

$$\text{Relative activity of promoter } \varphi \text{ (RPU)} = \frac{(a+\mu_{\varphi})S^{SS}_{cell,\varphi}}{(a+\mu_{J23101})S^{SS}_{cell,J23101}} \quad (4)$$

We further simplified Eq.4 by noting that:

$$\text{if } |\mu_{\varphi} - \mu_{J23101}| \ll a, \text{ then } \frac{(a+\mu_{\varphi})}{(a+\mu_{J23101})} \approx 1 \quad (5)$$

For example, we measured the growth rates of cells in our experiments to determine if the difference between the growth rates of cells containing the promoter test construct (μ_{ϕ}) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the maturation rate of GFP (that is, $|\mu_{\phi} - \mu_{J23101}| \ll a$). The cellular growth rates varied depending on the promoter being tested as well as on the experimental conditions: the fastest growth rate was observed in cells containing the BBa_J23113 promoter test construct grown in M9+glucose (μ = 0.9 hr^{-1}), and the slowest growth rate was observed in cells containing the BBa_R0040 promoter test construct grown in M9+glycerol (μ = 0.5 hr^{-1}). The maturation rate of the GFP variant used in the GFP reporter device (BBa_E0040) has been measured previously as a = 6.48 hr^{-1} (Canton *et al.*, 2008). Based on the worst-case assumption that cells containing the promoter test construct are the fastest growing cells (μ_{ϕ} = 0.9 hr^{-1}), and that cells containing the reference standard construct are the slowest growing cells (μ_{J23101} = 0.5 hr^{-1}) then:

$$\frac{(a+\mu_{\varphi})}{(a+\mu_{J23101})} = 1.06 \approx 1 \quad (6)$$

Therefore, we assumed that the difference between the growth rates of cells containing the promoter test construct (μ_{ϕ}) and cells containing the reference standard construct (μ_{J23101}) is negligible compared to the maturation rate of GFP, allowing Eq. 4 to be combined with Eq. 5 yielding:

$$\text{Relative activity of promoter } \varphi \text{ (RPU)} = \frac{S_{cell,\varphi}^{SS}}{S_{cell,J23101}^{SS}} \quad (7)$$

Taken together, by reporting promoter activity relative to a reference standard promoter (BBa_J23101) and choosing promoters with identical transcription initiation sites and identical sequences downstream of the initiation sites, researchers can quickly report measured relative promoter activities in compatible units without having to independently measure GFP maturation rates, mRNA degradation rates, protein production rates, or plasmid copy number for their specific experimental setup. We detail the precise numerical sensitivity of the quantitative model to each of the above assumptions in supplemental information available online.

We converted GFP synthesis rates measured across 7 different conditions and instruments (Figure C.1A; Coefficient of variation (CV) of the measurements is 39.1%) to relative promoter activity in RPUs (Figure C.1B; CV of the measurements is 17.5%). We noted that the coefficient of variation in promoter activity was reduced by approximately half when converted to RPUs from GFP synthesis rates. This reduction in variation suggests that relative promoter activity might be a useful property for characterizing promoters. However, care should be taken to note that while relative promoter activities remain fairly constant across some range of conditions, absolute promoter activities vary widely across these same conditions. Stated differently, a promoter that has an equivalent relative activity across multiple conditions might not produce

equal absolute activity (as measured in PoPS) across the same conditions (please see *Discussion*).

Laboratory-laboratory variation

Our initial success in characterizing relative promoter activity across different conditions and measurement instruments suggested a practical test. Specifically, we sought to determine whether multiple laboratories could work together to characterize promoters. To do this, we distributed a "reference promoter set" comprised of four strains, each containing one promoter test construct (BBa_J23113, BBa_J23150, BBa_J23151, or BBa_J23102) to researchers in six independent laboratories. Each researcher then measured the activity of the four promoters following a five-step procedure: (1) three independent cultures were grown from single colonies for each of the four promoters, (2) cells were collected in exponential phase, (3) GFP concentration per cell was measured using a flow cytometer, (4) flow cytometer data was gated based on forward and side scatter and the negative control, and (5) the geometric mean of the per cell fluorescence in the population was reported for each culture (Methods). We made no efforts to standardize the equipment (flow cytometers) or equipment settings beyond asking researchers to use typical settings for measuring GFP and by providing each lab with an example plot to guide gating of the flow cytometry data based on forward scatter, side scatter, and fluorescence (Kelly, 2008). As expected, there were slight differences in how the protocol was conducted in each laboratory, such as different culture conditions (rollers or shakers) and growth time. Since the measurements were reported in common units of RPUs we are able to compare the results of the interlaboratory promoter activity measurements directly (Figure C.2). The mean promoter activity measured by each lab is relatively consistent across all laboratories with less

than a 2-fold range of activities (min-max) across all measured promoters (BBa_J23150: 0.14–0.23 RPU; BBa_J23150: 0.38–0.61 RPU; BBa_J23103: 0.77–0.96 RPU). The activity of the weakest promoter, BBa_J23113, was equivalent to the negative control within error for all but one of the laboratories. These results suggest that relative promoter activity is an effective metric for making comparable measurements across multiple laboratories. Finally, we determined the coefficients of variation of the measured promoter activities across all labs to be 17.2%, 17.1%, and 8.5% for BBa_J23150, BBa_J23151, and BBa_J23101, respectively, setting a baseline for future improvements to the measurement kit and methods.

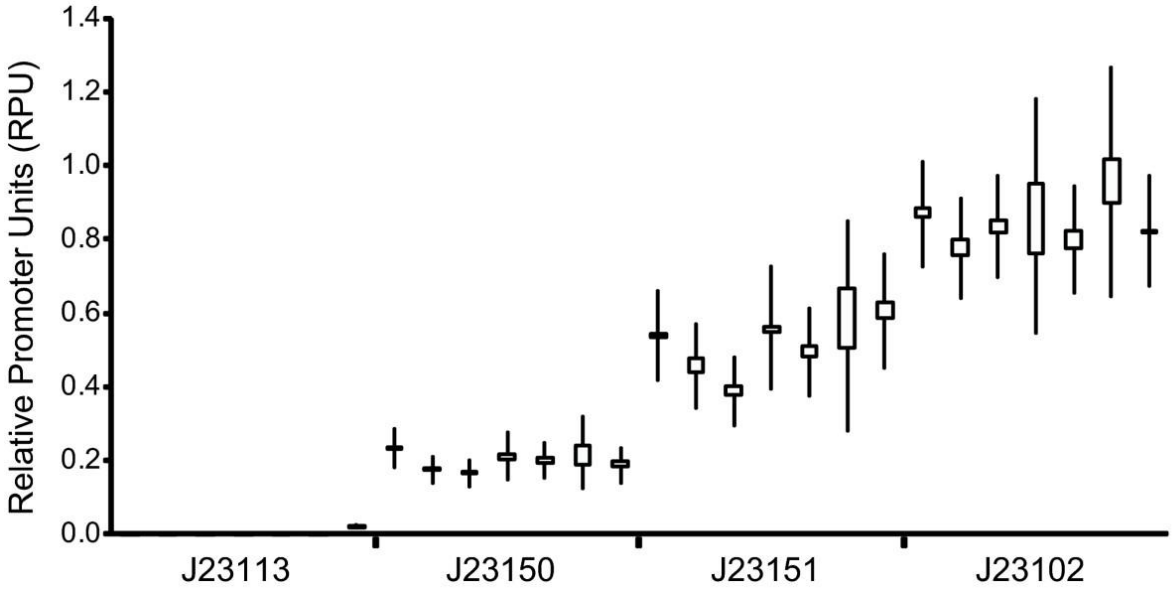

Figure C.2. Reference standards and units allow independent labs to make sharable measurements. Each laboratory followed the same measurement procedure, measuring relative promoter activities based on GFP concentration measured via a flow cytometer. Measurements were taken in triplicate; the boxes show the highest and lowest measured relative promoter activities and the whiskers show the 95% confidence interval of the mean of the activities. The large range in the 95% confidence interval (extending beyond the highest and lowest measured activities) is partially a function of the small number of replicates (three) that were conducted by each laboratory. The activity of BBa_J23113 was equivalent to the negative control within error for all but one of the laboratories. The measured activities of the other three promoters were fairly consistent across laboratories with less than a 2-fold range of activities measured for each promoter across all labs (BBa_J23150: 0.14 – 0.23; BBa_J23150: 0.38 – 0.606; BBa_J23103: 0.77 – 0.96).

Community-based measurement of promoter collections

Given that many laboratories could coordinate their measurement of promoter activities, we sought to prepare tools that would facilitate the widespread adoption of relative promoter activity measurements. To do this we first measured the relative activities of a set of seven representative promoters obtained from the Registry of Standard Biological Parts (Figure C.3) (www.parstregistry.org). These promoters included members of a constitutive promoter library (BBa_J23100 – BBa_J23119, constructed by JC Anderson) as well as the commonly used Tet repressor (BBa_R0040) and Lac repressor (BBa_R0011) regulated promoters 9Lutz and Bujard, 1997). The regulated promoters were tested in the absence of their cognate repressor proteins. Such libraries of characterized promoters have been shown to be valuable to researchers for tuning biochemical networks to optimize the synthesis of products of interest (Alper *et al.*, 2005; Basu *et al.*, 2005). We measured the relative promoter activities by calculating the steady-state GFP synthesis rates (*Methods*) and converting these rates to RPUs. Nine independent clones were characterized across three separate experimental runs for each promoter tested. The promoters ranged in activity from 0.026 ± 0.003 to 1.45 ± 0.095 RPUs (uncertainties represent 95% confidence interval of the mean). The GFP expression level from one promoter (BBa_J23113) was statistically equivalent within measurement error to the expression level of the negative control (TOP10).

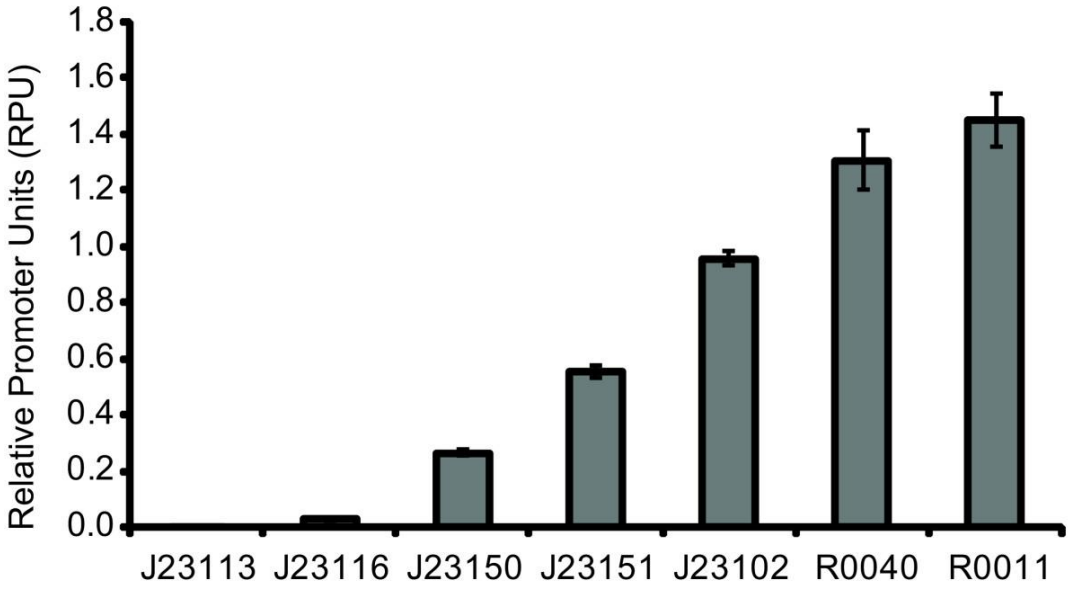

Figure C.3. Promoter collections can be readily characterized via Relative Promoter Units (RPUs). The five promoters labeled J23### are from a constitutive promoter library and R0040 and R0011 are tet- and lac-repressible promoters, respectively. The activity of the promoters was measured in relative promoter units (RPUs). This collection of promoter may itself be useful for tuning gene expression in engineered systems. The error bars represent the 95% confidence interval of the mean based on nine replicates.

To further support community-based standardized measurement of promoter activities, we developed a first generation measurement kit for characterizing the relative activity of BioBrick promoters in RPUs. Our overarching objective for the kits was to enable independent researchers to make comparable measurements of relative promoter activity in standard units. We developed instructions and a parts list for the promoter measurement kit (see supplemental information online). The promoter measurement kit contains measurement "instruments" such as a green fluorescent protein (GFP) reporter device (BBa_E0240) and backbone plasmid (pSB3K3), as well as a recommended *E. coli* strain (TOP10). The reference promoter (BBa_J23101) was inserted upstream of the GFP reporter device (BBa_E0240) and included in the kit as the reference standard construct (BBa_I20260). In order to measure the activity of a user-specified promoter, kit users assemble the user-specified promoter upstream of the GFP reporter device and insert this combined part into the backbone plasmid to form the promoter test construct. The

process for inserting a promoter upstream of the GFP reporter device is based on three-antibiotic BioBrick standard assembly (Shetty *et al.*, 2008), and is outlined in the instructions included with the kit (see supplemental file online).

Discussion

We found the absolute activity of promoters to vary under different experimental conditions and when using different measurement instruments. We chose a promoter (BBa_J23101) to serve as a reference standard and demonstrated that by measuring relative promoter activity (activity of a sample promoter divided by activity of the reference standard promoter BBa_J23101, measured under the same conditions) we could reduce reported variation in measured promoter activity across differing experimental conditions and equipment. We defined the Relative Promoter Unit (RPU) in order to enable researchers to report promoter characterization results in compatible units, and developed a measurement kit in order to more easily allow researchers to adopt an RPU-based measurement approach. We distributed a test set of 4 promoters to 7 independent labs and found good agreement in the measured relative promoter activities across the test set. Finally, we characterized the relative promoter activity of 7 BioBrick promoters in order to bootstrap a collection of promoters measured according to our initial RPU reference standard.

Absolute and relative promoter activities

The absolute activity of a promoter is defined by the number of elongating polymerases per second (PoPS) exiting the promoter. The same promoter under different environmental conditions can have widely varying absolute promoter activities (Figure C.1A). Moreover, it is challenging to relate an indirect measure of absolute promoter activity, such as per cell GFP synthesis rates, to PoPS since any variation across conditions in the functioning of the genetic measurement instrument may not be well correlated with variation in promoter activity across these same conditions. In contrast, the relative activity of a promoter is defined as the ratio of the absolute activity of the promoter to the absolute activity of a reference standard promoter

measured under the same conditions and with the same measurement instrument. We found that the relative activity of a promoter will remain fairly constant across a practical range of conditions (Figure C.1B).

Relative promoter activity as reported in RPUs enables the ranking of the activities of promoters but does not, by itself, provide information about the absolute activity of the promoter under particular conditions. For example, if the relative activity of a sample promoter remains constant across several conditions, it is not necessarily the case that the promoter is producing equal numbers of mRNA transcripts in each condition, only that the promoter activity is remaining proportional to the reference standard across the conditions. Nonetheless, relative promoter activity is a valuable property to measure and report since promoters can be rank ordered by relative promoter activity even if they were characterized by different researchers or across different environments (see supplementary information online). Furthermore, if both the relative and absolute activities of a promoter are measured, then a conversion factor can be established that defines the relationship between relative and absolute promoter activity under specific measurement conditions. For example, the reference standard promoter J23101, which has a relative activity of 1 RPU, has an estimated absolute activity of ~0.03 PoPS under specific conditions (supplementary information online). Therefore, the promoter J23151, which has a relative activity of approximately 0.5 RPU under these same conditions, would be predicted to have an absolute activity of ~0.015 PoPS. As more absolute measurements are made, an expanded set of conversion factors (or functions) could be developed, allowing for improved estimates of absolute promoter activities across a wider range of measurement conditions.

Interlaboratory measurement of promoter activity

Measurement of *in vivo* promoter activities across laboratories is challenging due to the sensitivity of results to both experimental conditions and measurement instruments (Figure C.1), as well as the lack of shared reference standards (Serebriiskii and Golemis, 2000). We demonstrated here that by using a relative promoter activity to characterize promoter strength based on a shared reference standard, seven independent laboratories could make comparable measurements of three promoters (Figure C.2). We expect that future improvements to the recommended measurement techniques and measurement kit components could further reduce the variation in measurements across laboratories, with the results reported here providing a practical baseline for judging proposed improvements.

Measurement procedures

The reference promoter BBa_J23101 and relative promoter activity measured in RPUs provide a shared platform for researchers to evaluate different measurement procedures. We deliberately have not advocated a single measurement procedure, and there should be many acceptable procedures for characterizing promoters in RPUs, just as units for length or mass are not tied to a single measurement approach. The choice of the best measurement procedure will be influenced by the particular group making the measurements. For example, laboratories without access to equipment for capturing high-throughput single-cell measurements of fluorescence might opt for a bulk fluorescence measurement using a fluorimeter. Other groups might prefer to obtain single-cell measurements using quantitative microscopy or flow cytometry, or to capture a time-course of fluorescence measurements from a growing culture. As different measurement procedures are likely to have merits within different communities we expect that a number of procedures will be

established. As an example, for the community of undergraduate teams using the promoter measurement kit during the International Genetically Engineered Machines (iGEM) competition, we have suggested a measurement protocol that can be easily carried out by novice researchers and that only requires two absorbance and bulk fluorescence measurements (supplementary file online) iGEM, 2008.

Engineering with characterized promoters

We anticipate that both absolute and relative promoter measurements will be useful in engineering genetic networks, however it is unclear to what extent one approach might be preferred over the other (presuming both types of measurements could be readily obtained). For example, we can imagine an engineering design framework in which the absolute activities of promoters and other functional genetic elements are tracked explicitly, in order to support detailed modeling and analysis of issues such as the absolute "load" placed on a host cell via recombinant gene expression. Such an ability seems likely to become more important as many-component engineered biological systems are attempted (dozens to hundreds of gene products), in which the absolute expression levels of individual genes must be well managed and might be kept low compared to the high-expression, protein production systems typically used today. However, we can also imagine a competing or complementary engineering framework, based on the idea that cells already provide self-adapting and robust environments within which the absolute activities of genetic elements such as promoters are finely regulated by overall environmental or culture conditions (Bremer and Dennis, Neidhart, F. C., 1996). In such a framework, relative measurements of promoter activities may be both easier to obtain and more relevant. Many natural biological systems already follow this model, and are robust to the

absolute properties of components so long as the relative relationships between subparts are maintained. For example, developmental body plans may vary in size with individual organisms having different overall sizes, however the ratio of the sizes of individual bones or organs to overall body mass is often tightly maintained (Huxley, 1932).

Standard promoter definition

The promoters tested here were practically standardized to have identical transcription initiation sites (predicted) and identical sequences downstream of the initiation site (supplemental information online). Thus, we expect that the mRNA expressed by each of the tested promoters is identical to the mRNA produced by the reference standard promoter, and that we can cancel the mRNA degradation rate and translation rate of immature GFP from mRNA terms in simplifying the model relating GFP synthesis rates to RPUs. This simplification allows the activity of promoters to be reported in comparable units (RPUs) without needing to directly measure mRNA levels. Going forward, an expanded definition for standard BioBrick promoters could be developed in order to ensure that all promoters share the same transcription start position and a fixed sequence downstream of the transcription start site (that is, 5' mRNA UTR). All promoters that adhered to such a standard could then be reliably measured using the kits described here, or via future kits based on gene expression reporters that adhered to any new standard, without the need for promoter-specific mRNA quantitation.

Distribution, use, and improvement of standardized measurement kits

Shared measurement tools and reference standards become more useful as they are broadly adopted. To facilitate such adoption, the Registry of Standard Biological Parts now includes our

promoter measurement kit and reference standard in the annual distribution of BioBrick parts. We also created a website in order to support the reporting and sharing of promoter activity measurements (www.partsregistry.org/measurement). This website contains instructions for use of the kit and summarizes previously characterized promoters. Finally, to enable discussion of proposed improvements to the kit and reference standard, and also the development of new kits and reference standards, we are supporting an open discussion of technical standards in synthetic biology (http://biobricks.org/standards).

Conclusion

Standard tools, techniques, and units for measurement are needed for a distributed community of biological engineers to independently characterize and share biological parts. We have defined a shared unit for measuring relative promoter activity (Reference Promoter Units, RPUs) and demonstrated that relative promoter activity can address some of the challenges in measurement across labs due to varying experimental conditions and measurement instruments. We developed a first-generation measurement kit for BioBrick promoters, and are freely distributing the kit via the Registry of Standard Biological Parts. Having demonstrated the feasibility and ease of use of the kit, we hope to encourage a community of users to adopt and improve these measurement tools and reference standard in order to characterize promoters via a comparable and common unit, the RPU. We expect that the shared experiences of biological engineers using common measurement tools and standards will help to identify new engineering challenges in improving the reliability and reuse of standard biological parts.

Methods

Strains and media

All measurement experiments and cloning were performed in *E. coli* TOP10 (Invitrogen) or W3110. Supplemented M9 minimal medium (M9 salts, 1 mM thiamine hydrochloride, 0.2% casamino acids, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) was used for all measurement experiments with either glycerol (0.4%) or glucose (0.4%) added as a carbon source and kanamycin (20 μg/ml) antibiotic added where appropriate. All oligonucleotides were purchased from Invitrogen and DNA modifying enzymes were purchased from New England Biolabs.

Promoter measurement kit contents

Sequences for all BioBrick plasmids (denoted pSB***) and BioBrick parts (denoted BBa_####) are available through the Registry of Standard Biological Parts. pSB3K3 contains a p15A origin of replication (copy number 10–12) and the kanamycin resistance marker, pSB4T5 contains the pSC101 origin of replication (copy number ~5) and the tetracycline resistance marker, and pSB3C5 contains the p15A origin of replication and the chloramphenicol resistance marker. Physical copies of the plasmids and parts are also available from the Registry via the annual Registry parts distribution. The details of the promoter measurement kit contents are described in Supplementary Box 1 and Supplementary Table 1 (see supplementary information online). The sequences for the preparative primers used to amplify pSB3K3 to generate backbone plasmid are: TACTAGTAGCGGCCGCTGCAG (forward primer) and CTCTAGAAGCGGCCGCG AATTC (reverse primer).

Assembly of test constructs

We built promoters by annealing synthesized oligonucleotides. The oligonucleotides were ordered with 5' phosphates and designed to leave an EcoRI overhang on the 5' end and a SpeI overhand on the 3' end so they could be used in subsequent ligation reactions without an intermediate restriction digest step. We inserted seven promoters: BBa_J23113, BBa_J23116, BBa_J23150, BBa_J23151, BBa_J23102, BBa_R0040, and BBa_R0011 into the promoter test construct and transformed into TOP10 according to the process outlined in Supplementary Box 1 (see supplementary information online). We found the optimal concentration of DNA for each of the three components in the ligation reaction (pSB3K3, BBa_E0240 or BBa_I13401, and the test promoter) was approximately 10 ng per uL. More detailed protocols and troubleshooting can be found at www.partsregistry.org/measurement. In the process of construction we found mutations in two of the promoters that we attribute to errors in the synthesis of the oligonucleotides that were annealed to construct the promoters. The two promoters were functional so we included them as additional members of the collection (BBa_J23150 and BBa_J23151). The method of part assembly described here is based on the three-antibiotic BioBrick standard assembly method (Shetty *et al.*, 2008).

Assay of promoter collection

The protocol described here will be referred to as the "original" protocol throughout the methods section and describes the measurement procedure used to characterize the set of seven promoters (Figure C.3). For each promoter construct three 17 mm test tubes containing 5 ml of pre-warmed (37 °C) supplemented M9 medium with kanamycin (20 μg/ml) were inoculated from single colonies of TOP10-DH5α containing the promoter test construct on the pSB3K3 vector

backbone. Cultures were grown in 17 mm test tubes for approximately 20 hrs at 37 °C with spinning at 70 rpm. We then diluted the cultures 1:100 into 5 ml of pre-warmed fresh media and the cultures were grown for approximately four hours under the previous conditions (17 mm tubes, 37 °C, spinning at 70 rpm). After four hours, we measured the OD_{600} of a 500 μl aliquot from each culture on a WPA Biowave Spectrophotometer. Based on this OD_{600} measurement, the cultures were diluted to the same OD_{600} (0.07) in 5 ml of pre-warmed fresh media and grown for one hour at 37 °C. We then transferred three 200 μl aliquots from each culture into a flat-bottomed 96 well plate (Cellstar Uclear bottom, Greiner). We incubated the plate in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer) at 37 °C and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter, 0.1 second counting time through 5 mm of fluid), fluorescence measurements (485 nm excitation filter, 525 nm emission filter, 0.1 seconds, CW lamp energy 12901 units), and shaking (3 mm, linear, normal speed, 15 seconds).

Background absorbance was determined by measuring wells containing only media. Background fluorescence was determined at different ODs from the fluorescence of TOP10 cells without a GFP expressing vector (Kalir *et al.*, 2001). After background subtraction, time-series fluorescence (*F*) and absorbance (*ABS*) measurements were used to calculate the ratio of the rates of GFP synthesis for the promoter test construct and the reference standard construct. Measurements were taken from an approximately 30 min period in mid-exponential growth (Setty *et al.*, 2003) (see supplemental information online). For example:

$$\text{RPU} = \frac{S^{SS}_{cell,\varphi}}{S^{SS}_{cell,J23101}} = \frac{(dF_{\varphi}/dt)/ABS_{\varphi}}{(dF_{J23101}/dt)/ABS_{J23101}} \quad (9)$$

Since we are calculating a ratio of the GFP synthesis rates we do not need to determine each rate in absolute units of GFP per second per cell, rather we can use the background-subtracted fluorescence (*F*) that is proportional to the number of GFP molecules and the background-subtracted absorbance (*ABS*) that is proportional to the number of cells in the culture to calculate the ratio of GFP synthesis rates (Ronen *et al.*, 2002; Canton *et al.*, 2008).

Assay of different measurement conditions

We measured the promoter activity of two promoters (BBa_J23101 and BBa_J23150) under seven different measurement procedures. The first of the seven procedures was identical to the "original" protocol described above for measuring the 7-member promoter collection except it was conducted at 30 °C in the strain W3110 with pSB3K3 as the vector backbone for the promoter test construct. The second procedure was identical to the original except it was conducted at 30 °C. The third procedure was identical to the original except that it was conducted at 30 °C and pSB4T5 was used as the vector backbone. The fourth procedure was identical to the original except that it was conducted at 30 °C and used pSB3C5 as the vector backbone. The fifth procedure was identical to the original. The sixth procedure was identical to the original except that instead of the second dilution into tubes followed by 1 hour of growth, the cells were diluted into 96 well plates and incubated for two hours before we started taking measurements. The seventh procedure was identical to the sixth except glucose was used instead of glycerol as the carbon source.

Assay of inter-laboratory variability

We distributed a set of four promoters (BBa_J23113, BBa_J23150, BBa_J23151, and BBa_J23102) to six laboratories to take independent measurements of promoter activity. The protocol each lab conducted was identical to the original protocol described, except that the cells were harvested after the first 1:100 dilution and 4 hours of growth (there was no second dilution step). The cells were then spun down, resuspended in PBS, and the fluorescence per cell was measured using a flow cytometer. The measurement equipment used (cytometer model, laser, emission filter) varied between the laboratories (see supplementary information, Table 2).

For all other experiments we measured RPUs from the GFP synthesis rates as described in Equation 9. However, for the inter-laboratory experiments we are unable to measure the GFP synthesis rates because these rates require a time series to calculate (dG/dt in Eq. 9) and we only requested a single time point, however we can use this single time point to find the background-subtracted per cell fluorescence at steady-state ($[F]$). The flow cytometer measures fluorescence per cell directly, thus (F) is calculated by taking the geometric mean of the population fluorescence per cell. We related the per cell GFP concentration ($[G]$) to RPUs by using a model described previously (Leveau and Lindow, 2001) (derivation in supplementary information online):

$$\text{RPU} = \frac{[G]_{cell,\varphi}}{[G]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} = \frac{[F]_{cell,\varphi}}{[F]_{cell,J23101}} * \frac{\mu_{\varphi}}{\mu_{J23101}} \quad (10)$$

$\mu_{\varphi}/\mu_{J23101}$ is a correction term based on differences in growth rate. Changes in the growth rate effect per cell GFP accumulation since loss of GFP per cell is largely due to dilution. Since we are calculating a ratio of the per cell GFP concentrations we do not need to determine each rate

in absolute units of GFP molecules per cell, rather we can use the background-subtracted per cell fluorescence ([*F*]) that is proportional to the number of GFP molecules per cell to calculate the ratio of GFP concentrations.

After background correction, the per cell fluorescence ([*F*]) was determined for each promoter and activities in RPUs were calculated using Eq. 10. We applied the growth rates measured previously (supplementary information online, table 3) across all laboratories when calculating RPUs, rather than requesting individual laboratories to measure growth rates. This approximation likely increased the variability in the promoter activity measurements across laboratories, as growth rates vary between laboratories due to differences in culture conditions and media.

Acknowledgments

The authors would like to thank Nishant Baht and Karen Wong for building and testing alternate designs of the measurement kits. The authors would like to thank Barry Canton for helpful discussions. The authors would like to thank Julius B. Lucks, Lei Qi, Weston Whitaker and Gavin Price for taking measurements in the multi-laboratory study. JC and AG would like to acknowledge: Stephanie Terrizzi and the Flow Cytometry and Cell Sorter Facility for flow cytometry (FACSAria funded by NCCR SIG grant # 1S10RR021051-01A2). CA would like to recognize that the work was performed in part at the Molecular Foundry, Lawrence Berkeley National Laboratory, with support from the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Funding for this work was provided by the NSF Synthetic Biology Engineering Research Center. JK received additional support via an NSF Graduate Research Fellowship.

References

1. Adams, C. W., and Hatfield, G. W. 1984. Effects of promoter strengths and growth conditions on copy number of transcription-fusion vectors. *J Biol Chem* **259**(12): 7399-7403.

2. Ajo-Franklin, C. M., Drubin, D. A., Eskin, J. A., Gee, E. P., Landgraf, D., Phillips, I., and Silver, P. A. 2007. Rational design of memory in eukaryotic cells. *Genes Dev* **21**(18): 2271-2276.

3. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. 2005. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**(36): 12678-12683.

4. Arkin, A. (1999) in *DARPA White Paper*

5. Barnett, E., and Nordin, B. E. 1960. The radiological diagnosis of osteoporosis: a new approach. *Clin Radiol* **11**166-174.

6. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature* **434**(7037): 1130-1134.

7. Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S., and Cohen, S. N. 2002. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**(15): 9697-9702.

8. Bremer, H., and Dennis, P. (1996). *Modulation of chemical composition and other parameters of the cell by growth rate. Escherichia coli and Salmonella.* ASM Press, Washington DC.

9. Canton, B., Labno, A., and Endy, D. 2008. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26**(7): 787-793.

10. Endy, D. 2003. 2003 Synthetic Biology study.

11. Endy, D. 2005a. Foundations for engineering biology. *Nature* **438**(7067): 449-453.

12. Endy, D. 2005b. Adventures in synthetic biology. *Nature* **438**(1): 449-453.

13. Eustice, D. C., Feldman, P. A., Colberg-Poley, A. M., Buckery, R. M., and Neubauer, R. H. 1991. A sensitive method for the detection of beta-galactosidase in transfected mammalian cells. *Biotechniques* **11**(6): 739-740, 742-733.

14. Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., Elston, T. C., and Collins, J. J. 2006. A bottom-up approach to gene regulation. *Nature* **439**(7078): 856-860.

15. Hawley, D. K., and McClure, W. R. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res* **11**(8): 2237-2255.

16. Huxley, J. (1932) *Problems of Relative Growth*, Methuen, London, UK

17. iGEM. (2008)

18. Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M. G., and Alon, U. 2001. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**(5524): 2080-2083.

19. Kelly, J. R. (2008)

20. Knight, T. (2002)

21. Knight, T. (2003)

22. Leveau, J. H., and Lindow, S. E. 2001. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *J Bacteriol* **183**(23): 6752-6762.

23. Lutz, R., and Bujard, H. 1997. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**(6): 1203-1210.

24. Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, USA

25. Pollack, A. (2006) in *New York TImes*, New York City

26. Ronen, M., Rosenberg, R., Shraiman, B. I., and Alon, U. 2002. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci U S A* **99**(16): 10555-10560.

27. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. 2007. Accurate prediction of gene feedback circuit behavior from component properties. *Mol Syst Biol* **3**143.

28. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235): 467-470.

29. Serebriiskii, I. G., and Golemis, E. A. 2000. Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system. *Anal Biochem* **285**(1): 1-15.

30. Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U. 2003. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A* **100**(13): 7702-7707.

31. Shetty, R. P., Endy, D., and Knight, T. F., Jr. 2008. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* **2**5.

32. Texas-Instruments. (1988) *TTL logic data book: standard TTL, Schottky, low-power Schottky*, Texas Instruments, Dallas, Texas, USA

33. Voigt, C. A. 2006. Genetic parts to program bacteria. *Curr Opin Biotechnol* **17**(5): 548-557.