2008

Conditional Protein Alleles and Regulation of Bacterial Transcription: A Thesis in Two Acts

Edmund Ching Schwartz

Conditional Protein Alleles and Regulation of Bacterial Transcription: A Thesis in Two Acts

A Thesis Presented to the Faculty of

The Rockefeller University

In Partial Fulfillment of the Requirements for

The degree of Doctor of Philosophy

By

Edmund Ching Schwartz

June 2008

CONDITIONAL PROTEIN ALLELES AND REGULATION OF BACTERIAL TRANSCRIPTION: A THESIS IN TWO ACTS

Edmund Ching Schwartz, Ph.D.

The Rockefeller University 2008

The use of small molecule modulators of protein function (drugs) has gained in popularity due to the speed and precision with which they can work. Unfortunately, a small molecule with the required specificity and potency is not always available. We have developed two methods to generate small molecule sensitive protein alleles without the need to screen for a new drug. In the first, which we term conditional protein splicing (CPS), an intein is split into two inactive fragments which are activated by heterodimerization. By fragmenting the target protein and fusing the pieces to this split intein, we were able to control the target protein's activity through splicing. In the second, which we term Split Ubiquitin for the Rescue of protein Function (SURF), a protein is expressed with a destabilizing N-terminal sequence (degron) that leads to its degradation. The degron is removed through heterodimerization induced ubiquitin complementation and cleavage. This stabilizes the protein and allows a build up of activity. These techniques should allow the manipulation of a variety of proteins for the study of biological processes.

The bacterial RNA polymerase σ factors are required for promoter specific transcription initiation. The σ factors responsible for driving most gene

expression during log phase growth, known as the group 1 σ family are autoregulated by a poorly understood mechanism. Group 1 σ factors are unable to bind to DNA in the absence of core RNA polymerase. The N-terminal domain of group 1 σ factors, known as region 1.1 is responsible for the auto-inhibition of DNA binding capabilities. However, since in all crystal structures of bacterial RNA polymerase, region 1.1 is either absent or not resolved, it is not clear how region 1.1 works. We used NMR to determine the structure of region 1.1 and intramolecular crosslinking to provide evidence that region 1.1 acts by directly binding to the DNA binding domains of group 1 σ factors.

Acknowledgements

I would like to thank :

Tom Muir for his unwavering support and encouragement from failed cloning to glowing fly and from precipitated protein to high resolution structure.

Michael Young for letting a first year student from a chemistry lab wander in and slaughter thousands of his flies and for always helping me keep things, both scientifically and personally in perspective.

Seth Darst for patiently waiting through a myriad of false starts and incorrect structures and for getting me that beer even as the bartender was yelling at him.

Shai Shaham who, as the only one present whose research was not being discussed, provided a much needed voice of reason at my FAC meetings

Tom Kodadek for serving as my external committee member.

Matthew Pratt, without whom the SURF project would never have happened, the σ project would not have been as definitive, and lunch would have been a lot more boring.

Lino Saez for teaching me everything I know about flies, including how to spot a virgin.

The Muir, Young and Darst labs for creating a fantastic environment to work (and play) in.

My wife, Sarah Smith for keeping me going when I didn't think I could cut it, for keeping me humble when I thought I could and for always being there for me.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

Section 1.1 : Overview

Small molecule agonists or antagonists of protein function (drugs) have a number of attractive features for the study of biological processes [1]. The most prominent is the speed by which they work. Upon binding, small molecules can effect an almost immediate knock-out or knock-in of protein function. Thus, in cell culture the effects are often limited only by diffusion and in animal studies, limited only by the rate at which the drug is distributed to the tissue type of interest. This speed allows the detailed study of very rapid processes such as mitosis or cell signaling which are difficult to investigate using slower genetic methods.

Sometimes, it is not simply the presence or absence of a protein's activity that is required for a biological response, but a specific level of activity. Genetic techniques such as mutations, knock-outs or RNAi do not allow a titration of protein activity. Gene copy number is not linearly correlated with protein activity and RNAi gives a binary response. In contrast, by varying the drug dose, one can achieve anywhere from a minor perturbation to an almost complete knock-out.

Finally, the effects of small molecules are reversible or at least can be confined to a specific time window. This can simply require waiting for the drug to

be metabolized or excreted. In cell culture, one can achieve an even more rapid drug clearance by washing the cells with drug-free media. In some cases, an antidote small molecule is available that can attenuate the drug's effects without the need for a wash. This can allow one to, for instance, isolate the effects of a protein during a specific day (or potentially even hour) in development.

The biological processes that have been studied using small molecules are too numerous to mention here. The approach, whether going by the name chemical biology, pharmacology or simply biology has proven itself versatile and powerful. However, drugs with the required specificity and potency for biological studies are not available for all the proteins in the cell. If one requires the speed, dosing and reversibility of a small molecule inhibitor but one is not available for the protein in question, one is faced with a choice. One can try to find a new drug, whether by design, by screen or by both or one can modify the target protein to be responsive to a drug that is already available.

The most direct way to design a protein to be sensitive to a drug is what is known as the "bump and hole" strategy (**Figure 1-1**). Sometimes, a potent inhibitor is available, but it is not specific enough, most commonly when the protein of interest is a member of a large protein family with diverse roles such as the kinases. Since the substrate binding pockets of all kinases are broadly similar, an inhibitor designed against one kinase will often inhibit others leading to off target effects that make interpretation of the results difficult. However, by making mutation(s) in the binding pocket of the target kinase and compensating

"mutations" in the drug, one can achieve a specific knock-down of only the mutated kinase, even in the presence of multiple other kinases. The Shokat group has used this approach to develop specific inhibition of a variety of modified kinases [1-4]. The bump and hole approach was first used on GTPases [5, 6] and has also been used on kinesins [7] and nuclear hormone receptors [8] among other proteins. The bump and hole approach will be described in more detail in section 1.2.

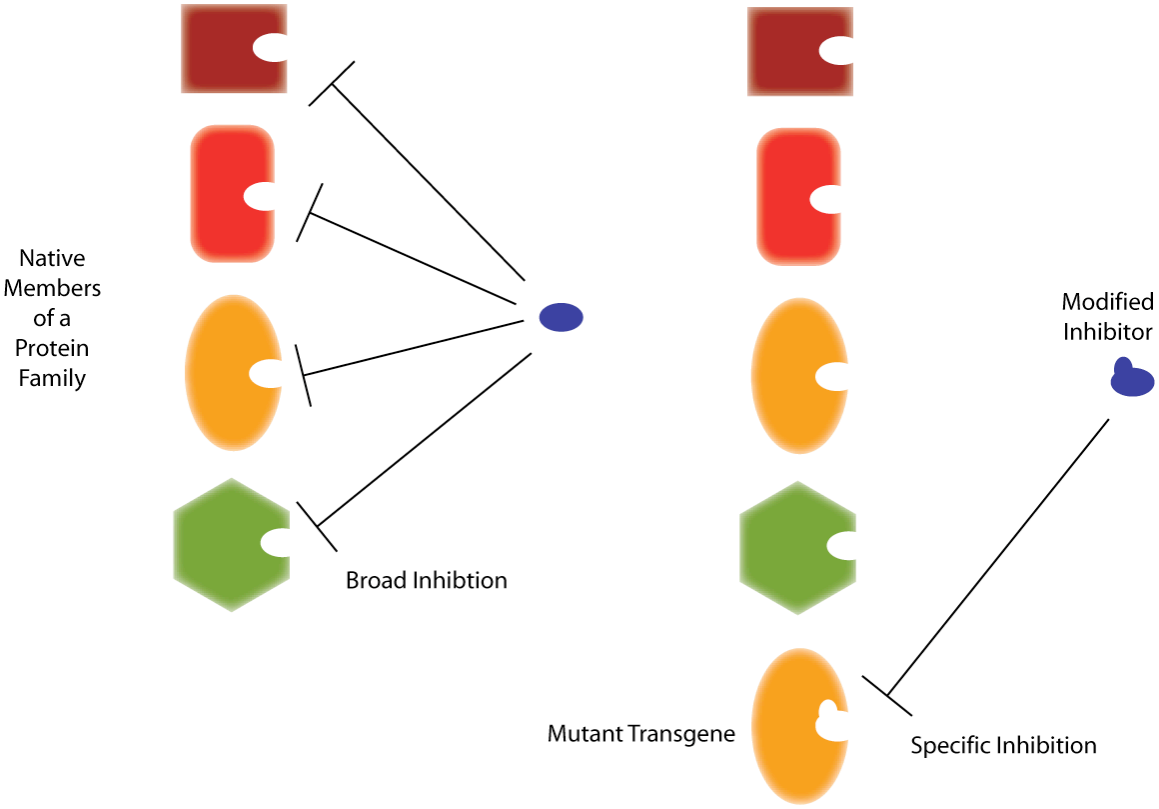


Figure 1-1 : The Bump and hole approach (Section 1.2). A small molecule inhibitor will often have off target effects that arise from binding to related proteins (left). By simultaneously mutating the target protein and chemically modifying the small molecule inhibitor, one can generate a drug specific allele (bottom right) even in the presence of similar proteins and the wild type version of the protein (right).

Figure 1-2 : Chemical Inducers of Dimerization for the control of protein function (Section 1.3). Chemical Inducers of Dimerization (CIDs) can be used to regulate protein function in a variety of ways. Proteins that are naturally regulated by dimerization, such as membrane receptors can be activated through forced dimerization [9-11] (Top). Proteins with two independent domains, such as transcription factors can be activated by bringing the two domains together [12] (Bottom Left). Other proteins can be controlled through dimerization through cellular localization [13-15] (Bottom Right). Finally, some proteins can be split into inactive fragments that complement and regain activity upon dimerization [16-27] (Middle).

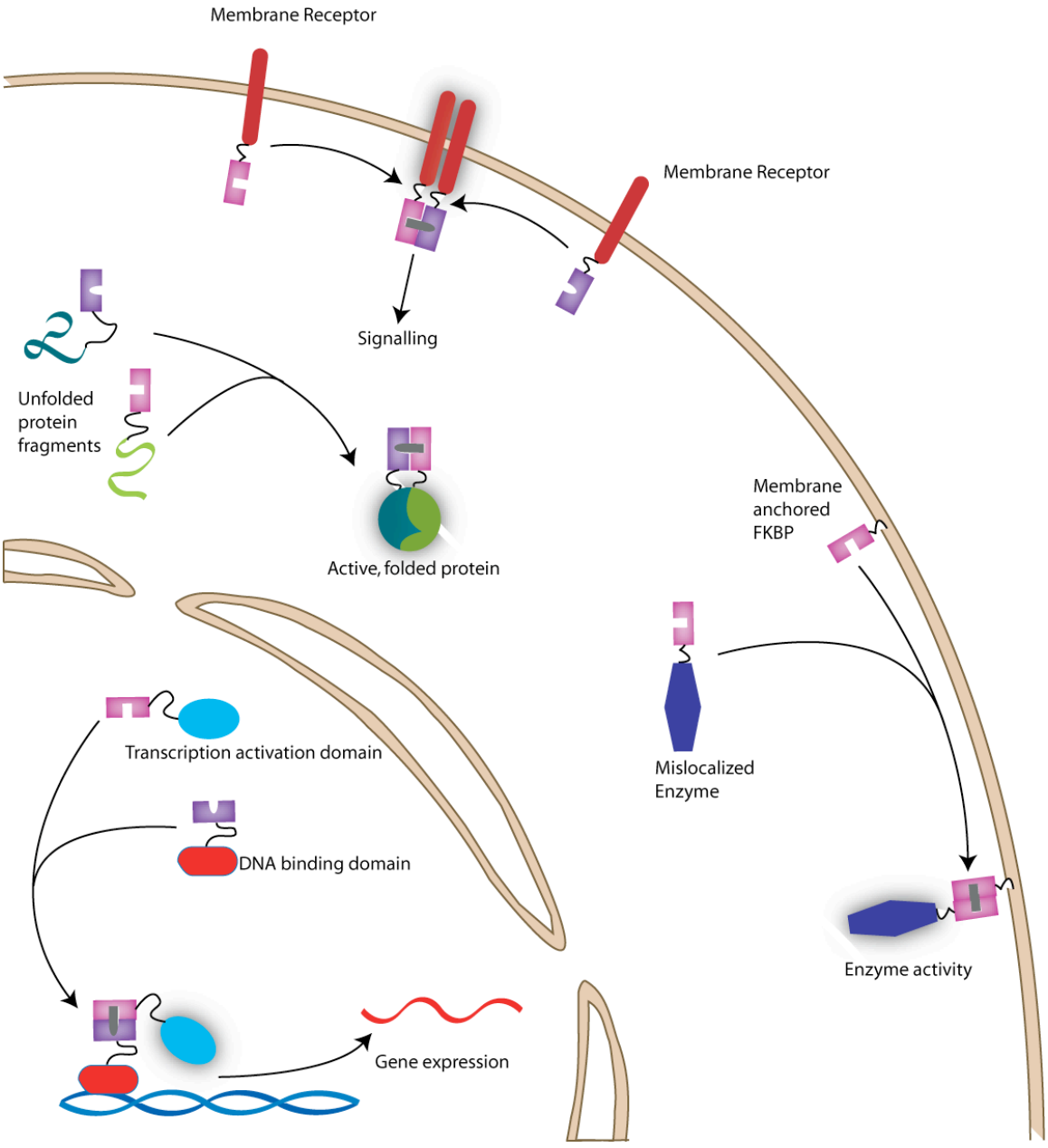


Figure 1-2 : Chemical Inducers of Dimerization for the control of protein function (Section 1.3).

Another popular means to sensitive a protein to an already available drug is the use of Chemical Inducers of Dimerization (CID, **Figure 1-2**). Protein domains that form dimers in the presence of a small molecule can be fused to the target protein or fragments of the target protein to gain control over its function. This can most easily be applied to proteins whose function is already regulated by its oligomeric state. Crabtree, Schreiber and coworkers were able to regulate the T cell antigen receptor complex through fusion to the FKBP domain which dimerizes in the presence of the molecule FK1012 [9]. This approach was also successfully applied to the Fas death receptor [28].

Other proteins can be divided into two domains that require one another for proper cell function. Fusion of these domains to dimerization domains allows regulation of the target proteins function with the use of a CID. A variety of drug sensitive transcription factors have been generated in this way by separation into DNA binding domain and a transcriptional regulation domain [12].

Sometimes CIDs can be used to regulate an enzyme's activity simply by controlling its localization within a cell. For instance, Src kinase requires membrane localization for enzymatic activity. Again, Crabtree, Schreiber and coworkers were able to regulate the activy of Src kinase by using FKBP and FK1012 to control the association of the enzymatic domain of Src with a myristoylated, membrane anchored FKBP [13]. The use of CIDs to manipulate gene function will be discussed in more detail in section 1.3.

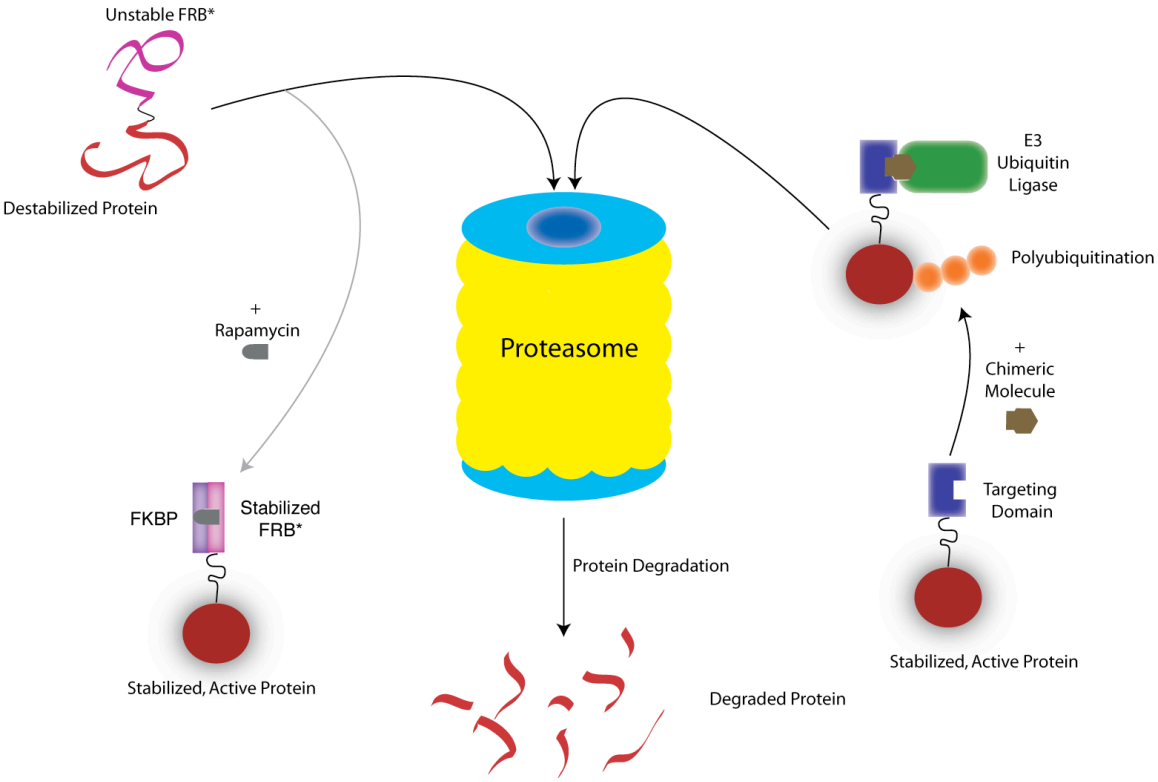


Figure 1-3 : Small molecule control of protein stability (Section 1.4). A protein's activity can be regulated by targeting it for or rescuing it from degradation. A protein can be targeted for degradation by recruiting an E3 ubiquitin ligase with an engineered chimeric molecule (Right). Alternatively, by fusion to an unstable FRB*, a protein can be destabilized and targeted to the proteasome. Binding to rapamycin and cellular FKBP stabilizes the protein and allows the build up of protein activity (Left).

As an alternative to directly controlling a protein's activity with dimerization, some more recent techniques control a protein's stability (and thus, indirectly, its activity) with CIDs as both a knock-in and a knock-out (**Figure 1-3**). The Crews and Deshaies labs have developed a technique known as PROteolysis TArgeting Chimeric moleculeS (PROTACS) in which E3 ubiquitin ligases are recruited to a target protein through the addition of molecular

chimeras of a peptide that binds to a ubiquitin ligase and either the protein target's substrate or a molecule that binds to a fused targeting domain. This results in the ubiquitination and subsequent degradation of the target protein, allowing a chemical knockout [29]. Crabtree and Wandless have developed a destabilized FRB domain (FRB*) that, when expressed as a fusion with a target protein results in the degradation of the target protein [30]. Binding to rapamycin and cellular FKBP stablilizes both the FRB* and the fused protein allowing it to build up and gain activity, resulting in a chemical knock-in. This technique has been applied to a variety of proteins and has even been shown to work in mice [31].

A protein's stability can also be controlled through fusion to a single domain rather than through controlled dimerization. The Wandless group has described a destabilized FKBP* that works in much the same manner as the FRB* domain works. However, by the use of an engineered FKBP ligand, the FKBP* can be stabilized without dimerization, resulting in a smaller, potentially less invasive "tag" on the affected protein [32]. Techniques that exploit protein stability for control of activity are discussed in section 1.4.

Fusion with a hormone receptor ligand binding domain (HR-LBD) is another method by which to build small molecule sensitivity into a target protein as this can result in activity that is responsive to hormones. Conditional CRE recombinases have been generated in this manner, allowing genetic knock-in and out in a tamoxifen controlled manner [33-35]. The Liu lab has developed a

tamoxifen controlled intein/ER fusion [36, 37]. By insertion of this conditional intein into a target protein, the target protein's activity can be indirectly controlled through protein splicing. HR-LBD fusions will be discussed further in section 1.5.

The engineering of a target protein to make it responsive to a known drug is known as generating a conditional protein allele since the modified protein is generally genetically encoded and expressed in cells or animals using standard genetic techniques. The ideal conditional protein allele is one whose "active" form is nearly identical to the wild type protein in both structure and activity. The ideal means to generate conditional protein alleles is one that is fully general in that it should be applicable to virtually any target protein with a minimum of optimization. Unfortunately, the methods described above, while powerful, all either suffer from a lack of general utility or do not generate a wild type protein.

The bump and hole approach requires that an inhibitor can be chemically modified and that the binding pocket of the target enzyme is amenable to mutation. Not all proteins can be controlled by dimerization, and when they can be, the resulting protein is present as a fusion with the dimerization domains rather than in its native form. Although it appears that nearly any protein's stability can be regulated by fusion to FRB* or FKBP*, once again, the rescued protein is a fusion protein. PROTACS either requires that a ligand for the target protein be available and amenable to chimerization or that the protein be expressed as a fusion with a targeting domain. Fusion with an HR-LBD does not always result in a conditional allele and, once again, results in a tagged protein.

The conditional intein described by Liu eliminates the protein tag through splicing, but may not be general and has not been validated in animals at this time.

The majority of my thesis work has been spent on developing general conditional protein alleles that generate wild type (or nearly wild type) proteins. One approach that I have taken is to modify a technique known as conditional protein splicing (CPS). In CPS, a protein splicing element known as an intein is split into two inactive fragments which are activated by heterodimerization. By fragmenting the target protein and fusing the pieces to this split intein, I was able to control the target protein's activity through splicing. This work will be described in Chapter 2.

A second approach I have worked on relies on the control of the target protein's stability. During the development of the CPS approach, I found that 3 tandem copies of the FRB domain destabilizes a protein when present as an N-terminal fusion. This triple FRB can be used in a similar manner to the FRB* developed by Crabtree and coworkers. As a further evolution of the approach, an additional fusion with split ubiquitin was made. Varashovsky and coworkers have previously shown that split ubiquitin can be reconstituted by forced dimerization [16] resulting in C-terminal hydrolysis by a cellular ubiquitin hydrolase. The cleavage of ubiquitin removes the destabilizing triple FRB from the target protein, simultaneously regenerating the wild type, untagged protein and rescuing it from degradation. This work will be described in Chapter 3.

In addition, during my thesis work, I have worked on another, unrelated project. The bacterial RNA polymerase sigma factors are required for promoter specific transcription initiation [38]. The sigma factors responsible for driving most gene expression during log phase growth, known as the Group 1 σ family are autoregulated by a poorly understood mechanism. Group 1 σ factors are unable to bind to DNA in the absence of core RNA polymerase. The N-terminal domain of group 1 σ factors , known as region 1.1 is responsible for the auto-inhibition of DNA binding capabilities [39, 40]. However, since in all crystal structures of bacterial RNA polymerase, region 1.1 is either absent [41, 42] or not resolved [43], it is not clear how region 1.1 works. I used NMR to determine the structure of region 1.1 and intramolecular crosslinking to provide evidence of an intrasteric mode of autoregulation. This work will be described in Chapter 4.

In the following sections, I will review the current techniques available for the generation of conditional protein alleles.

Section 1.2 : Bump and Hole

In 1987, Hwang and Miller described a mutation (D138N) in Elongation Factor Tu (EF-Tu) which reduced its affinity for GDP but allowed it to bind to xanthosine 5'-diphosphate [5]. EF-Tu aids in the binding of charged tRNAs to ribosomes during mRNA translation and binds GDP or GTP and a charged tRNA. In crystal structures, the D138 side chain hydrogen bonds to the 2-amino group of guanine. The D138N mutation abolishes this interaction by replacing a

hydrogen bond acceptor with a hydrogen bond donor. Xanthosine differs from guanosine in the substitution of a carboxyl group for the 2-amino group. Thus, while xanthosine cannot form a hydrogen bond with D138, it can form one with D138N allowing binding to the mutant but not the wild type protein. It was further shown that XTP could stimulate the formation of an Ef-Tu • ribosome • tRNA complex with the mutant protein but not with the wild type protein. Without setting out to, Hwang and Miller had in essence developed one of the first allele specific inhibitor – target pair.

Building on this finding, Powers and Walter deliberately made a similar mutation in another GTP binding protein, *E. coli* FtsY with the goal of isolating the effects of GTP binding FtsY as opposed to GTP binding to an FtsY interacting partner, Ffh [6]. Neither FtsY nor Ffh displays much GTPase activity in isolation but when mixed together, strong GTPase activity is observed. However, since both are capable of binding GTP, it was difficult to determine which, if not both of the two proteins is actually responsible for this activity. To answer this question, Powers and Walters generated an FtsY with a homologous mutation (D441N) to the one described by Hwang and Miller, creating an allele specific sensitivity to XTP. The FtsY(D441N) • Ffh complex exhibited GTPase activity only in the presence of both GTP and XTP, which was also hydrolyzed. This suggests that both FtsY and Ffh are acting as GTPases in complex and that they act to stimulate one another's activity.

Kapoor and Mitchison used the bump and hole approach to generate allele specific activators and inhibitors of kinesin [7]. Kinesins are ATPase motor proteins that aid in the movement of cellular components and in cytoskeletal organization. The addition of a bulky cyclopentyl group at the N6 position of ATP creates steric clash with the ATP binding pocket of kinesin. By mutating an arginine in the ATP binding pocket to an alanine, Kapoor and Mitchison were able to relieve this steric clash which resulted in the stimulation of the mutant's but not the wild type protein's motor activity by cyclopentyl-ATP. A non-hydrolyzable cyclopentyl-ATP was also able to inhibit the mutant protein, but not the wild type protein allowing both allele specific activation and inhibition.

Similarly, Shi and Koh showed that functionalizing estrogen with a carboxylic acid greatly reduced its ability to bind to the estrogen receptor (ER), likely from a combination of sterics and electrostatic repulsion. However, a compensatory E→A mutation in the ER LBD allowed binding to and stimulation by the modified estrogen [8]. The modified receptor/ligand pair could be used to drive gene expression without the concern of binding to native ER LBD.

The Shokat group has taken full advantage of the power of the bump and hole approach to study kinases. Highly specific agonists or antagonists of protein kinases are rare due to their structural similarity, especially in the ATP binding pocket. However, by adding a bulky subsituent to ATP and making a compensatory mutation in the target kinase, Shokat and colleagues have generated a variety of allele specific activators or, by making non-hydrolyzable

analogues, inhibitors [1]. These target/inhibitor pairs have proven to be powerful tools for the study of the roles and activities of kinases. In one report, a bump and hole CaMKII/ATP analogue pair was used in transgenic mice to study its role in memory formation [4]. Through selective inhibition of the sensitized trans-gene, it was determined that the activity level of CaMKII is crucial for the formation of long term memory only in the first week after "learning."

In another study, the Shokat group used the Bump and Hole approach to identify the subtrates of the cyclin-dependent kinase Cdk1 [3]. Determining which proteins were phosphorylated by a specific kinase is normally impossible in a cell lysate. However, by using a Cdk1 allele that was sensitized to be able to utilize N^6-(benzyl) ATP, Shokat and co-workers were able to isolate the effects of a single kinase in an entire cell lysate. Radiolabelled N^6-(benzyl) ATP was added to a cellular extract along with the sensitized Cdk1. Since only the sensitized Cdk1 was able to use N^6-(benzyl) ATP, only proteins that were substrates of Cdk1 were labeled with radioactive phosphate, allowing their isolation and identification.

The Bump and Hole approach has proven applicable to a variety of proteins but it is not fully general. Most obviously, it requires a molecule that binds to the wild type protein to use as a starting point. The binding pocket of the protein must also be amenable to mutations without drastically affecting the protein's activity. Furthermore, especially for animal studies, the modification of the drug may hamper its pharmokinetic properties. As a more practical concern,

many biology labs lack the synthetic capabilities to synthesize the drug analogs and many chemistry labs lack the molecular biology knowledge to generate the mutant proteins, limiting the accessibility of the technique.

Section 1.3 Controlling Dimerization

A wide variety of proteins are naturally regulated through dimerization. The fact that the dimerization domain or surface is normally distinct from the functional domain of a protein has been exploited to experimentally control these proteins in two main ways. The first is the use of dominant negative proteins or peptides that contain the dimerization surface but not the functional domain. The dominant negative species can bind to the wild type monomers, preventing them from forming productive dimers. Dominant negative alleles of this variety can be found through mutagenesis screens, but some have also been rationally designed.

Perhaps the earliest report of this approach was by Laudano and Doolittle when, in 1980 they described the design and synthesis of peptides based on the N-terminus of fibrin that bound to and prevented the polymerization of fibrin [44]. Fibrin is formed by thrombin proteolysis of fibrinogen. Fibrin monomers spontaneously polymerize and are crosslinked by Factor XIII to form blood clots. The peptides synthesized by Laudano and Doolittle masked the binding surface

of a fibrin monomer, capping it and disrupting the poymerization process, effectively halting clotting.

The platelet derived growth factor (PDGF) receptor forms dimers upon binding to PDGF. The intracellular tyrosine kinase domains then cross-phosphorylate one another leading to receptor activation and signaling. Ueno, *et. al.* generated a dominant negative mutant simply by removing the intracellular domain [45]. Since a PDGF monomer is not able to phosphorylate itself, this truncated PDGF receptor was able to knock-out wild-type PDGF receptor activity by forming non-productive dimers in which tyrosine phosphorylation could not occur. In a similar study, Kirschner and coworkers generated a dominant negative fibroblast growth factor (FGF) receptor by deletion of the intracellular tyrosine kinase domain [46]. Xenopus embryos expressing the truncated FGF receptor displayed developmental defects in mesoderm formation due to an inhibition of FGF signalling.

The design of dominant negative like alleles continues to grow in sophistication. In recent work, Degrado and coworkers used computer modeling to design transmembrane helical peptides that could tightly bind cellular transmembrane helices and disrupt their interactions with other proteins [47]. These peptides were used to generate a dominant negative ToxR in bacteria by preventing its productive dimerization. They could also be used to disrupt the intramembrane interaction between two integrin subunits, which in the case of

integrins actually results in their activation. So in some cases, a dominant "negative" can actually be an agonist rather than an antagonist.

The second widely used approach to control proteins by affecting their dimerization is the use of Chemical Inducers of Dimerization (CID). CIDs are small molecules that have two protein binding surfaces and are thus able to bind to two separate proteins simultaneously, inducing their dimerization (**Figure 1-4**). When the domains that bind to a CID are expressed as fusions with other proteins, the fusion proteins can also be induced to dimerize, allowing their function to be controlled in a variety of ways.

In work published in 1993, Schreiber, Crabtree and coworkers used chemically induced oligomerization to activate the T-cell antigen receptors [9]. Antigen receptors are naturally activated by binding to foreign antigens that are presented on the surface of antigen presenting cells. Binding leads to aggregation of the receptors and intracellular signaling events which govern the immune response. The ζ-chain of the antigen receptor was fused to a 3 copies of the FKBP domain. The FKBP domain dimerizes in the presence of the CID FK1012, so the addition of FK1012 led to the oligomerization and activation of the modified receptors.

A similar approach was used to investigate the role of the Fas receptor in apoptosis. The Fas receptor is involved in inducing apoptosis in B and T cells. Mice without functional Fas have excess, abnormal B and T cells and an autoimmune syndrome that is similar to lupus. A fusion of a myristoylated

intracellular Fas domain and FKBP was generated that was dimerized and activated with FK1012. By doing so, it was demonstrated that dimerization is necessary and sufficient to trigger apoptosis. The conditional Fas allele was also used in transgenic mice to show that Fas only plays an apoptotic role in CD4+ CD8+ cells [11, 28]. The authors followed up this work by designing and synthesizing a novel CID/dimerized protein pair, namely a covalent dimer of Cyclosporin A, which can dimerize cyclophilins [48]. This system was also able to induce apoptosis through dimerization of the Fas receptor.

The power of CIDs is that with a little imagination, they are not limited only to the regulation of proteins that are naturally controlled through dimerization. Many transcription factors have a modular architecture in which DNA recognition and binding and transcriptional activation or inhibition are carried out by separate, autonomous domains [49, 50]. In fact, some transcription factors do not even exhibit DNA binding on their own, rather, they bind to other proteins which either directly or indirectly bind to DNA. Often, all that is really required for transcriptional activation is to bring the activation domain close enough to the DNA to allow the recruitment of the RNA polymerase machinery. This feature has been exploited in two-hybrid screens to identify protein-protein interactions [51, 52], and has also allowed the widespread use of CIDs to activate transcription factors. A number of CID inducible transcription factors have been reported, in systems from bacteria to monkeys and with a variety of CIDs. Furthermore, because of the modular nature of transcription factors, the

activation domain of one transcription factor can be dimerized with the DNA binding domain of another and still be able to effect gene expression, further increasing the versatility of the system [12]. Looking at a cell through a high powered microscope, it is easy to get lost in the multitude of organelles and other substructures, small and large. Sometimes activating a protein is a simple matter of helping it find where it is supposed to be in the mazelike cellular milieu. The Src family of tyrosine kinases require membrane proximity for activity. Schreiber and Crabtree were able to use their FKBP-FK1012 system to selectively activate Src kinase by inducing the dimerization of a membrane anchored FKBP and an FKBP-Src that lacked a myristoylation sequence [13].

Cdc42 is a Rho GTP-binding protein that controls actin through the actions of the Wiskott-Aldrich syndrome protein (WASP). The Chavrier group generated conditional alleles of both Cdc42 and WASP by CID mediated recruitment to the plasma membrane [53]. A chimeric receptor of the transmembrane and extracellular portions of the CD25 receptor and two copies of FKBP was made. Cdc42 or WASP were expressed as fusions with FRB. The addition of rapamycin caused Cdc42 or WASP to localize to the cell membrane, resulting in actin polymerization and the formation of filopodia.

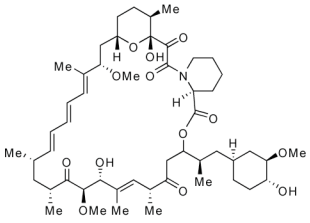
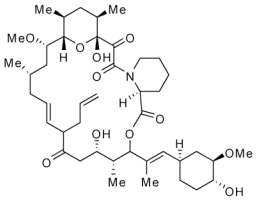
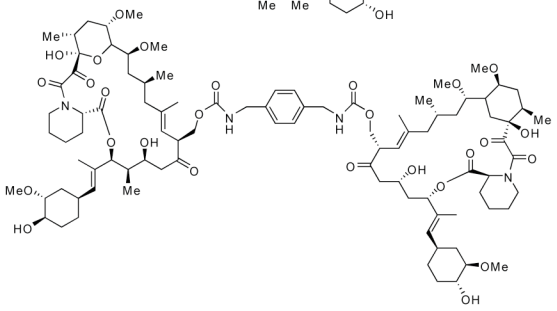
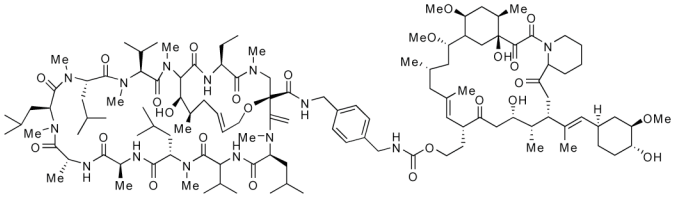
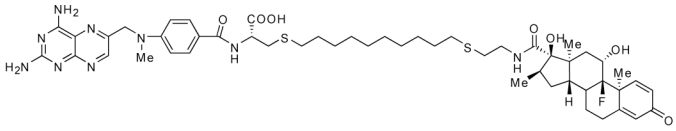
Name	Structure	Dimerized Domains
Rapamycin		FKBP + FRB
FK-506		FKBP + Calcineurin B
FK-1012		FKBP + FKBP
FKCsA		FKBP + Cyclophilin
Dexa-MTX		DHFR + Glucocorticoid receptor

Figure 1-4 : Sample CIDs and their targets. Shown here are the structures and protein targets of a number of CIDs. Some are natural products such as rapamycin and FK-506, others are artificial fusions between small molecules such as FKCsA [54] and Dexa-MTX [55].

Proteins destined for the cell surface or secretion traffic through the Golgi apparatus where they acquire a number of post-translational modifications. Many of the Golgi resident enzymes responsible for these processes display modular architecture with a transmembrane domain responsible for localization and a catalytic domain that carries out the post-translational modification. The Bertozzi lab took advantage of this property to generate a CID controlled protein allele of a fucosyltransferase [14]. The transmembrane domain was fused to FKBP and the catalytic domain to 3 copies of FRB. Cells expressing these constructs displayed rapamycin dependent protein glycosylation. This method was also shown to work on golgi sulfotransferases [15].

A remarkable feature of some proteins is that they can be split into inactive fragments, but upon forced dimerization, the fragments can refold and regain activity, a phenomenon known as protein complementation. Protein complementation of fluorescent or luminescent proteins as well as those required for cell survival is often used as an indicator of protein-protein interactions[56]. Proteins that are capable of complementation, can also be fused to dimerization domains such as FKBP and FRB and activated by the addition of a CID. A number of conditional protein alleles have been generated this way, including GFP [17-20], luciferase [21, 22], β-lactamase [23, 25, 27], β-galactosidase [24], ubiquitin [16] and dihydrofolate reductase [26].

Although dimerization has been exploited in a number of creative ways to control protein function, a conditional allele for every protein cannot be generated

by controlling dimerization. Dominant negative alleles are limited to proteins that are naturally regulated by dimerization. In order to make a CID dependent protein allele, the protein must be generated as a fusion with the dimerization domains which may affect the protein's activity. Additionally, every CID dependent protein allele must be independently designed as there is no completely general way to activate a protein through dimerization.

Section 1.4 : Controlled Degradation

An underappreciated means by which the cell regulates protein activity is through protein degradation, usually via the ubiquitin/proteasome pathway. For example, *Drosophila* are able to adjust their circadian rhythms to the rising and setting of the sun through the light mediated degradation of the TIMELESS protein [57]. *Salmonela* confine the membrane ruffling effects of the SopE protein to a narrow time window with an N-terminal domain that targets the protein for rapid degradation [58]. Defined more broadly to include partial as well as complete proteolysis, partial degradation is used to control such proteins as the caspases [59] and signaling molecules such as the NF$\kappa\beta$ family [60].

Protein stability can be exploited to gain experimental control over protein activity. In 2000, Howley and coworkers demonstrated that proteins could be destabilized by fusion to a domain that is targeted by the cell's ubiquitination machinery [61]. The Crews and Deshaies labs exploited this finding to generate

conditional protein alleles using a system they have termed PROteolysis TArgeting Chimeric moleculeS (PROTACS) [29]. In PROTACS, engineered CIDs recruit ubiquitin ligase to the targeted protein, resulting in its ubiquitination and subsequent degradation. In some cases, the targeted protein has a substrate that can be incorporated into the chimeric molecule, allowing a traceless knockout. If not, the target protein can also be generated as a fusion with a ligand binding domain. The Church lab has also described a similar method of targeted degradation [62].

Crabtree, Wandless and coworkers have also used protein stability to control protein acivity, but in the opposite way. An unstable, mutant FRB domain (FRB*) was fused to a variety of proteins resulting in their degradation. However, when the FRB* domain is bound to rapamycin and cellular FKBP, it adopts a stable, folded conformation. The rapamycin induced stability of FRB* also stabilizes the target protein allowing it to build up and generate activity [30]. This technique was used to activate GSK-3β during defined 48 hour periods in developing mice embryos, defining a role for GSK-3β in both hard palate and sternum formation [31].

Wandless and coworkers have reported a further refinement of the system. Instead of using a destabilized FRB*, a destabilized FKBP* was developed [32]. Proteins containing an FKBP* are degraded in a similar manner to those with FRB*. However, rather than requiring dimerization with cellular FKBP for stabilization, the FKBP* is directly stabilized through binding to an

engineered ligand that was also developed by the Wandless group. The resulting, stabilized protein is more similar to the wild type protein and is thus likely to behave in a more natural manner.

Chemically controlled degradation has the potential to be at least as generally applicable as techniques such as RNAi. PROTACS, FRB* and FKBP* have all been applied to multiple proteins and as more labs begin using the technologies will no doubt be applied to many more in the future. Unfortunately, a traceless method of conditional degradation is not yet available. In some cases, a natural ligand can be incorporated into a chimeric molecule allowing PROTACS on a wild type protein, but for the most part PROTACS requires the target protein be tagged with a ligand binding domain. Both FRB* and FKBP* based protein alleles, of course, require tagging the target protein with the destabilization tag.

Section 1.5 : Hormone Receptor Ligand Binding Domains

In 1988, Yamamoto and coworkers described an artificial transcription factor made by the fusion of the steroid binding domain of the glutocorticoid receptor with the E1A adenovirus transcription factor [63]. In the absence of hormone binding, the artificial transcription factor is bound by HSP90 and associated proteins, inhibiting activity. Following binding to Dexamethasone, the

fusion protein dissociates from the HSP90 complex and regains activity (**Figure 1-5A)**.

Since then, a variety of conditional protein alleles have been generated through fusion to hormone receptor ligand binding domains, from a wide array of transcription factors, to kinases and β-galactosidase [64]. Two more recent examples are particularly interesting since they can be used to indirectly generate conditional alleles of other proteins.

The development of Cre and Flp recombinase based transgenics has allowed the generation of specific knock-outs and knock-ins in mice and yeast. Despite the versatility of such techniques, they are still limited in their temporal precision. The placement of Cre recombinase under a tissue or development specific promoter can improve the spatial and temporal resolution of the knock-out, but this strategy is limited by the available promoters. In 1995, Chambon and co-workers described a Cre recombinase • ER-LBD fusion that was active only in the presence of 4-hydroxytamoxifen [33] (**Figure 1-5B**). Since then, other Cre • HR-LBD fusions have been made, responsive to a variety of steroids [34, 35, 65-67]. Similar work has been performed on Flp recombinase for use in yeast genetics[68, 69].

Figure 1-5 : Generating conditional protein alleles by fusion to hormone receptor ligand binding domains. (**A**) The target protein is expressed as a fusion with a hormone receptor ligand binding domain (HR-LBD). In the absence of hormone binding, the target protein is rendered inactive through binding to HSP90 and associated proteins. Hormone binding causes dissociation from the HSP90 complex, producing an active protein. (**B**) Fusion of CRE recombinase to the estrogen receptor ligand binding domain (ER-LBD) allows gene recombination to be triggered by tamoxifen. (**C**) Insertion of the ER-LBD into an intein results in tamoxifen controlled protein splicing.

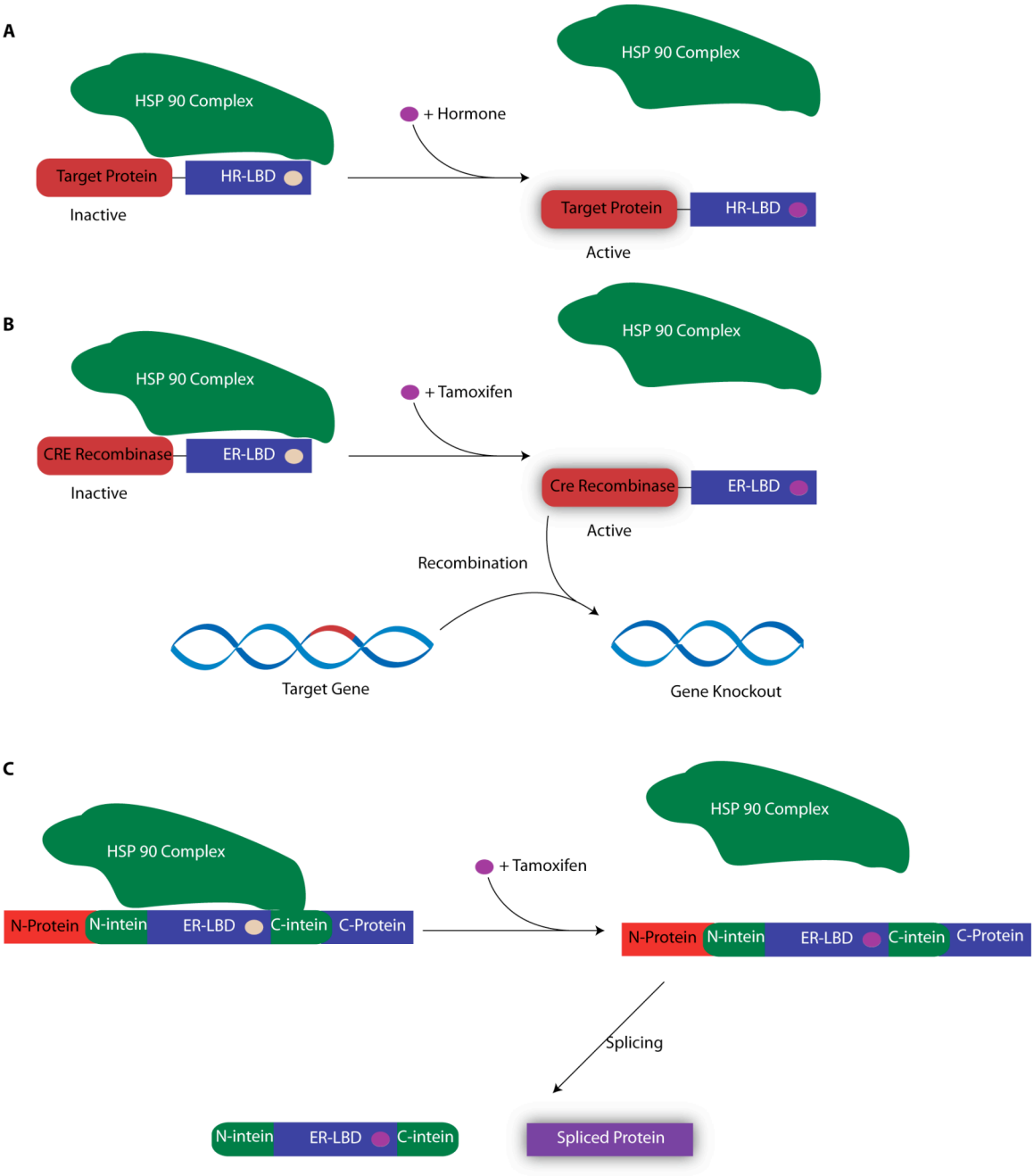


Figure 1-5 : Generating conditional protein alleles by fusion to hormone receptor ligand binding domains.

Intein mediated protein splicing and the related Expressed Protein Ligation (EPL) have already become widely used for biochemical preparations[70, 71]. Since protein splicing results in a dramatic change in primary protein sequence, it has attracted interest as a means of generating conditional proteins alleles. Unfortunately, intein splicing does not appear to be regulated. The Liu group generated a chemically controlled intein by the insertion of the ER-LBD into the RecA intein [36]. The resulting intein fusion required tamoxifen binding for activity and, when inserted into other proteins was able to generate conditional alleles of a variety of proteins.

Fusion with a hormone receptor ligand binding domain can imbue proteins of many classes with small molecule sensitivity. Unfortunately, the ligand binding domain remains fused to the protein which, as with other techniques that require a tag, may or may not affect the target protein's activity. Additionally, the unliganded LBD does not directly inactivate the target protein, rather it recruits the HSP90 apparatus which may not fully inactivate the target protein. The conditional intein described by Liu has the advantage that the LBD is removed during splicing, leaving a wild type or nearly wild type protein. Splicing, however, has a degree of sequence dependence so the conditional intein may not be fully general.

Section 1.6 : The need for a general, traceless conditional allele

To be truly confident in the validity of the phenotypes generated with a conditional allele, the protein must be as close to wild-type as possible. Otherwise, a variety of controls must be performed to ensure that what is being analyzed is not, in fact, a conditional mutant allele. Most of the techniques described above generate tagged proteins rather than wild-type proteins. In some cases, such as degradation from FKBP* the tag is relatively small and presumably have relatively minor effects, but larger tags can be cause for concern.

Ideally, a technique for manipulating protein function should be so straightforward and general that it can be used as “kit” science – a box with a vector, an aliquot of the drug and an instruction manual that any lab can pick up and, after a few subclonings be able to control their favorite protein in a rapid, dose dependent manner. Unfortunately, many of the methods described above do not fall into this category. With the possible exceptions of the FRB*/FKBP* controlled degradation systems, the design of a conditional protein allele has to be performed on a case by case basis and may not work at all.

Much of my thesis work has been spent developing methods of generating conditional protein alleles with special emphasis on minimizing the differences between the conditional and wild type alleles and on maximizing the generality of the technique. I have exploited two techniques for this purpose. The first,

conditional protein splicing (CPS) uses a split intein to splice together two inactive fragments of a protein to generate an active protein. This technique has the advantages of being nearly traceless and very fast, but unfortunately suffers from a lack of easy generality. The second technique relies on ubiquitin complementation to release a protein from a degradation tag, resulting in its stabilization and activity. This technique has the advantages of being easily generalized and traceless although, since it relies on protein translation to build up activity, it is not as fast as CPS based protein activation. CPS will be discussed in Chapter 2. Controlled protein degradation will be discussed in Chapter 3.

Chapter 2 : Conditional Protein Splicing

This work was performed in collaboration with Lino Saez and Michael Young of The Rockefeller University.

Section 2.1 : Introduction

Control over the timing, location and level of protein activity *in vivo* is crucial to a full understanding of biological function [1]. Living systems are able to respond to external and internal stimuli rapidly and in a graded fashion. One method by which the cell can achieve this is by maintaining a pool of proteins whose activity is altered through post-translational modifications [72]. In this manner, cells can adapt to environmental cues within minutes or even seconds. Thus, one means by which precise experimental control over cellular processes could be achieved is by directly altering post-translational protein modifications. Here we show that the post-translational process of protein trans-splicing [73] can be used to modulate enzymatic activity both in cultured cells and in an animal, specifically, *Drosophila melanogaster*. An optimized conditional protein splicing [74] (CPS) system was used to trigger the *in vivo* ligation of two inactive fragments of firefly luciferase. Enzymatic activity generated by trans-splicing appears rapidly and is tunable. This system provides a means of controlling enzymatic function with greater speed and precision than with standard genetic techniques and represents a useful tool for probing biological processes.

Section 2.1.1 : Protein splicing

Protein Splicing is a post-translational process in which a protein domain known as an “intein” is able to effect its own removal from a polypeptide and splice together the flanking regions or “exteins” with a native peptide bond [70, 71, 73, 75]. Over 350 intein or intein like proteins domains have been identified, from a variety of archae and eubacteria as well as in lower eukaryotes such as yeast [76]. Inteins have not been found in multicellular organisms although they bear some similarity to the hedgehog autoprocessing domain [73]. The machinery required for the splicing reaction is contained almost entirely within the intein, requiring only an additional cysteine or serine at the N-terminus of the C-extein (**Figure 2-1**). As such, inteins can be inserted into a variety of proteins without affecting the splicing reaction.

The relative promiscuity of inteins has allowed them to be exploited for a variety of biochemical purposes [70]. Mutant inteins which do not splice but rather exhibit N- or C-terminal cleavage can be used to purify recombinant proteins. The recombinant protein is expressed as a fusion with an affinity tag and a mutant intein. The affinity tag is used to immobilize the protein on a column. Following intein cleavage, the recombinant protein is released from both the column and the affinity tag, allowing the elution of a pure recombinant protein.

Since the mutant inteins can be cleaved by thiolysis, they can be used to generate recombinant proteins with a C-terminal thioester. This allows the facile

generation of proteins that can participate in the native chemical ligation (NCL) reaction [77]. In NCL, an N-terminal cysteine from one polypeptide selectively reacts with a C-terminal thioester from another polypeptide to generate a native peptide bond, joining the two polypeptides together. When one or both of the polypeptides is a recombinantly expressed protein, the NCL reaction is known as Expressed Protein Ligation (EPL) [78].

EPL has proven tremendously versatile for the site or domain specific incorporation of a wide variety of moieties [71]. If one of the two ligation fragments is made synthetically, nearly anything can be incorporated at any site in that piece. Ligation fragments that are expressed recombinantly can also be modified – either globally by changing the growth conditions of the cells or more specifically by chemically modifying the protein after expression. The site or domain specific modifications that can be made to proteins through EPL are limited only by imagination. To name a few, isotopic labels, fluorescent tags, phophorylation, ubiquitination, photocages, and unnatural amino acids have all been specifically incorporated into proteins using EPL [71, 75].

A remarkable subset of inteins have been identified which, rather than splicing in *cis*, splice in *trans* [73]. The intein itself has been split into two fragments during evolution, each of which is expressed, along with an extein as a separate protein. The intein fragments have a high affinity for one another [79], however, and when present in the same cell are able to associate, re-fold and splice together their exteins. In addition to the naturally split inteins that have

been identified, several inteins which normally splice in *cis* have been artificially split and shown to retain the ability to splice in *trans.* However, artificially split intein fragments have only a weak affinity for one another and thus, in order to force the *trans*-splicing reaction to occur, very high concentrations of protein are required.

Nonetheless, *trans*-splicing has found a place in the biochemist's toolbox. The *trans*-splicing reaction can be used in a similar manner to EPL to generate semisynthetic proteins [75]. The obvious drawback to using *trans*-splicing instead of EPL is that the intein must be incorporated into the protein fragments. This is trivial when expressing the proteins recombinantly, but can be challenging when done synthetically. However, since the split intein fragments can only react with their complemenatary intein fragment (indeed, even fragments from different inteins will not react with one another) the *trans*-splicing reaction can be performed in crude cell lysates [79] and even in living cells [80] without the need for purification.

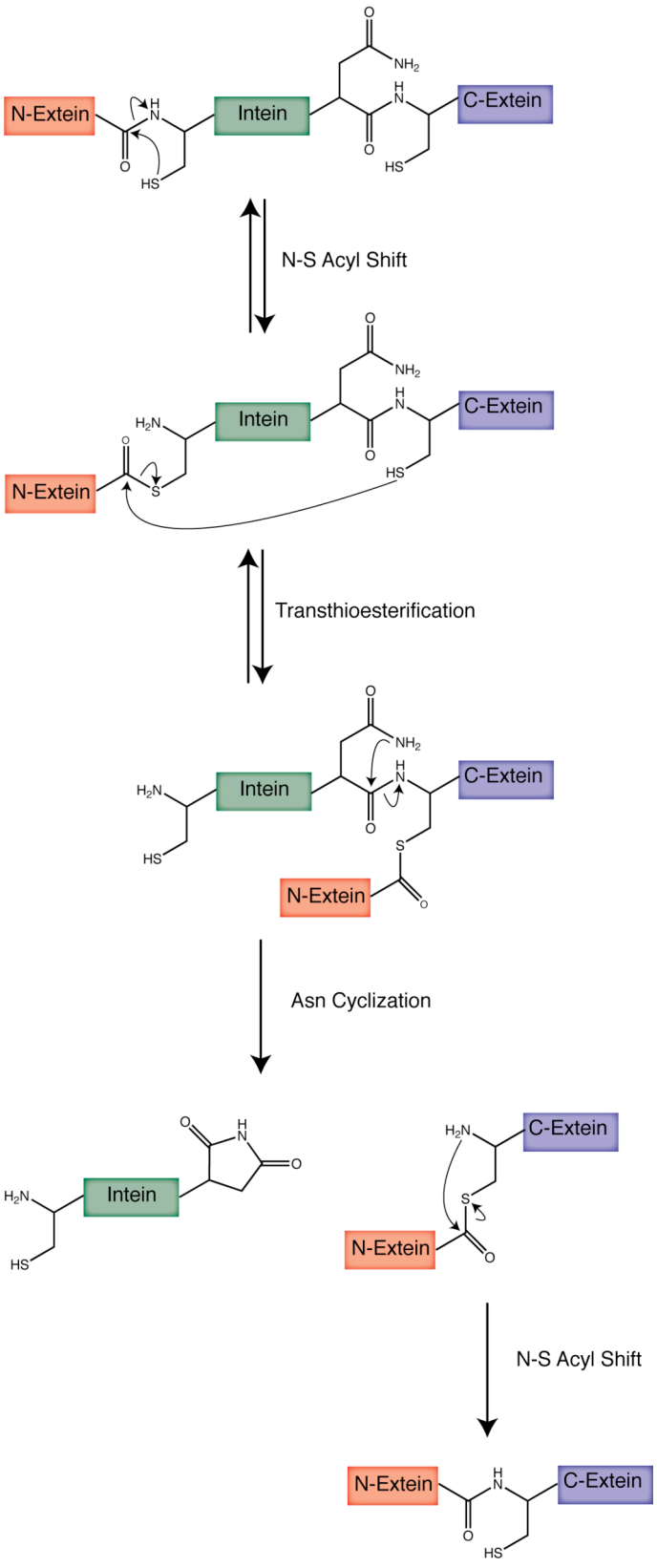


Figure 2-1 : The protein splicing mechanism.

Section 2.1.2 : Controlling protein splicing

The dramatic changes in sequence and structure that accompany splicing can affect the activity of a protein, but since inteins themselves are not regulated, this is of little experimental utility. As a result, there has been some effort put into generating conditional inteins that can in turn be used to conditionally activate or inactivate a target protein.

Adam and Perler have developed a temperature sensitive (TS) intein by inserting the *Mycobacterium* GyrA intein into *E. coli* DNA gyrase and screening for splicing at restrictive temperatures [81]. The Perrimon group has also reported a TS intein [82]. A library of intein mutants was screened for splicing at different temperatures resulting in the isolation of an intein that splices efficiently at 18 °C but not at 30 °C. This TS intein was used to generate TS alleles of Gal4 and Gal80. When used in combination with Gal4 enhancer trap lines, this allows the control of a wide array of genes' activities with simple temperature shifts in both cultured cells and living *Drosophila*.

As noted in chapter 1, the Liu group has developed an Estrogen Receptor Ligand Binding Domain (ER-LBD)-intein fusion in which the intein's activity requires the binding of tamoxifen to the ER-LBD (**Figure 1-5C**). The ligand binding domain of the ER was inserted into the RecA intein. The ER-intein fusion was then mutagenized by error prone PCR and subjected to screening and selection to evolve intein splicing only in the presence of 4-Hydroxy-tamoxifen.

The tamoxifen sensitive intein has been used to generate a variety of conditional protein alleles, including GFP, β-galactosidase, and the transcription factors Gli1 and Gli3T [36, 37].

Previous members of the Muir lab had developed a Conditional Protein Splicing (CPS) system in which fragments of an artificially split *S. cerevisieae* vacuolar ATPase (VMA) intein complement through rapamycin induced heterodimerization of the FKBP and FRB proteins (**Figure 2-2A**). The original CPS report demonstrated the rapamycin triggered splicing of the purified model proteins MBP and Hisx6 [74]. This was rapidly followed up by a report in which the process was shown to work in cultured mammalian cells [83]. The process was also shown to be rapid and dose dependent. In later work, the activity of purified protein kinase A was modulated using CPS [84]. A modified protein kinase A was generated as a fusion with the C-terminal CPS cassette and a substrate mimic peptide inhibitor (**Figure 2-2B**). Although the peptide inhibitor has a relatively weak K_I, since it is tethered to the kinase, it is able to efficiently inhibit enzymatic acitivity. Following either splicing or C-terminal cleavage, the inhibitor is released from the kinase, allowing it to recover activity. Finally, CPS has been used in conjunction with the naturally split DnaE intein to perform 3 piece protein ligations for biochemical studies [79].

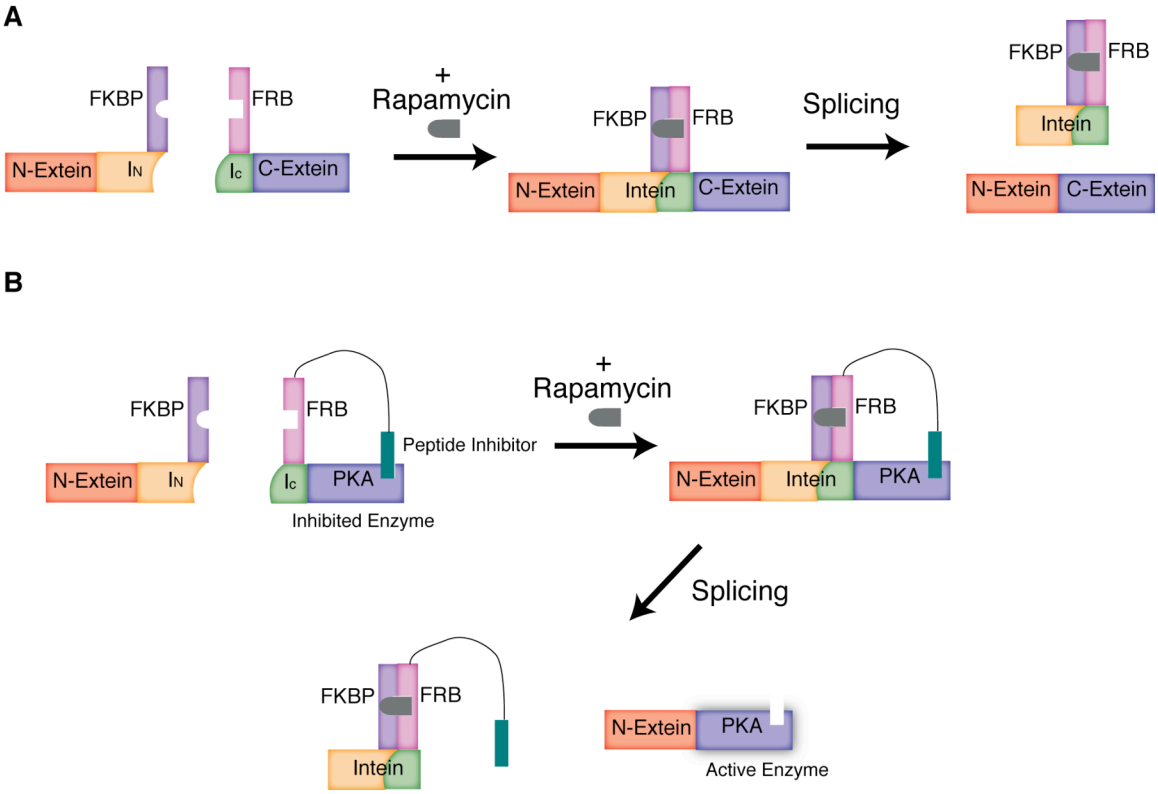


Figure 2-2 : Conditional Protein Splicing. **(A)** Fragments of the *S. Cerevisiae* VMA intein (I_N and I_C) are fused to the FRB and FKBP domains and to two proteins or peptides (N-extein and C-extein). In the presence of rapamycin, FKBP and FRB dimerize, which brings the intein fragments together, thereby leading to complementation and splicing. **(B)** CPS has been used to activate an auto-inhibited kinase. Protein Kinase A (PKA) was generated as a fusion with FRB, the C-terminal intein fragment (Ic) and a peptide inhibitor of PKA. Upon splicing, PKA was released from the peptide inhibitor, generating an active enzyme.

Section 2.2 : Results

Section 2.2.1 : Generating conditional protein alleles with CPS

We were interested in using CPS as a general means of producing conditional protein alleles both in cultured cells and in living animals. The only other use of CPS to regulate protein activity was the auto-inhibited kinase described above. Unfortunately, this method could not be applied to proteins which could not be reversibly inhibited in *cis* by a peptide. Additionally, although this may be because of the kinase chosen, CPS regulation of protein kinase A was never successfully performed in cultured cells, much less in living animals.

We reasoned that the most general way of inactivating a protein would be to split it into fragments. Upon splicing, the full length, wild type protein would be regenerated as well as, presumably, activity. To test this, we chose firefly luciferase as a model enzyme. Luciferase has a number of attractive features for developing and optimizing CPS as a means of generating conditional protein alleles[85]. The bioluminescent activity of luciferase is easily detected and quantified allowing one to quickly determine if the system was working and with what efficiency. Luciferase requires no post-translational modifications for activity, so it should be active as soon as it has refolded following splicing. Finally, since luciferase is a monomeric protein, we would not have to worry about potential dominant negative effects from unspliced luciferase fragments.

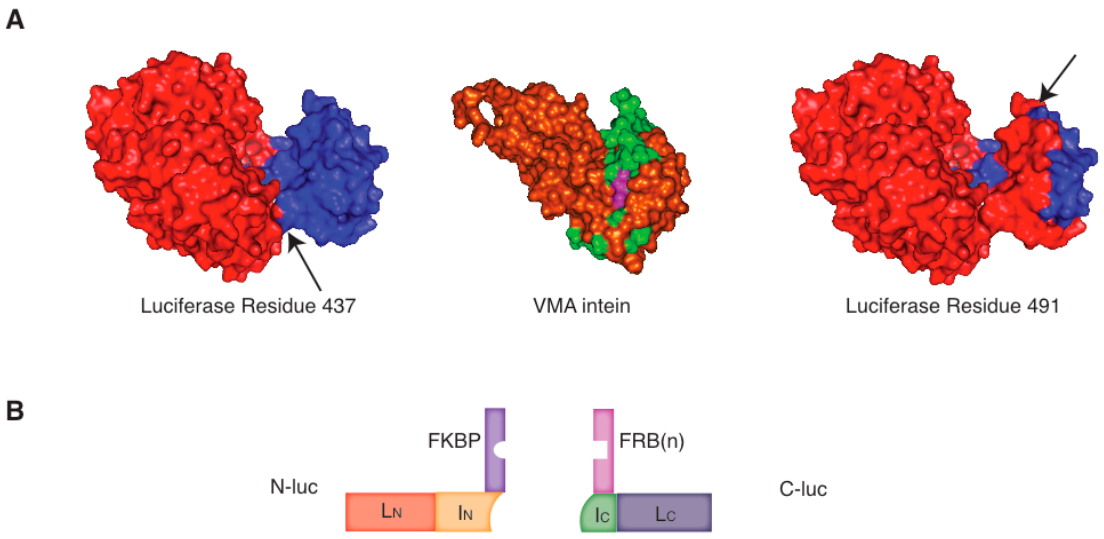


Figure 2-3 : Design of a conditional allele of luciferase. (A) Space-filling models of firefly luciferase depicting the different splicing sites we tested. The N- terminus of luciferase is in red, the C- terminus is in blue, and splice sites are indicated by arrows. Also shown is the structure of the VMA intein with the active splicing residues shown in magenta (N-terminus, orange; C-terminus, green). We did not observe splicing from the split at Arg437, but we were able to generate conditional splicing at Lys491. **(B)** The constructs used in this study. In N-luc, the N-extein (L_N) is luciferase$_{1\text{-}490}$ with the mutation H489K. In C-luc, the C-extein (L_C) is luciferase$_{491\text{-}551}$ with the mutation K491C and tandem copies of FRB (FRB(n), where n = 1-3).

When deciding where to split luciferase for the generation of a CPS based conditonal protein allele, we were aided by the fact that a crystal structure has been determined. Luciferase is a 61kDa monomeric protein made up of a large N-terminal lobe and a small C-terminal lobe (**Figure 2-3A**). Another group had reported protein *trans*-splicing on luciferase using a different intein by splitting the

protein in between the N- and C-terminal lobes. Thus, we initially split the protein at the same site, residue Arg437. Surprisingly, despite robust expression in *Drosophila* (Schneider's line 2, S2) cells, no *trans*-splicing or luciferase activity was observed (**Figure 2-4**). Modeling studies suggested that steric crowding within the trans-splicing complex might account for the lack of splicing and enzymatic activity (**Figure 2-3A**); each structured component would act like a wedge preventing functional complementation of the other. Consistent with this, the construct encoding luciferase 437-551 was able to splice with other proteins such as MBP (**Figure 2-5**).

With this in mind, we returned to the crystal structure to design a new split site. We reasoned that since it was likely the tertiary structure of luciferase that was interfering with the complementation and splicing of the intein, the split site should be chosen to minimize the interaction between the intein and luciferase. This led us to choose Lys491, which is located on a surface exposed loop within the smaller C-terminal lobe (**Figure 2-3A**) and should be compatible with the presence of the folded intein. This gave rise to the CPS constructs referred to as N-luc (encoding (Flag)-(Luciferase[1-490])-(N-VMA)-(FKBP)) and C-luc (encoding (Flag)-(FRB)-(C-VMA)-(Luciferase[491-551]), **Figure 2-3B**).

Figure 2-4 : Attempted splicing at luciferase R437. Attempts to splice at Luciferase R437. (**A**) Drosophila S2 cells were transfected with constructs encoding Luciferase1-436 (D436G) – N-VMA – FKBP and FRB – C-VMA – Luciferase 437-551 (R437C) and allowed to express for 40 hours. Cells were then treated with 0.1% DMSO (D) or with 1 mM rapamycin (R) for 5 h before being analyzed by western blot using an anti-luciferase antibody. The molecular weight of luciferase is 60.7 kDa. (**B**) *Drosophila* S2 cells were transfected with the indicated plasmids and a renilla luciferase transfection control. Following 16 hours of expression, cells were treated with 0.1% DMSO or rapamycin to a final concentration of 100 nM and 0.1% DMSO. Rapamycin treatment was allowed to proceed for 4 hours before cells were harvested and assayed for chemiluminescence. A CPS inducible split luciferase (N-luc and C-luc) was included as a positive control. Error bars represent standard deviation (n=3).

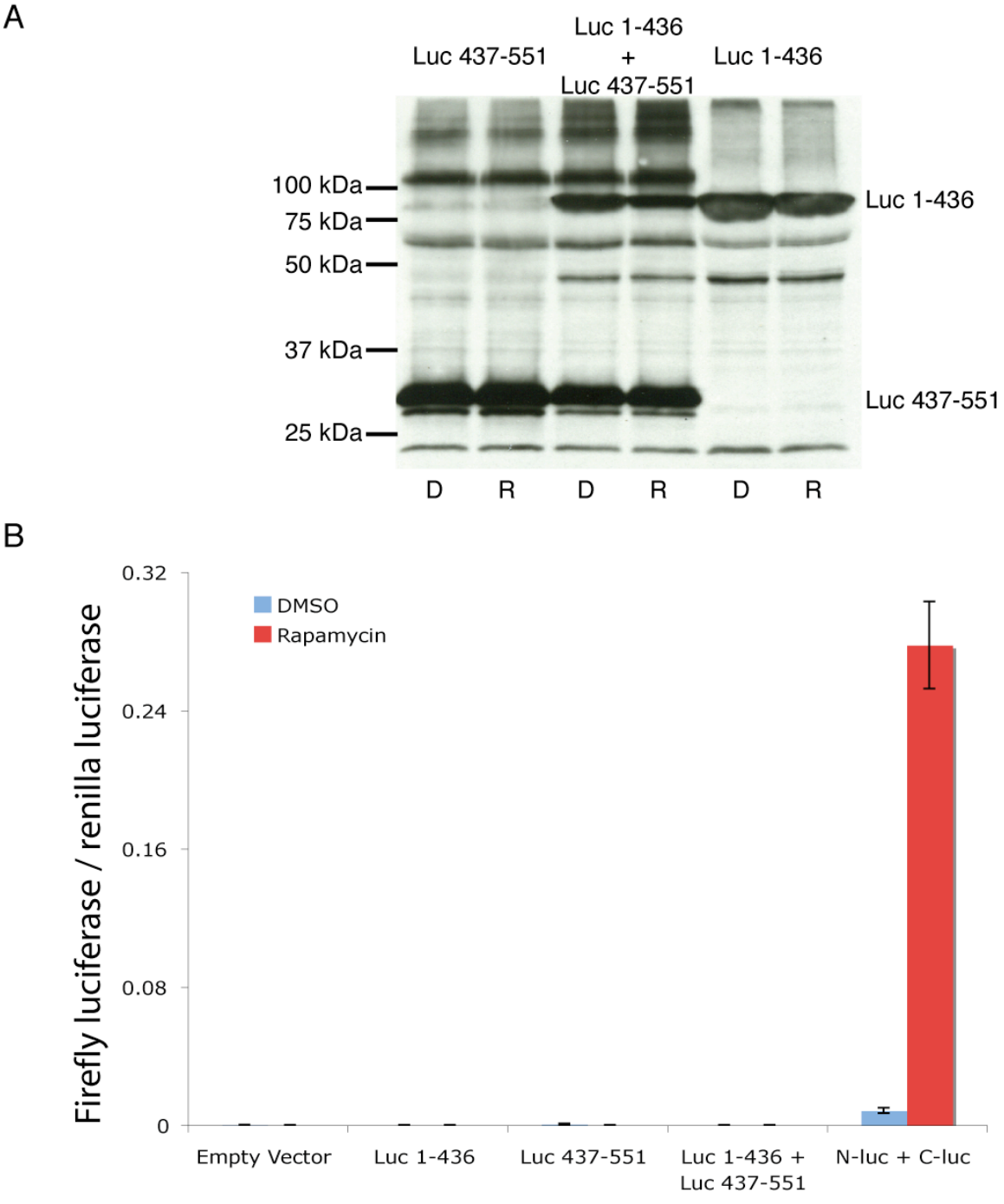


Figure 2-4 : Attempted splicing at luciferase R437.

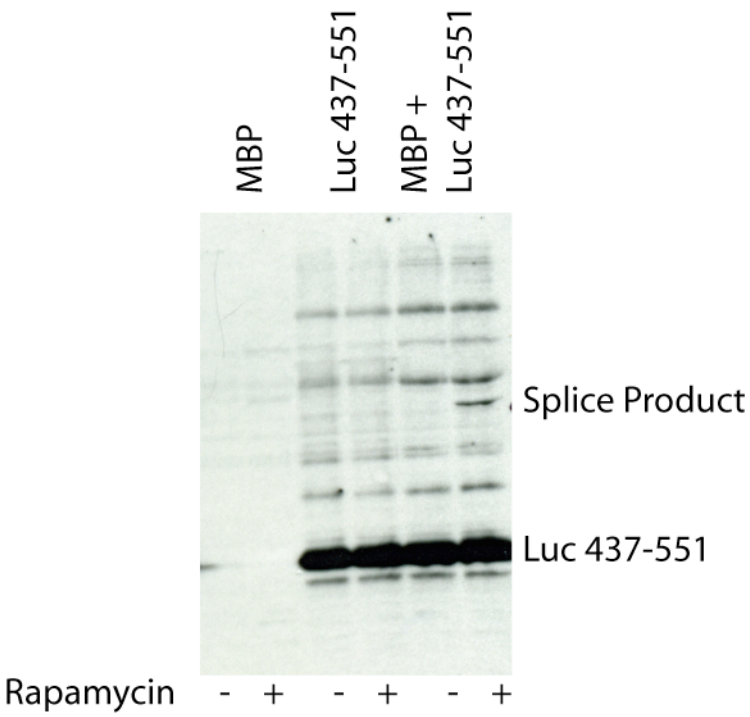


Figure 2-5 : Splicing of Luc 437-551 with Maltose Binding Protein (MBP). Constructs encoding Luc 437-551 and one encoding MBP-I_N-FKBP (MBP) under the control of the heatshock promoter were used to transfect S2 cells. Two days after transfection, cells were incubated at 37 °C for 5 minutes followed by 5h at room temperature. Cells were then treated with 0.1% DMSO or 100 nM rapamycin for 2h then harvested and analyzed by western blot using a luciferase antibody.

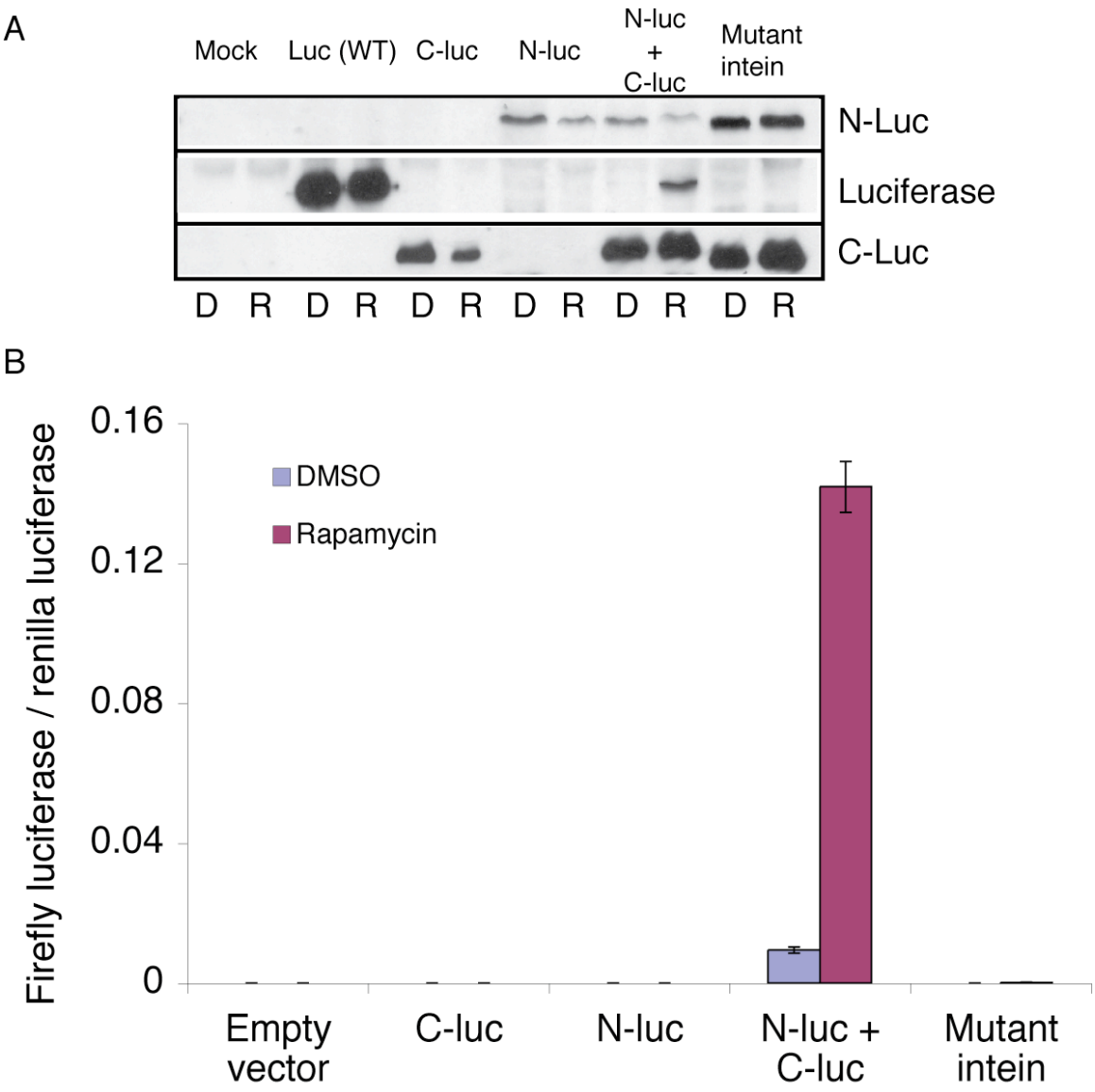


Figure 2-6 : Splicing and activation of luciferase in cultured cells. *Drosophila* S2 cells were transfected with the indicated constructs and treated as described in **Figure 2-4b** (**A**) Western blot using an anti-luciferase antibody against lysates from cells treated with DMSO (D) or rapamycin (R). Loading was normalized through renilla luciferase activity levels. (**B**) Luciferase activity in cell lysates expressed as a ratio between firefly luciferase activity and renilla luciferase activity. Error bars represent standard deviation (n=3). The mutant N-Luc and C-Luc constructs contain inactivating C1A and N454A mutations, respectively.

Cultured S2 cells were transfected with N-luc and C-luc. Following protein expression, cells were treated with rapamycin or DMSO vehicle. Western blotting confirmed the rapamycin dependent generation of full-length luciferase (**Figure 2-6A**). However, since presumably the split at Lys491 would result in an unfolded C-terminal lobe, it was quite possible that the splice product would not have enzymatic activity.

To determine if the spliced luciferase could fold and gain enzymatic activity, S2 cells expressing N-luc and C-luc were treated with rapamycin and assayed for luciferase activity. As expected, neither luciferase fragment displayed any enzymatic activity on its own (**Figure 2-6B**). However, when cells expressing both N-luc and C-luc were treated with rapamycin, luciferase activity increased dramatically (**Figure 2-6b**), and in a dose-dependent manner (**Figure 2-7**). In order to control for possible differences in transfection efficiency, firefly luciferase activity was normalized to renilla luciferase activity, which was co-transfected. To avoid potential complications from rapamycin side effects, we also tested a rapamycin analog (AP21967, Ariad). The rapamycin analog and the FRB (containing the mutation T2098L) used in this study were designed using the bump and hole approach described above. The modified rapamycin is able to bind to the mutant FRB, but not to the native TOR protein, preventing off target effects. The rapamycin analog was able to induce splicing nearly as well as rapamycin (**Figure 2-7**).

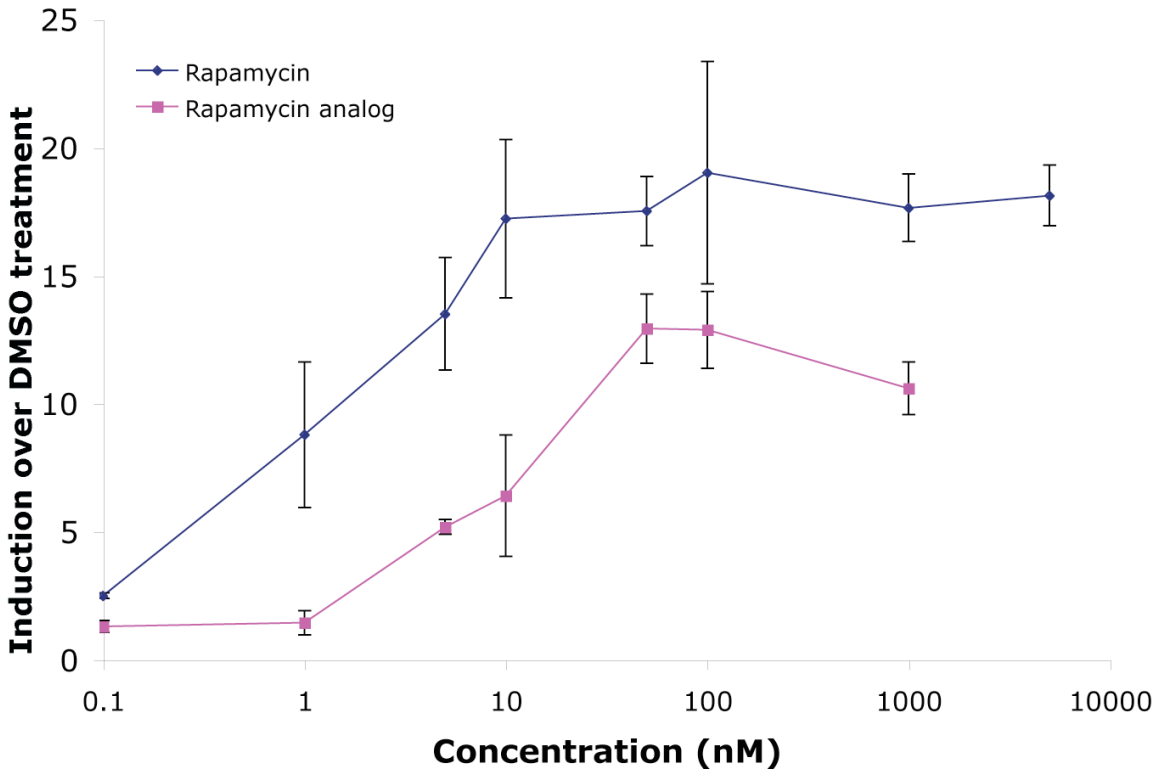


Figure 2-7 : Dose dependence of splicing in response to rapamycin or rapamycin analog. S2 cells were transfected with N-luc and C-luc as well as renilla luciferase transfection control. After 16 hours, cells were treated with the indicated level of rapamycin or rapamycin analog (Ariad AP21967). After 4 hours, cells were assayed for luciferase activity and plotted against the level of luciferase activity in cells treated with DMSO control. Error bars represent standard deviation (n = 3-4, except 100 pM and 5 mM data points where n = 2).

To ensure that the observed luciferase activity was due to protein splicing and not complementation of the luciferase fragments, we generated versions of N-luc and C-luc with intein mutations rendering them splicing incompetent. Two residues are required for the splicing reaction – the first (cysteine) and last (asparagine) residues of the intein (**Figure 2-1**). By mutating these to alanine,

the N-S acyl shift and the transthioesterification steps are blocked, preventing splicing. No significant splicing or chemiluminescence activity could be detected from cells expressing these mutant constructs, either with or without rapamycin treatment (**Figure 2-6B**). Thus, splicing is necessary for the generation of luciferase activity, and the small amount of luciferase activity observed when the N-luc and C-luc constructs are co-expressed is due to a low level of rapamycin-independent splicing.

Although rapamycin increased luciferase activity by more than an order of magnitude, we wanted to improve the ratio of induced to background splicing. There were two main ways we could go about this – attempting to improve the splicing reaction itself or improving the association between the two fragments. We chose to try to optimize the splicing reaction through temperature effects and to improve the association between the fragments through avidity.

Figure 2-8 : Optimization of the luciferase CPS system. (**A**) The effects of temperature on the splicing reaction. S2 cells transfected with N-luc and C-luc were incubated for 6h at 25 °C and then transferred to the indicated temperature for 10h. Cells were then treated with 0.1% DMSO or 100 nM rapamycin for 4h at the indicated temperature before being lysed and assayed for luciferase activity. (**B**) Effect of multiple copies of FRB on the induction of luciferase activity. S2 cells were transfected with N-luc and C-luc with 1-3 copies of FRB. Following 16h of gene expression, cells were treated with 0.1% DMSO or 100 nM rapamycin for 4 hours and then assayed for luciferase activity. (**C**) Kinetics of activation through splicing. S2 cells transfected with N-luc and C-lucFRB2 were treated with 0.1% DMSO or 100 nM rapamycin after 16h of gene expression. Cells were harvested at the indicated time points and assayed for luciferase activity. (**D**) Use of ascomycin competition to regulate splicing. S2 cells expressing N-luc and C-lucFRB2 were treated with 0.1%DMSO or 5 nM rapamycin. Ascomycin was added to a final concentration of 500 nM, 15 minutes or 1 hour after rapamycin treatment without washing out the rapamycin. Cells were harvested at the indicated time points and assayed for luciferase activity. Error bars in all cases represent standard deviation (n=3).

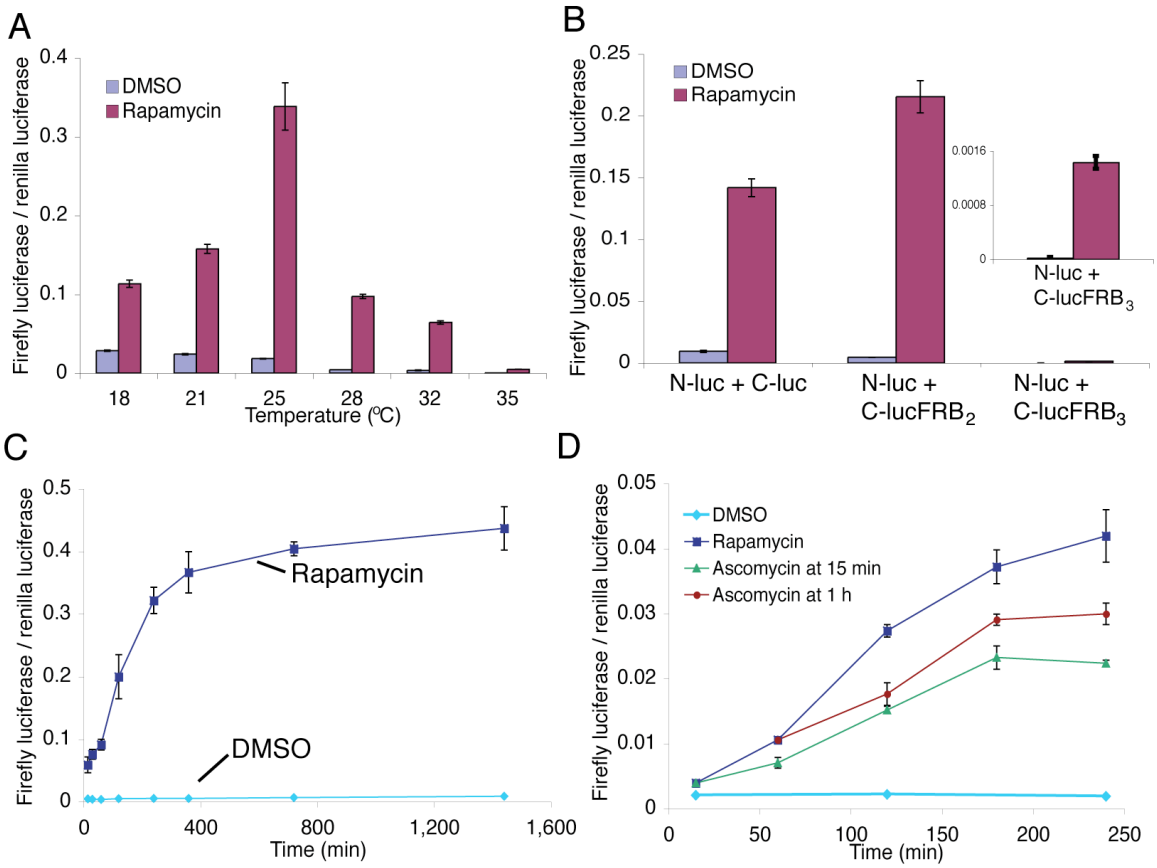


Figure 2-8 : Optimization of the luciferase CPS system.

The splicing efficiency of the VMA intein is known to be dependent on temperature [86, 87], so we reasoned that there may be an optimum temperature that minimized background splicing while still allowing rapamycin induced splicing. N-luc and C-luc were co-expressed and the cells treated with rapamycin at different temperatures. Background splicing decreased with increasing temperature while the level of rapamycin induced luciferase activity was highest at 25 °C (**Figure 2-8A** and **Figure 2-9**). The ratio between background and rapamycin induced luciferase activity was highest at elevated temperatures, for example 6-fold at 21 °C compared to 22-fold at 28 °C. The lower levels of activity at elevated temperatures are likely due to S2 cell physiology. For example, 35 °C is not optimum for the growth of this cell line. This is supported by the fact that CPS induction of luciferase is efficient in mammalian cells at 37 °C (**Figure 2-9**).

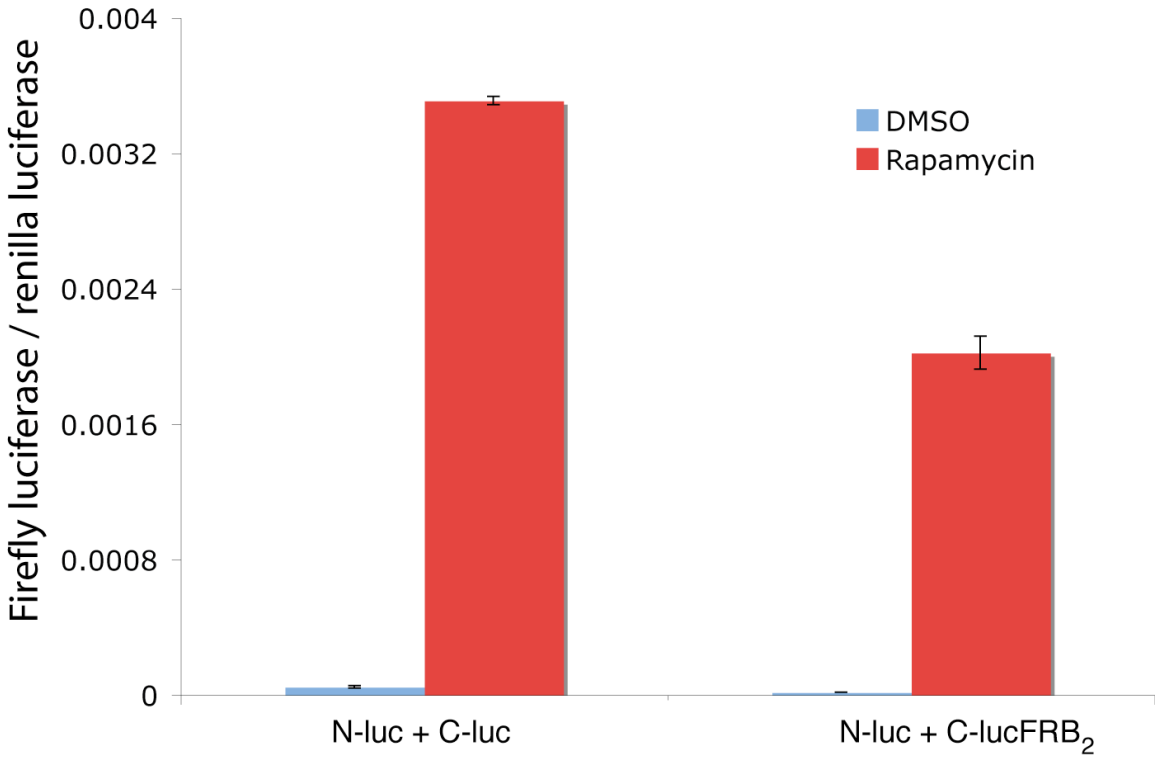


Figure 2-9 : CPS activated luciferase in mammalian cells. N-luc, C-luc and C-lucFRB$_2$ were cloned into the pCDNA3.1 vector (Invitrogen). Hela cells were transfected with these constructs as well as pRL-SV40 (renilla luciferase control, Promega). Following 16 h expression, cells were treated with DMSO or 100 nM rapamycin for 4 hours. Cell lysates were then analyzed for luciferase activity. Luciferase activity increased 75 and 130-fold with 1, 2 copies of FRB respectively. Error bars represent standard deviation.

Additional copies of the heterodimerization domains are known to improve rapamycin-induced association [9, 14]. Thus, we generated versions of C-luc containing 2 and 3 tandem copies of FRB (C-lucFRB$_2$ and C-lucFRB$_3$, **Figure 2-2C**). In both cases, the level of rapamycin induced splicing over background was better than with a single FRB domain (**Figure 2-8B**). Luciferase activity increased

15, 50 and 75-fold over background with 1, 2 and 3 copies of FRB respectively. However, C-lucFRB$_3$ expresses poorly (**Figure 2-10**), leading to very low levels of luciferase activity albeit with negligible background splicing. Thus, the C-lucFRB$_2$ construct was used for further characterization of the system, although for certain applications, the C-lucFRB$_3$ version may be preferable.

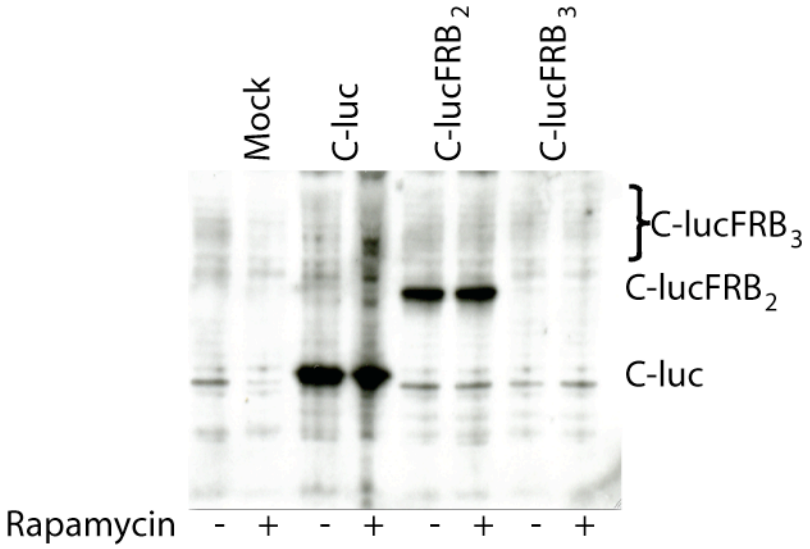


Figure 2-10 : Poor expression of C-lucFRB$_3$. Cells were transfected with the indicated constructs and treated as described in Figure 3. Cells were analyzed by western blot using an anti-luciferase antibody. C-lucFRB$_3$ should migrate above C-lucFRB$_2$.

One of the major advantages of small molecule based approaches is the speed with which they work. To evaluate the kinetics of the CPS reaction, S2 cells expressing N-luc and C-lucFRB$_2$ were treated with rapamycin and monitored for luciferase activity at a number of time points over 24 hours. By fifteen minutes, the earliest time-point taken, luciferase activity increased by an order of magnitude over DMSO treated cells and continued to increase over the

next 24 hours (**Figure 2-8C**). Additional experiments were performed to examine whether the kinetics of splicing reactions could be tuned by adding a second small molecule, ascomycin, which acts as a competitive inhibitor of rapamycin induced dimerization [74]. S2 cells expressing N-luc and C-lucFRB$_2$ were treated with rapamycin followed by ascomycin either 15 minutes or 60 minutes later. Addition of the antagonist led to the attenuation of trans-splicing (**Figure 2-8D**).

Next, we asked whether the CPS approach would work in a living animal. Transgenic flies were generated containing genomically integrated copies of the luciferase CPS expressed under the control of the *timeless* promoter. Adult flies were placed in individual chambers with food containing luciferase substrate and either DMSO or rapamycin. Luminescence was then monitored for 1-2 days. Flies fed DMSO had very similar luminescence levels to non-transgenic flies, whereas flies fed rapamycin exhibited easily detected luciferase activity (**Figure 2-11A-B** and **Appendix 1**). Western blotting confirmed the generation of full-length luciferase upon feeding the transgenic flies rapamycin (**Figure 2-12**). The jagged nature of the luciferase traces are a product of the assay since the same profile is seen with positive control flies expressing full-length luciferase under the control of the *timeless* promoter (**Figure 2-11C and Appendix 1**). Similar traces have been obtained in other studies using real time luciferase imaging in fruit flies[88, 89]. CPS induced luciferase activity could be detected by the time the first measurements were taken, as early as 10-20 minutes after flies were placed on rapamycin containing food. However, in some cases the kinetics were on the

order of a few hours, which likely reflects differences in feeding behavior (**Appendix 1**). CPS induced luciferase activity could also generated when the constructs were expressed under the control of the *armadillo* promoter, albeit with slower kinetics (**Appendix 1**).

Figure 2-11 : Activation of luciferase in living *D. melanogaster*. Flies were deprived of food for 1 hour before being anesthesized with CO_2 and placed in individual chambers with food containing 200 μM luciferin and either 2% DMSO (light blue) or 100 μM rapamycin (dark blue). Luciferase activity was then monitored from individual flies every 12 minutes with a 1 minute acquisition in a luminometer. (**A**) Non-transgenic flies (*yw*). (**B**)Flies expressing N-Luc and C-LucFRB$_2$ under the control of the *timeless* promoter (yw;*tim*-Gal4,N-luc;C-lucFRB$_2$). (**C**) Flies expressing full length luciferase under the *timeless* promoter (*tim*-Luc;CyO;Sb/TM6B-Tb).

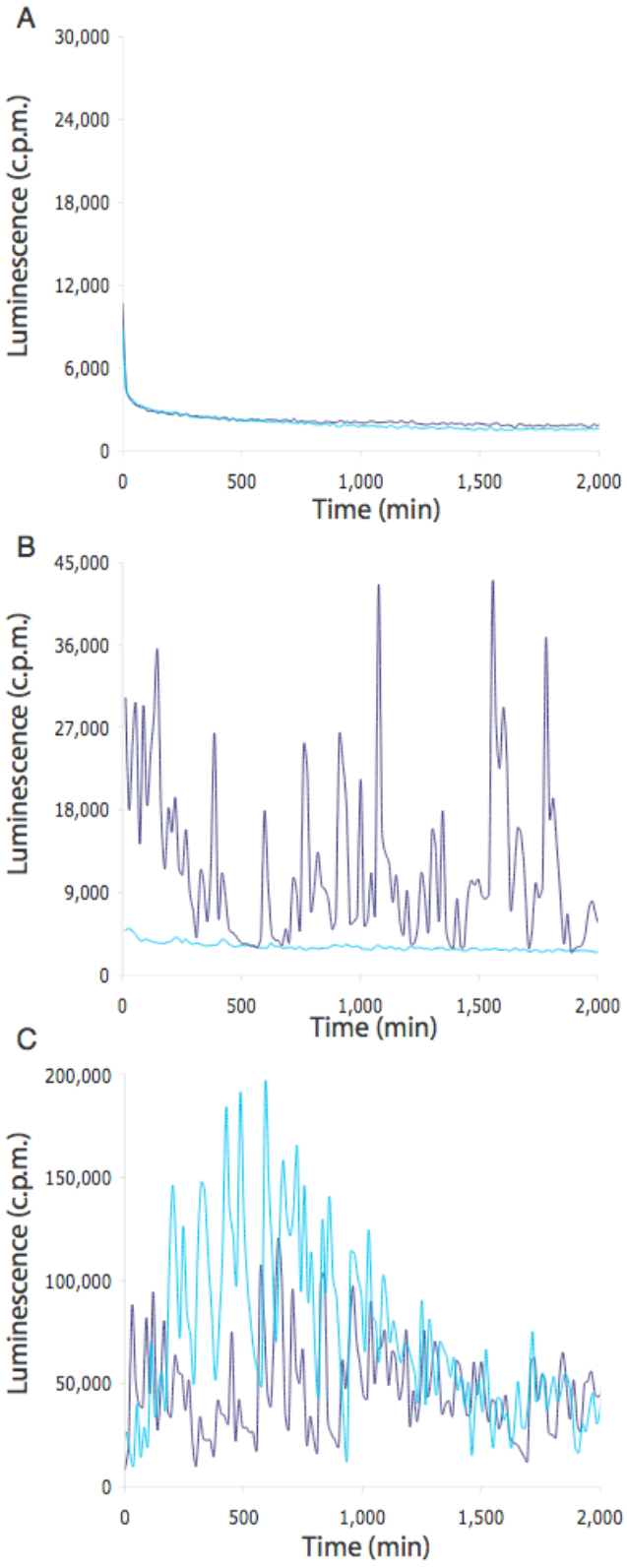


Figure 2-11 : Activation of luciferase in living *D. melanogaster*.

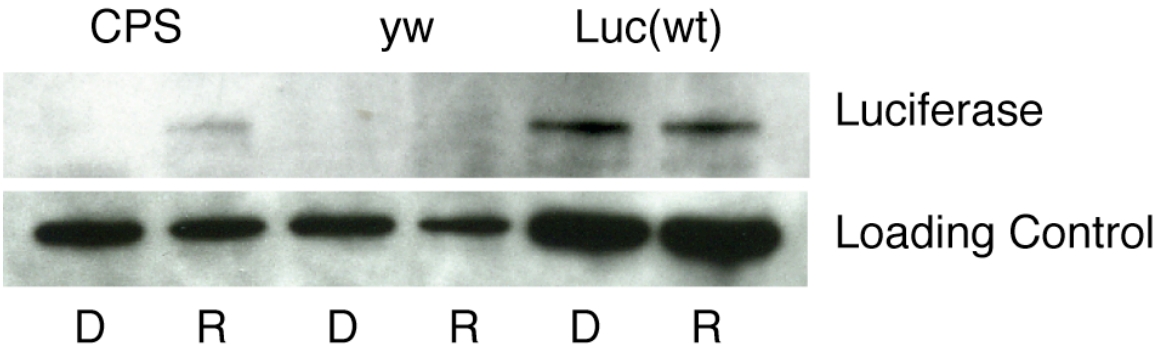


Figure 2-12 : Generation of full length luciferase through CPS in adult *Drosophila*. Adult flies of genotypes *yw*;*arm*-Gal4,N-luc;C-luc (CPS), *yw* and *tim*-Luc;CyO;Sb/TM6 (Luc(wt)) were deprived of food for 1 hour and then placed on 3MM filter paper soaked in 40% glucose supplemented with 100 μM rapamycin (2% DMSO) for 12h. Individual flies were homogenized in PLB with 1 mM $ZnCl_2$. Protein concentration in fly lysates was determined by the Bradford assay (Biorad). 5 μg of total protein was loaded from each fly lysate. A non-specific band served as a control for transfer efficiency.

Section 2.2.2 : How good is CPS at generating conditional protein alleles?

As a means of generating conditional protein alleles, CPS has a number of attractive features. First and foremost is the speed by which it functions. The induction of activity can occur in less than 15 minutes in both cultured cells and in living animals. By way of comparison, using the FKBP-FRB dimerization system to drive transcription through association of fragments of a transcription factor can take days to generate gene products *in vivo.* Speed also sets CPS apart from techniques that rely on protein stability as a mode of control since those processes are limited either by the rate of protein synthesis or degradation.

Like most conditional protein alleles, those generated by CPS are titratable. In the CPS system, activity levels can be in a variety of ways. Most obvious is rapamycin dosage. The use of ascomycin competition or temperature changes can further tune the response. Finally, FRB copy number can be used to set the upper and lower limits of protein activity that will be generated.

Another major advantage over many other techniques is the relatively traceless nature of the knock-in. The final, spliced protein differs from wild-type luciferase in only 3 point mutations, only one of which we are certain is required for protein splicing and recovery of function. Although a flag tag was added for the purposes of western blotting, this was simply for experimental convenience and likely could be removed without affecting the protein's activity. Although

admittedly a tag, no matter how big does not necessarily affect a proteins activity indeed, often does not, using a traceless technique such as CPS could serve to eliminate any concerns of non-wild-type activity from the start of an experiment.

Unfortunately, where CPS falls short as a means of generating conditional proteins alleles is in its generality and its lack of reversibility. Splicing can be attenuated by washing out rapamycin or by the addition of ascomycin, but it cannot be reversed. Although splicing in *cis* has been reported to be almost completely promiscuous with regard to flanking extein sequences, our experience with conditional *trans* splicing has shown a very strong sequence dependence on splicing efficiency. We have already mentioned above one luciferase splice site that was tried but failed. In addition to this, 5 other splice sites were cloned and tested, only one of which worked (**Figures 2-13 and 2-14**).

The most obvious difference between the site that worked (K491) and the site that failed (R437) is that the site that succeeded is located in a surface exposed loop while the one that did not is located between two folded lobes. Consistent with the idea that surface exposure is required for splicing to occur, the second site that successfully spliced (T508) is also located in a surface exposed loop. However, a third surface loop residue (S185) was also tested and displayed neither splicing nor activity. Thus, while surface exposure may be necessary for splicing, it does not appear to be sufficient.

Figure 2-13 : Splicing at luciferase residue 508. *Drosophila* S2 cells were transfected with the indicated constructs and treated as described in **Figure 2- 4a**. (**A**) CPS constructs of comprising luciferase 1-507 and luciferase 508-551 were able to splice in the presence of rapamycin to generate full length luciferase. Luciferase 508-551 was also able to splice with a CPS construct encoding Maltose Binding Protein (MBP). The constructs used for splicing at luciferase 491 and at luciferase 508 migrate almost identically on SDS-PAGE and are thus labeled only once as "N-luc" and "C-luc." (**B**) Luciferase generated by the splicing of luciferase 1-507 and luciferase 508-551 is active. Raw luciferase counts are shown from equal volumes of lysate as the Renilla luciferase control was not yet available at the time of this experiment.

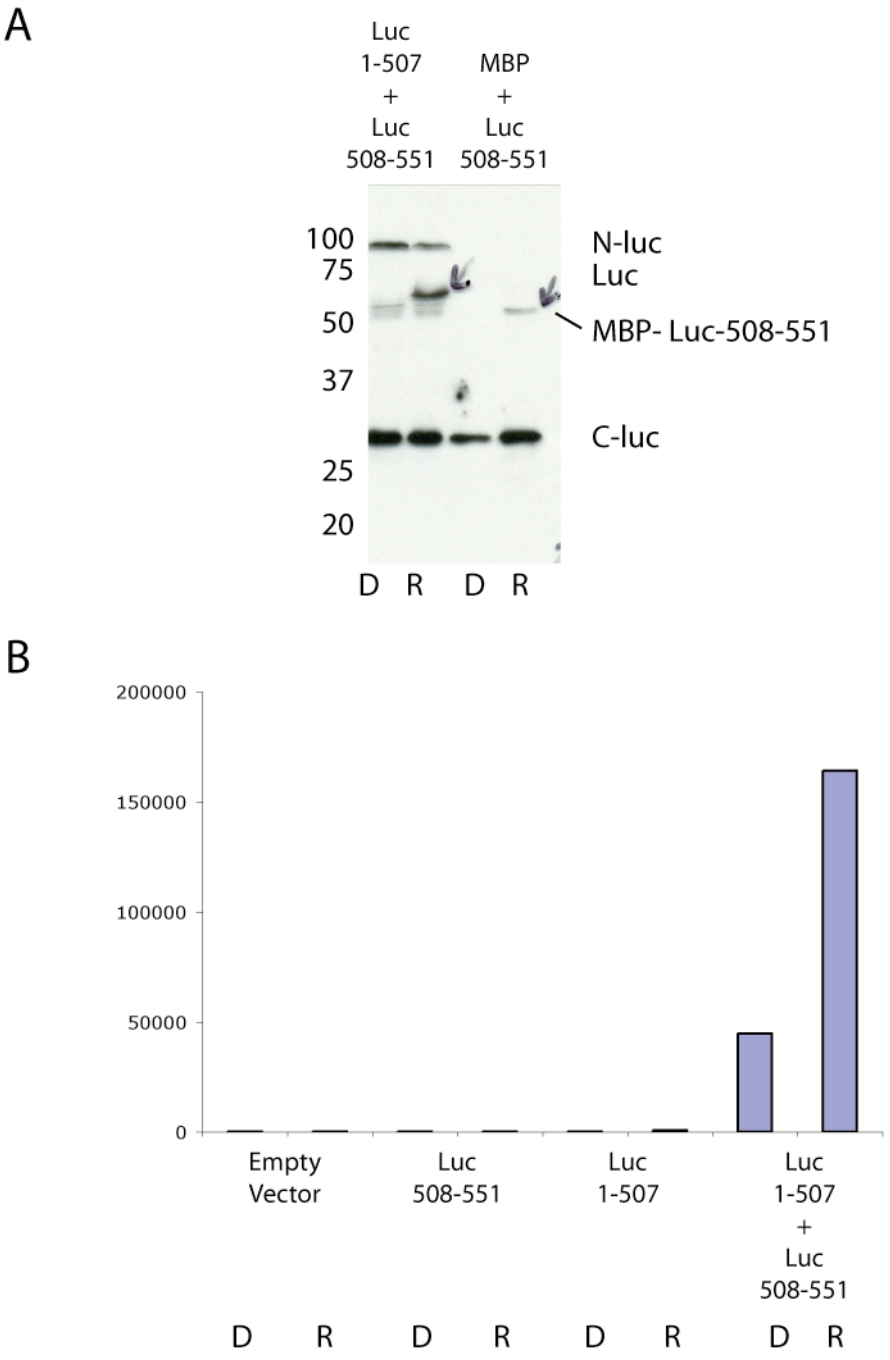


Figure 2-13 : Splicing at Luciferase residue 508.

In addition to the 3 splice sites located in surface loops tested, one that was located in a surface helix (S298) and two that were located in the hydrophobic core of the protein (C81, C216) were also generated and all failed. This is despite the fact that in some cases, the expression of both fragments could be detected by western blot (**Figure 2-14**). Interestingly, while in no case could splicing be detected by western blot, the C216 constructs displayed luciferase activity that was not dependent on the presence of rapamycin. This may be the result of one of the fragments retaining some activity. Alternatively, the two pieces may have an affinity for one another and may be able to spontaneously complement in *trans*.

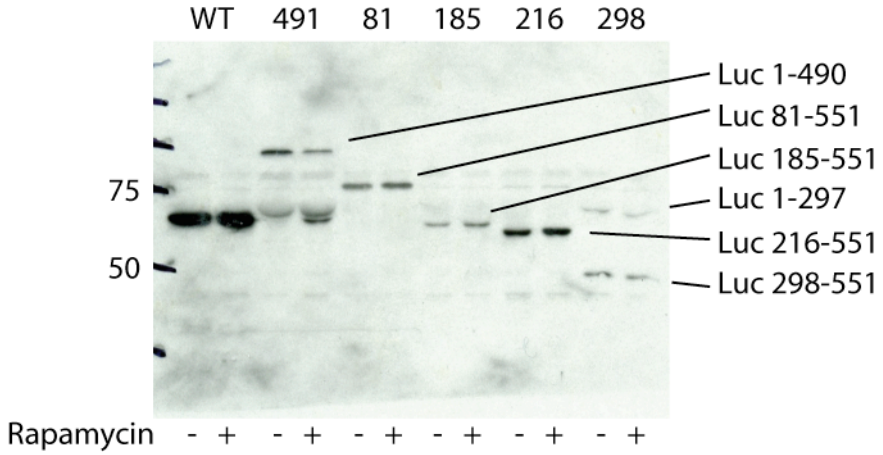


Figure 2-14 : Failed splicing at 4 split sites in luciferase. S2 cells were transfected with constructs encoding CPS pairs split at the indicated residue and treated as described in **Figure 2-4B**. Full length luciferase (WT) and the N-luc/C-luc CPS pair (491) used previously in this section are included as controls. No splicing could be detected in any of the alternative splicing sites.

The original reason for attempting to use CPS in *Drosophila* was our desire to apply CPS to the study of circadian rhythms. Circadian rhythms in *Drosophila* are regulated by a series of interlocking feedback loops which result in a tightly defined rise and fall in the level and activity of a number of proteins which in turn give rise to stereotyped 24 hour rhythmic sleep-wake cycle. Although genetic techniques have allowed the elucidation of many of the participating proteins as well as provided insight into how they interact to give rise to circadian rhythmicity, the lack of speed, tunability and reversibility of genetic methods have left many questions about circadian rhythms unanswered [57].

We reasoned that CPS would be ideal for the study of circadian rhythms. The speed with which it works would allow the manipulation of the cycling proteins faster than they naturally rise and fall. For example, PERIOD or TIMELESS could be activated out of phase with the natural cycle. The resulting phenotype would allow the determination of which, if any of these proteins is the actual “pace setter” of circadian rhythms. Ideally, a non-natural sleep-wake cycle could be forced through the use of rapamycin, allowing elucidation of the molecular and cellular events that precede the sleep or wake phase.

Alternatively, a number of the proteins involved, notably DOUBLETIME(DBT) [90, 91] and SHAGGY(SGG) [92] are necessary for normal development. Null mutations of DBT and SGG are lethal, making it difficult to specifically study their roles in the regulation of circadian rhythms. Since active protein is only generated in the presence of rapamycin, CPS could be used to

generate a delayed knockout of DBT or SGG. Rapamycin, or more likely, rapamycin analog could be provided during larval growth but withheld from adult flies. Presumably, drug metabolism and protein turnover would then result in an adult knockout of SGG or DBT allowing a more detailed study of the role that sgg and dbt play in circadian rhythms.

Despite the difficulty in locating additional splicing sites in luciferase, we attempted to generate a CPS allele of DBT. Unfortunately, the structure of DBT has not yet been determined making the design of a splicing site difficult. DBT is a two domain protein [91]. The N-terminal domain is the fly homolog of casein kinase, of which a crystal structure has been generated and a C-terminal domain that bears no homology to any other sequence in the NCBI database (protein BLAST search http://www.ncbi.nlm.nih.gov) and is characterized by repetitive stretches of alanine, glycine and glutamine. Compounding the difficulty in designing a splice site was the lack of a convenient and reliable *in vitro* or cell based assay for DBT function. Since we would not be able to test the effects of any mutation made without generating a transgenic fly, we were reluctant to make many mutations, especially in the kinase domain in which many functional mutations have been isolated.

We chose 6 sites to attempt to splice DBT. They fall into 2 categories – native cysteines that did not require mutation (C96, C267), and residues following a glycine that were mutated to cysteine (F282C, A311C, S352C, S376C, **Figure 2-15**). Of these, only one (C96) was within the core of the kinase domain.

Unfortunately, in no case could full length DBT be generated through splicing, as analyzed by western blot (**Figure 2-16**). Interestingly, the C-terminal fragment of the C96 splice site appears to be able to splice with the N-luc fragment used for the *Drosophila* work, much like some of the luciferase fragments that could not splice with each other could splice with MBP or a Hisx6 tag.

Figure 2-15 : DOUBLETIME. The protein sequence of DOUBLETIME (DBT). Shown in blue is the region that shows sequence homology to casein kinase. Shown in red is the region that has no homology to any sequence in the NCBI database. Residues at which splicing was attempted are shown in bold, black letters.

In hindsight, DBT may have been a poor choice as the second target of CPS. The lack of structural information has already been mentioned. Additionally, the C-terminal domain, which is required for DBT's role in the regulation of circadian rhythms may have also been interfering with splicing. The C-terminal domain's sequence is unlike anything else in the NCBI database. It likely has little to no secondary structure which may affect solubility and may be involved in protein-protein interactions which may affect the localization of the

CPS construct. Perhaps in the future, when more is known about the rules governing CPS sites or when a structure of DBT is available, a CPS allele of DBT can be successfully designed.

Figure 2-16 : Attempts to generate a conditional allele of DBT. S2 cells were transfected with the indicated CPS constructs. All CPS constructs were generated as N-terminal MBP fusions for Western blotting purposes. Following overnight expression, cells were treated with 0.1% DMSO or 100nM rapamycin for 24 hours. Cells were then harvested and analyzed by Western blot (anti-MBP or anti-DBT as indicated). Spliced DBT could not be detected in any of the samples.

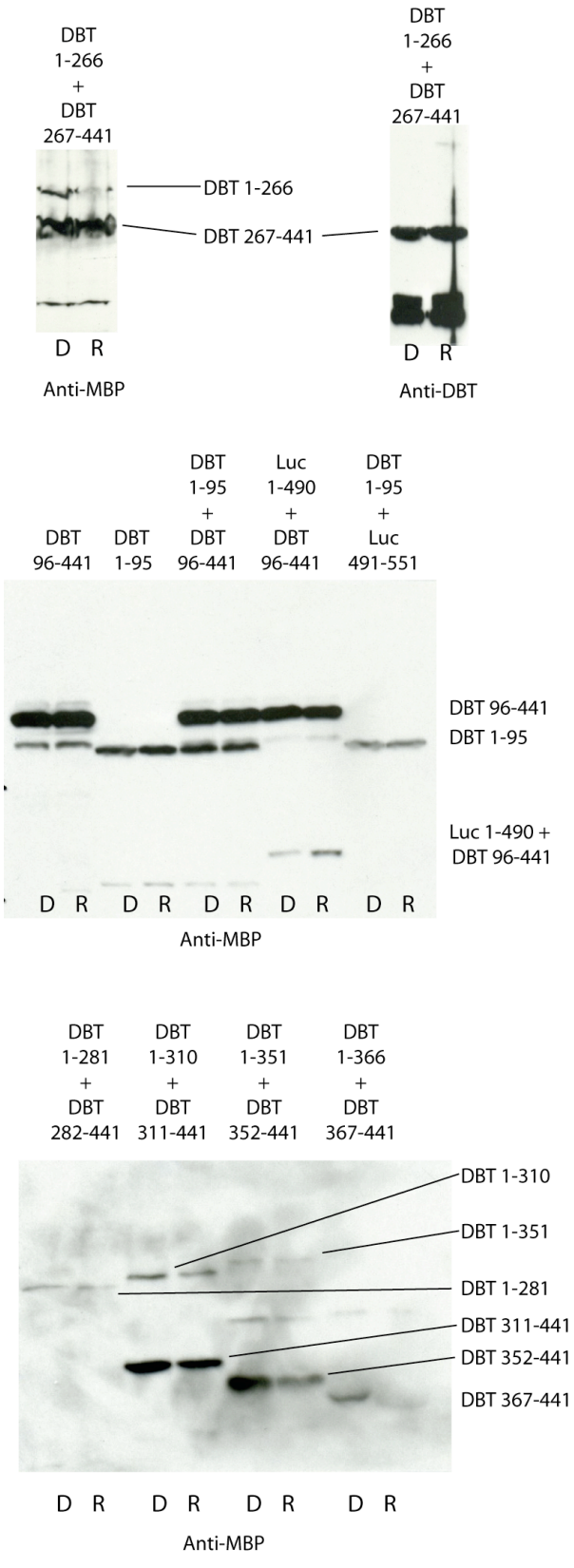


Figure 2-16 : Attempts to generate a conditional allele of DBT.

Section 2.2.3 : Improving CPS for the generation of conditional protein alleles

It is fairly clear that, while CPS is very powerful when successfully applied, it is difficult to apply it to new proteins. One approach that we proposed to alleviate this difficulty was to generate a method to screen rapidly for new splicing sites. Thus, rather than designing, cloning and testing each splice pair individually, one could start with a library of split protein pairs and simply screen for ones that exhibit rapamycin dependent splicing. Unfortunately, one cannot simply generate an N-terminal fragment library and a C-terminal fragment library and then mix them together for the screen since this would result in the vast majority of splicing "hits" not generating wild type protein. Indeed, our suspicion is that the majority of CPS partners found this way would be the result of splicing a small peptide from the N-terminus of the protein to a small peptide from the C-terminus. Thus, the library must be generated in a manner that the N-terminal fragment is genetically linked to the C-terminal fragment that encodes the remainder of the protein.

There are two potential methods for creating a library of CPS constructs to screen for splicing – a transposon based screen and a circular permutation based screen (**Figure 2-17 and Figure 2-18**). Neither is perfect – the transposon method necessarily adds flanking residues that are part of the

transposon itself, and the gel sizing step in the circular permutation method will result in proteins that are only approximately full length. Nonetheless, at the very least, the screens should be capable of identifying regions of the protein that are amenable to intein insertion. Given that the two luciferase splicing sites that worked were only 17 residues apart, finding such a region may be all that is needed to generate a CPS protein allele that results in wild type protein simply by removing or replacing the residues that are inserted or deleted in the library cloning.

Although the generation of a screen would greatly streamline the application of CPS to new systems, screening for CPS sites in a new protein could still take a long time and ultimately be unsuccessful. Unfortunately, it appears that there is and always will be a degree of sequence dependence in CPS which may not always be overcome. As such, we turned to a new approach which does not require splicing and is described in the next chapter.

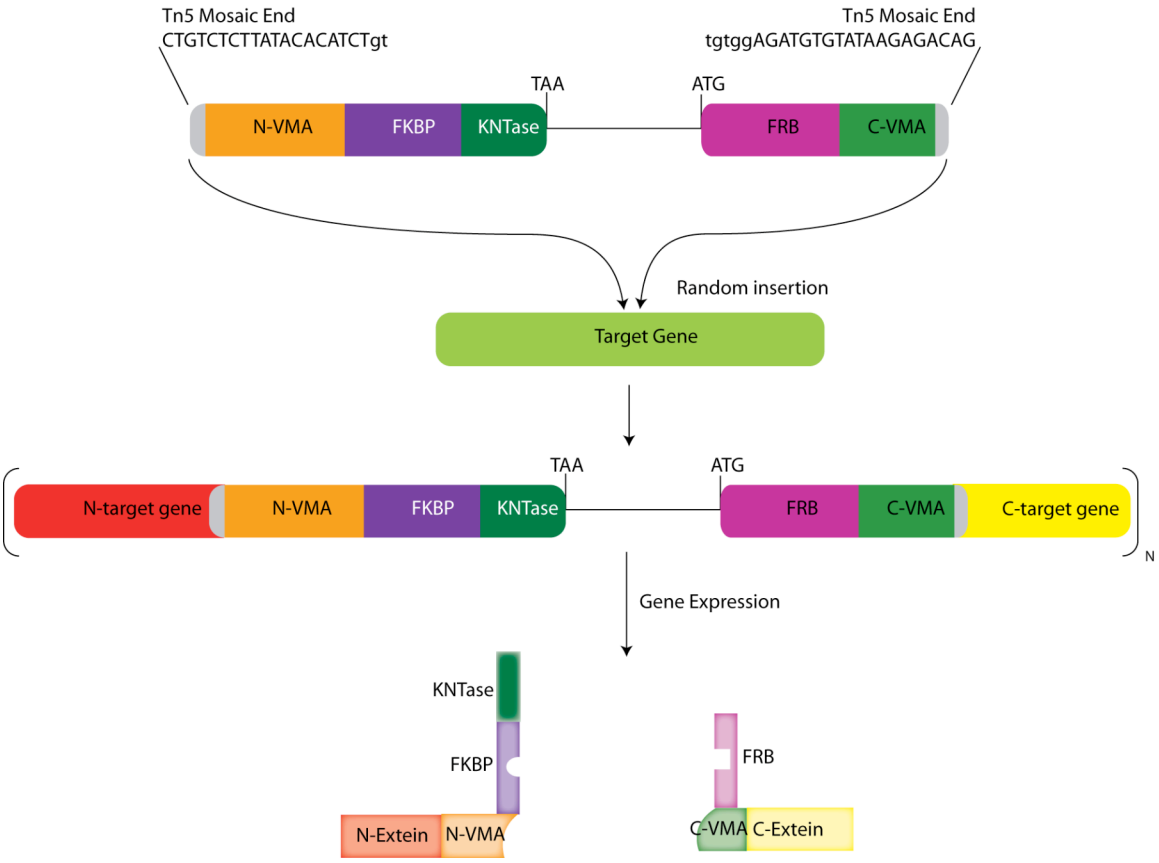


Figure 2-17 : A transposon screen for splicing sites. A high copy plasmid encoding for ampicillin resistance and the target gene will be mutagenized *in vitro* using the transposon CPS-Tn5. Since Kanamycin nucleotidyltranferase (KNTase) is present as a C-terminal fusion, it will be correctly translated and confer kanamycin resistance only if N-VMA/FKBP is inserted in frame into a gene. Transformed bacteria will also acquire kanamycin resistance if the transposon inserts into the ampicillin resistance marker. Thus, transformants will be selected for both ampicillin and kanamycin.

Figure 2-18 : A circular permutation screen for CPS sites. In a circular permutation screen for CPS accessible sites, the target protein would first be cloned as a head to tail permutation. DNA ligation will result in a circularized gene which is then randomly cut under single hit conditions with DNAse. A gel sizing step then removes genes that have been cut multiple times. The full length genes can then be inserted into a plasmid encoding the N- and C- CPS cassettes through blunt end ligation generating a library that can be screened for splicing partners.

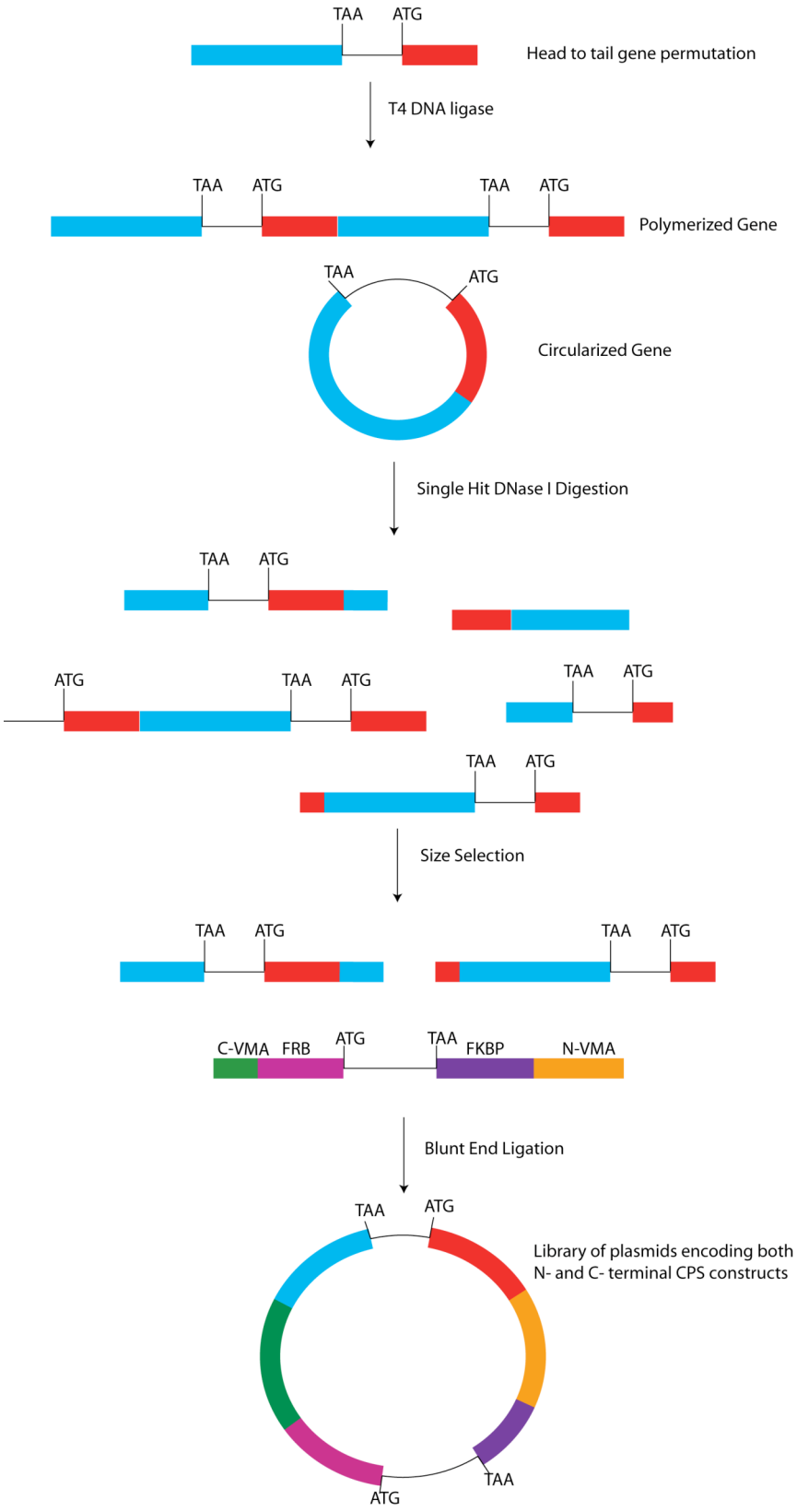


Figure 2-18 : A circular permutation screen for CPS sites.

Chapter 3 : Controlled Protein Degradation

This work was performed in collaboration with Matthew Pratt of the Rockefeller University.

Section 3.1 : Introduction

The allele of luciferase that we generated using CPS was an almost perfect conditional allele. The induction of activity was rapid and titratable, appearing within minutes and in a dose dependent manner. The activated protein was virtually wild type, with only a few point mutations. Finally, small molecule mediated protein activation was possible in a variety of contexts – cultured cells, both mammalian and insect as well as living *Drosophila*. However, we were unable to generate a conditional allele of another protein using CPS, bringing the generality of CPS into question.

Taking advantage of this property of ubiquitin, we turned to an alternative strategy that takes advantage of endogenous cellular processes, by placing the stability of a protein of interest under pharmacological control. Other work in this area was discussed in Section 1.4.

We reasoned that an ideal degradation-mediated technology would result in the release of a native protein from the degradation signal (or "degron") following small molecule rescue of a chimeric protein-degron fusion. To achieve release from the degron, we somewhat counter intuitively turned to a known mediator of protein degradation, ubiquitin. Ubiquitin plays several roles in both

protein trafficking and stability, primarily through its posttranslational conjugation to lysine side chains of target proteins [93]. Unlike this branched ubiquitin structure, linear ubiquitin fusions are known to be translational products where the N-terminal ubiquitin moiety is rapidly and specifically cleaved from its fusion partner by dedicated ubiquitin proteases [94]. These proteases have broad specificity for the residues following ubiquitin, with proline being the only amino acid resistant to cleavage. Johnsson and Varshavsky have utilized this feature in a two-hybrid technology based on complementation and subsequent cleavage of a genetically split ubiquitin [16]. Briefly, ubiquitin was split into N- and C-terminal fragments, corresponding to residues 1-37 and 35-76 respectively. When a point mutation (I13G/A) was also introduced, the ubiquitin fragments would only complement and fold, allowing for cleavage of fusion constructs, when a dimerization signal brought them into close proximity.

Thus, as an alternative to CPS, we developed a new posttranslational, small molecule-mediated, technology for the manipulation of protein function. This system, termed SURF (Split-Ubiquitin for the Rescue of Function), places the complementation of ubiquitin under the control of the FKBP/rapamycin/FRB heterotrimerization system (**Figure 3-1**). Before complementation a protein of interest is targeted for destruction by the proteasome through the introduction of an N-terminal degron. Small molecule induced dimerization results in ubiquitin complementation and folding followed by cleavage of the protein of interest, thereby releasing it from the degron and rescuing its function.

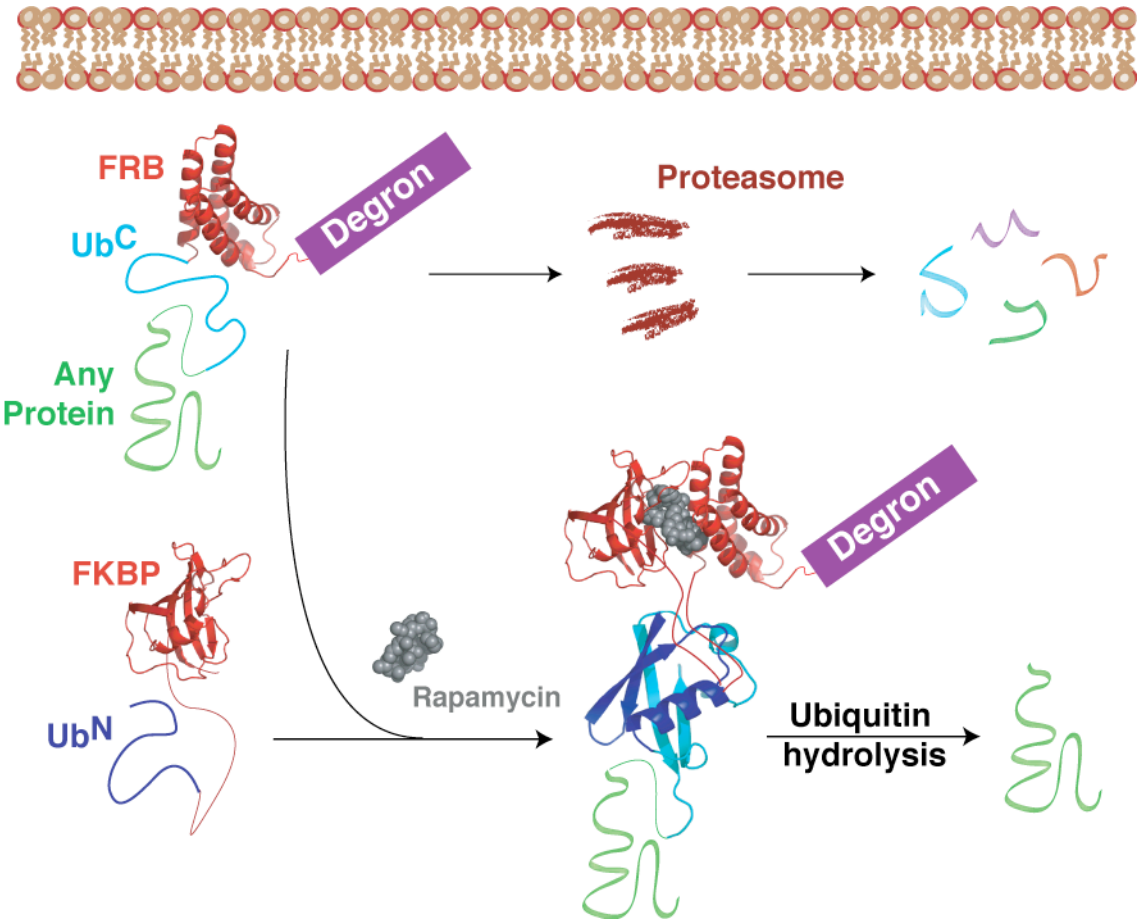


Figure 3-1 : Overview of the SURF technology. A protein of interest is genetically fused to a degron that destines the protein for destruction by the proteasome. In between the protein under investigation and the degron are FRB and the C-terminal fragment of ubiquitin (Ub^C). Upon addition of rapamycin a complex is formed between the small molecule and another engineered protein containing FKBP and the N-terminal ubiquitin fragment (Ub^N). This complex creates a shunt away from degradation by allowing dimerization and complementation of the ubiquitin moiety through interactions between FKBP and FRB. After ubiquitin folding, a protease releases the protein of interest, effectively rescuing it from destruction.

Section 3.2 : Rescue of destabilized luciferase by ubiquitin hydrolysis

For this strategy to be successful, we needed a degron that was powerful enough to keep the level of background activity low, but not so powerful that it would work faster than ubiquitin hydrolysis could rescue the target protein. Two general protein constructs were generated (**Table 3-1, Figure 3-2**). The first construct (**1**) contains maltose-binding protein (MBP) fused to the conditional complementation mutant of the N-terminal fragment of ubiquitin (I13A, residues 1-37) followed by FKBP. The second construct, bearing the protein of interest, has the following architecture: a degron, followed by FRB, the C-terminal fragment of ubiquitin (residues 35-76), and the protein under investigation. Initially we chose firefly luciferase as our model system, due to the ease and sensitivity of detection. We chose three degrons, which we theorized would have different degradation kinetics, to examine the gross requirements for protein rescue. The first degron (**2**) is a fragment (residues 1-100) of the protein SopE from *Salmonella typhimurium*, which has been shown to reduce the cellular half-life of proteins to which it is fused[58]. The two other degrons (**3**,**4)** are FRB mutants originally designed to allow the use of non-toxic derivatives of rapamycin. FRB* contains the three point mutations K2095P, T2098L, and W2101F and has already seen utility as a degron [30]. The other FRB derivative, which was used in the CPS work from the previous chapter, is a single point mutant (T2098L) and is here referred to simply as FRB. Furthermore, we chose

to use three copies of these FRB derivatives to increase the avidity for the FKBP/rapamycin complex, thus driving complementation. Additionally, the fusion of three copies of FRB to the C-luc fragment used in the previous chapter resulted in poor expression, suggesting that three copies of FRB may act as a degron.

HeLa cells were transfected with the complementary pairs of the above constructs and treated with rapamycin (100 nM) or DMSO vehicle for 36 h. Cleavage from the degron and rescue of luciferase function were then measured by western blotting and luminescence activity, respectively (**Figure 3-2**). When attached to the strong degron SopE (**2**), luciferase activity could not by rescued by the addition of rapamycin and no cleavage product was observed. To ensure that **2** was being expressed, cells were treated with the proteasome inhibitor ZL_3VS [95], and luciferase activity was indeed observed (**Figure 3-3**). From this we concluded that the kinetics of degradation of SopE fusions must be faster than the kinetics of rapamycin induced complementation and/or ubiquitin cleavage. Unlike the SopE fusion, detectable levels of constructs **3** and **4** were observed even in the absence of rapamycin (**Figure 3-2**). Surprisingly, three copies of FRB resulted in a more robust degron than the corresponding FRB* construct, even though a previous study would suggest the opposite, at least for a single copy [30]. Upon addition of rapamycin, cleavage of luciferase from both constructs **3** and **4** was efficient, with **4** giving a better induction of activity.

Therefore, the general architecture of construct **4** was used as a blueprint for all subsequent SURF controlled proteins.

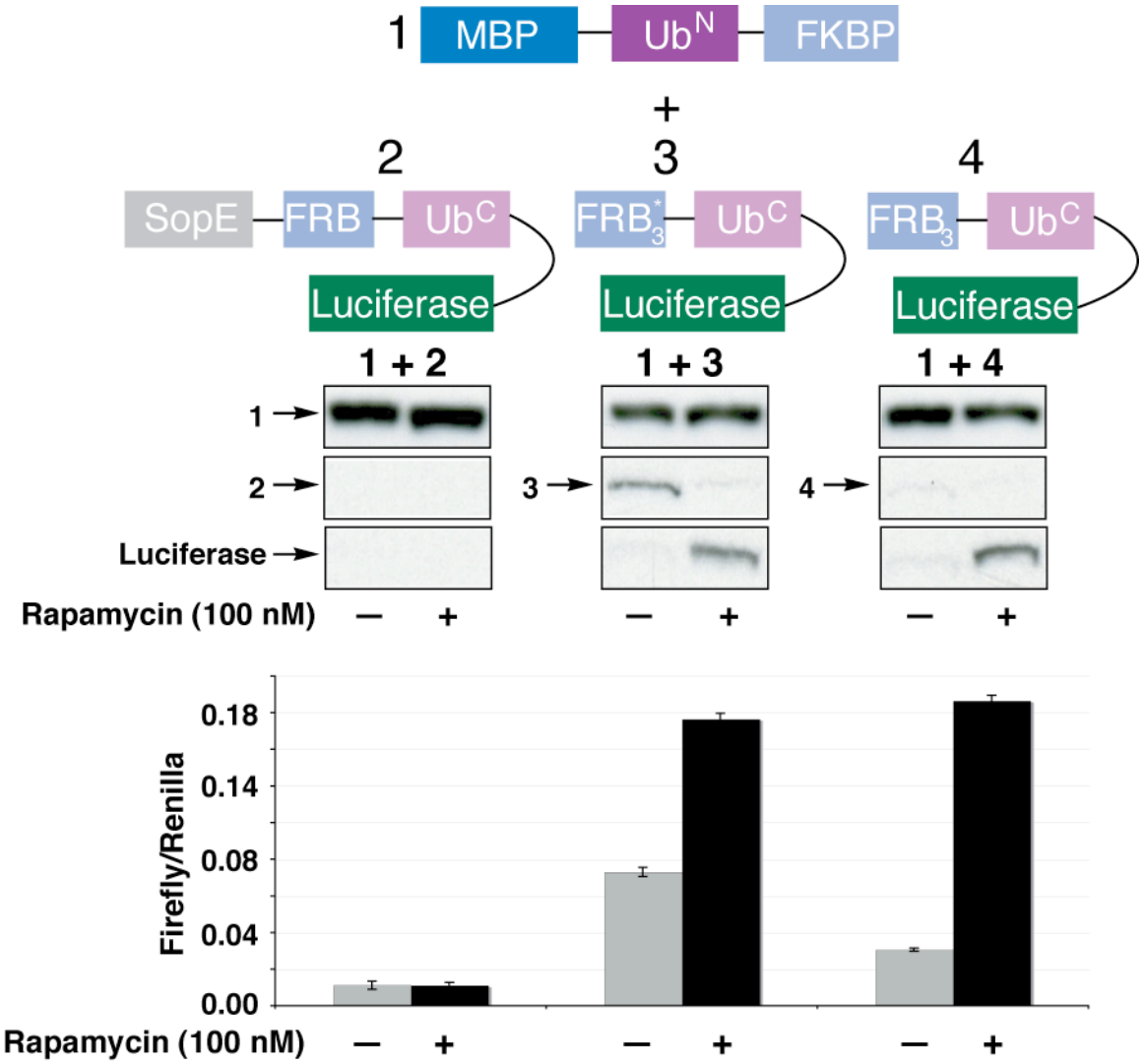


Figure 3-2 : Evaluation of degrons. Constructs **1** and **2**, **3**, or **4** were transfected into HeLa cells and the ability of rapamycin to rescue luciferase activity was measured by western blotting (anti-HA) and luciferase activity. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate.

Table 3-1 : Architecture of SURF constructs

Construct	Sequence
MBP-UbN-FKBP (**1**)	MBP-EFEGGST-Ubiquitin (I13A, 1-37)-GGSTMAAA-FKBP
SopE-FRB-UbC-Luc (**2**)	SopE (1-100)-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase
FRB_3*-UbC-Luc (**3**)	FRB*-ASLEGSTCT-FRB*-ASLEGSTCT-FRB*-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase-HA tag
FRB_3-UbC-Luc (**4**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase-HA tag
FRB_3-UbC-Casp3 (**5a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Caspase-3-HA tag
FRB_3-UbC-Casp3 (**5b**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-Caspase-3-HA tag
FRB_3-UbC-vSrc (**6a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-vSrc-HA tag
FRB_3-UbC-vSrc (**6b**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-vSrc-HA tag
FRB_3-UbC-Smad3 (**7a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-HA tag-Smad3
FRB_3-UbC-Smad3 (**7b**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-HA tag-Smad3

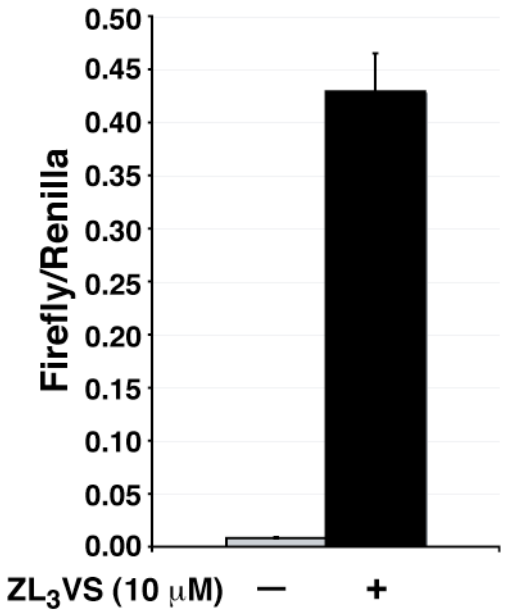


Figure 3-3 : Rescue of SopE induced degradation with a proteasome inhibitor. Cells were transfected with constructs **1** and **2** and treated with the proteasome inhibitor ZL_3VS [95] for 6 h. At this time expression of **2** was measured by luciferase activity.

To determine the kinetics of rescue, HeLa cells were transfected with **1** and **4**. After 24 h, rapamycin (100 nM) or DMSO was added to the cells and luciferase activity and cleavage were measured at different time points (**Figure 3-4A**). Appearance and activity of luciferase were observable in as little as 2 h and grew until a plateau was reached after 24 h of treatment. These kinetics are governed by the rate of protein synthesis in the cell and compare well with other degradation technologies [30, 32].

To explore the effects of varying rapamycin concentration, we transfected HeLa cells with our SURF constructs and treated them with different concentrations of rapamycin for 36 h (**Figure 3-4B**). Both luciferase activity and

western blotting analysis confirmed that the amount of protein rescue could be altered by simply changing the rapamycin concentration. The rapamycin mediated interaction of FRB and FKBP can also be reversed through the addition of another small molecule, ascomycin. We treated HeLa cells expressing the SURF constructs with rapamycin (10 nM) and differing amounts of ascomycin (**Figure 3-4C**). Western blotting analysis and luciferase assays clearly show that ascomycin is able to competitively reduce the levels of rescue. Therefore, more precise control over levels of protein rescue can be achieved through the use of combinations of rapamycin and ascomycin concentrations. Additionally, by washing away rapamycin and adding ascomycin the SURF system should be effectively turned off. However, unlike other destabilized protein technologies where the protein of interest remains fused to the degron, the rate of disappearance of rescued protein will be dependent upon its intrinsic cellular stability.

Figure 3-4 : Characterization of the SURF system. *(A)* HeLa cells were transfected with constructs **1** and **4** and rapamycin was added. Luciferase cleavage and activity was determined by western blotting (anti-HA) and luciferase activity at different time points. *(B)* Cells were transfected and analyzed as in (A) to determine the effect of treatment with differing amounts of rapamycin. *(C)* Cells were transfected and analyzed as in (A) to determine the effect of treatment with rapamycin and different amounts of ascomycin. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate.

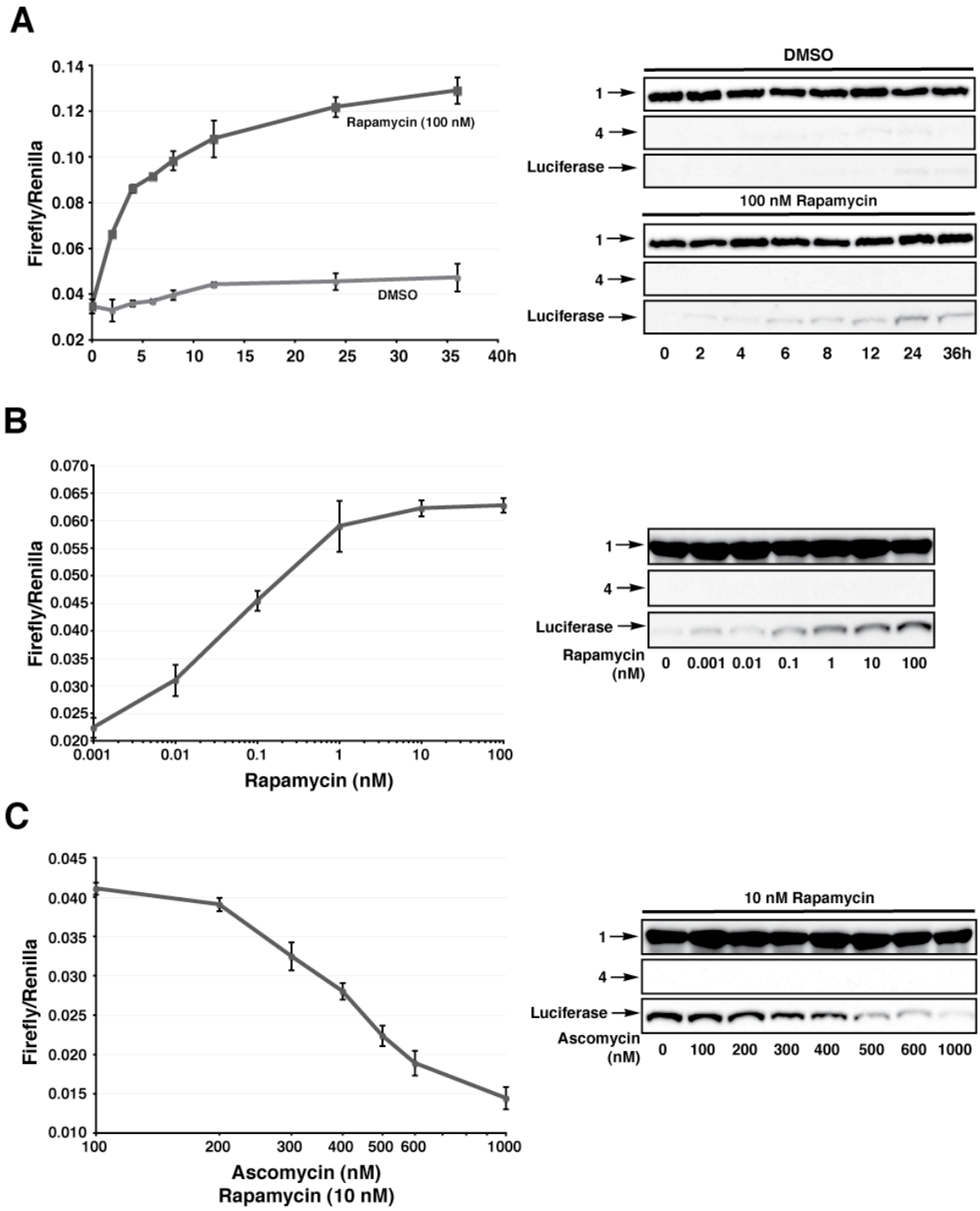


Figure 3-4 : Characterization of the SURF system.

Section 3.3 : Application of SURF to other protein targets

We next explored the generality of SURF by applying it to members of three important classes of proteins: a protease, a kinase, and a transcription factor.

Section 3.3.1 : SURF control of Caspase-3

We chose caspase-3 as a candidate for SURF because, as a major executer of apoptosis, it is an important factor in cancer biology, with many types of cancer resistant to its effects [59, 96]. Although small molecule activators of apoptosis are available, a convenient method to titrate caspase-3 levels does not exist. Therefore, SURF control of caspase-3 may help illuminate the roles of this protease.

Towards this end, we generated a SURF construct containing human procaspase-3 (**5a**), as well as a mutant thereof (**5b)** bearing a proline as the residue immediately following ubiquitin, thereby precluding ubiquitin hydrolysis (**Table 3-1**). We hypothesized that **5b** would be stabilized and avoid degradation utilizing a mechanism similar to previously described systems. Constructs **1** and **5a** or **5b** were then expressed in HeLa cells in the presence or absence of rapamycin (100 nM) and the rescue analyzed by western blotting (**Figure 3-5A**). Procaspase-3 was rescued with similar efficiencies in both the cleavage (**5a**) and stabilization (**5b**) constructs. As a qualitative measure of the ability of these

inducible caspase-3 systems to actively partake in an apoptotic cascade, HeLa cells expressing these SURF constructs were treated with rapamycin (100 nM) for 24 h to allow adequate rescue. At this time the cells were either treated with the apoptosis activator Staurosporine or the vehicle DMSO for 4 h and the extent of proteolysis, and procaspase processing, was interrogated (**Figure 3-5B**). Cleaved procapsase-3 was efficiently proteolyzed as judged by a significant decrease in the intensity of the procaspase-3 band. Stabilization construct **5b** was also processed upon staurosporine treatment; however, the level of proteolysis appears to be somewhat reduced when compared to cleavage construct **5a**.

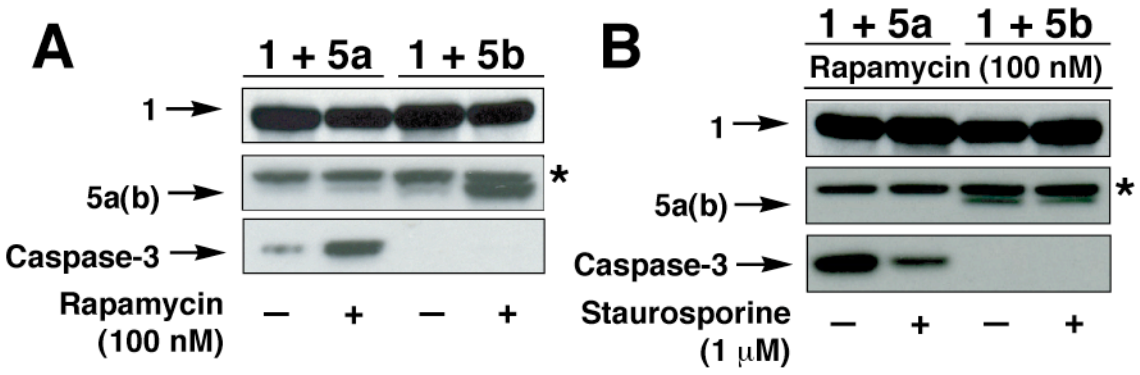


Figure 3-5 : SURF can rescue the protease Caspase-3. HeLa cells were transfected with **1** and either a cleavage competent SURF construct **5a** or incompetent mutant **5b**. *(A)* The capacity for rapamycin rescue was then examined by western blotting (anti-HA). *(B)* Processing of cleaved and stabilized caspase-3 during stauroporine induced apoptosis was analyzed by western blotting (anti-HA). * indicates a background band.

Section 3.3.2 : SURF control of v-Src

Rous Sarcoma virus Src (v-Src) is a constitutively active tyrosine kinase endowed with cellular transformative activity [97]. v-Src is a quintessential oncogene and the first tyrosine kinase discovered, and it's cellular homologue (c-Src) is involved in many cellular signaling pathways, many of which are still not fully understood [98]. Thus, we chose v-Src as a model protein to explore SURF's utility for controlling kinase activities. We generated a v-Src SURF construct and the corresponding stabilization mutant, **6a** and **6b** respectively (**Table 3-1**). Constructs **1** and either **6a** or **6b** were coexpressed in SYF cells, which lack endogenous cellular Src kinase [99], and treated with rapamycin (100 nM) or DMSO vehicle (**Figure 3-6A**). In the presence of rapamycin both the cleaved and stabilized forms of v-Src were efficiently rescued from degradation.

The activity of these v-Src proteins against a wide array of substrates was then analyzed by a global anti-phosphotyrosine western blot (**Figure 3-6B**). Despite the near complete destruction of full-length **6a** and **6b**, two bands reacted strongly with the phosphotyrosine antibody in the absence of rapamycin. Furthermore, these bands were not present in mock-transfected cells. The full-length v-Src constructs (**6a** and **6b)** could become self-phosphorylated and contribute to this signal. However, because of the intensity of these bands, it seems likely that additional phosphorylated proteins are present and account for most of the signal. While the identities of these phospho-proteins are currently

unknown, we speculate that they may be members of the ubiquitin-proteasome pathway involved in the recognition and trafficking of our degron. Additional studies will be needed to test this idea.

More important for the current study is that when treated with rapamycin, the cleavage construct showed rescue of v-Src function against a variety of cellular targets. Furthermore, the pattern of phosphorylated proteins compares well to cells transfected with v-Src alone, albeit with much less intensity due to much larger amounts of v-Src in v-Src transfected cells (data not shown). The stabilization construct **6b** also showed rescue of v-Src activity. However, interestingly this construct appears to phosphorylate a much smaller pool of substrates (compare lanes 2 and 4, **Figure 3-6B**). Thus the activity of v-Src appended to the degron is distinct from the cleaved v-Src.

To further determine the functional consequences of our v-Src constructs, we performed a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) cell proliferation assay [100]. SYF cells were transfected with **1** and **6a** or **6b**, empty vector, or a plasmid containing v-Src. After treatment with rapamycin (100nM) for 24 h, the cell proliferation rates were determined. Rescue of v-Src from the cleavage construct **6a** resulted in a similar proliferation rate as cells expressing v-Src alone (**Figure 3-7**). In contrast, exposure of the stabilization mutant **6b** did not result in an increase in cell proliferation over mock-transfected cells.

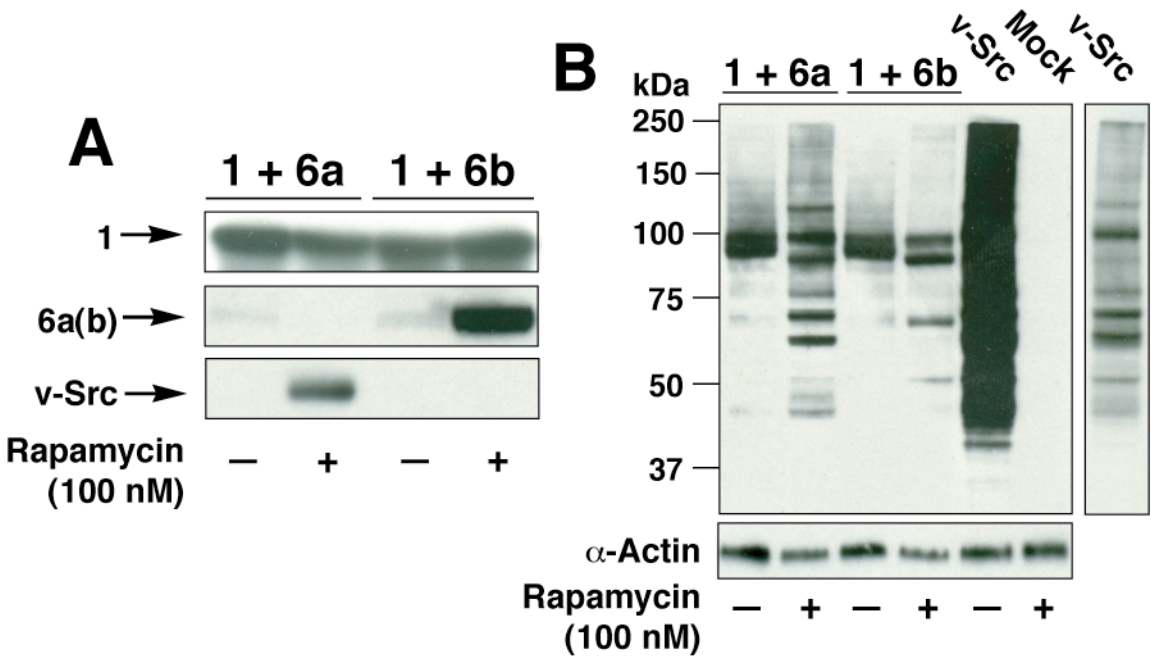


Figure 3-6 : SURF can rescue the tyrosine kinase v-Src. SYF cells were tranfected with **1** and either the cleavage construct **6a** or stabilization mutant **6b**. *(A)* Rescue of v-Src was explored by western blotting (anti-v-Src). *(B)* Recovered v-Src activity was examined by a global tyrosinephosphate western blotting (anti-pY). Cells transfected with wild-type v-Src express much higher levels of v-Src than SURF transfected constructs resulting in a very high level of cellular phosphorylation ('v-Src' lane). A shorter exposure is shown on the far right for comparison to the SURF induced v-Src (**1 + 6a**).

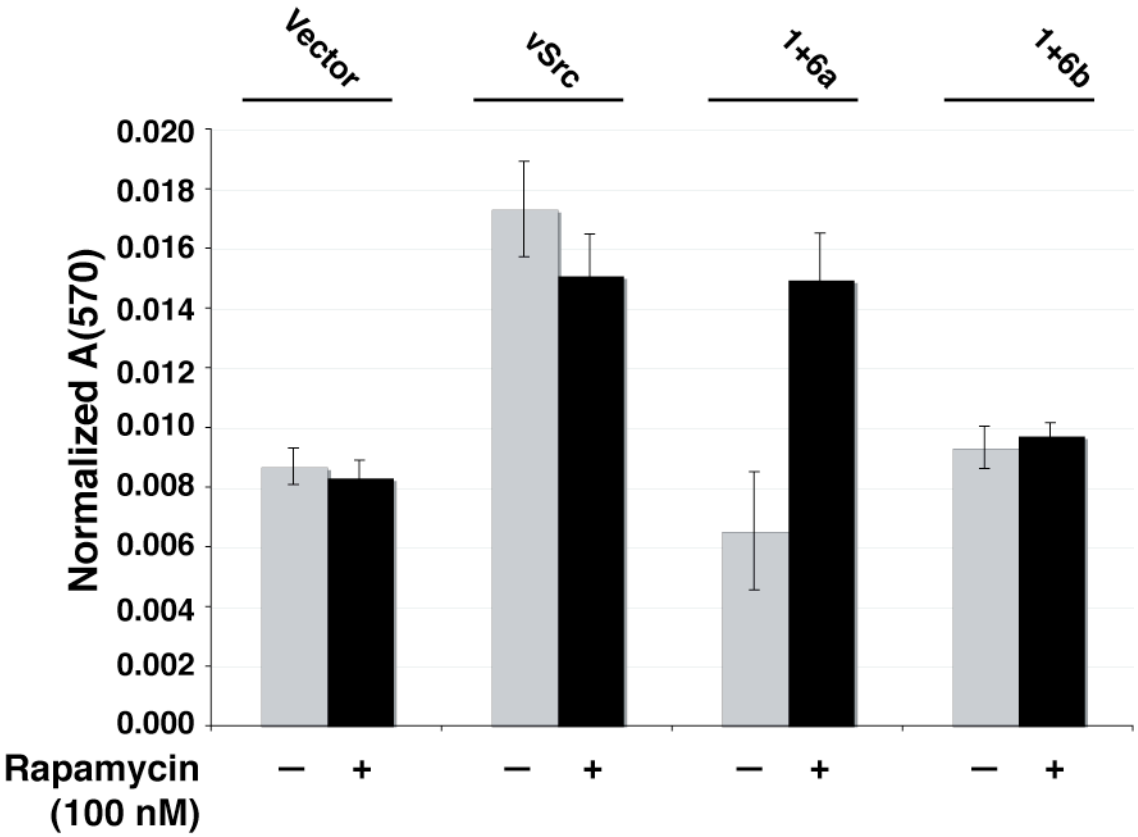


Figure 3-7 : MTT cell proliferation assay. SYF cells were tranfected with **1** and either the cleavage construct **6a** or stabilization mutant **6b** or the appropriate control plasmids. The cells were then treated with Rapamycin or DMSO vehicle for 24h. Cell proliferation was then quantified in triplicate by absorbance of reduced 3(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazonlium bromide (MTT) at 570 nm.

Section 3.3.3 : SURF control of SMAD3

Smad3 is a latent transcription factor involved in the transforming growth factor β (TGFβ) signaling pathway [101, 102]. In the basal state, Smad3 is unphosphorylated and primarily cytoplasmic. Activation of the TGFβ receptor (which is composed of two receptor kinases TβR-I and TβR-II) results in phosphorylation of Smad3 on two serine residues at its extreme C-terminus. Phospho-Smad3 then accumulates in the nucleus where it forms heterotrimers with a related protein, Smad4, and regulates gene transcription. Activated TGFβ receptors are also involved in Smad-independent signaling cascades[102]. Thus, stimulation of cells with TGFβ can lead to pleiotropic signaling effects making it difficult to understand the precise mechanistic basis of a phenotype. Therefore, SURF control of Smad3 might provide a convenient tool to isolate the canonical TGFβ pathway from other receptor-mediated signaling cascades.

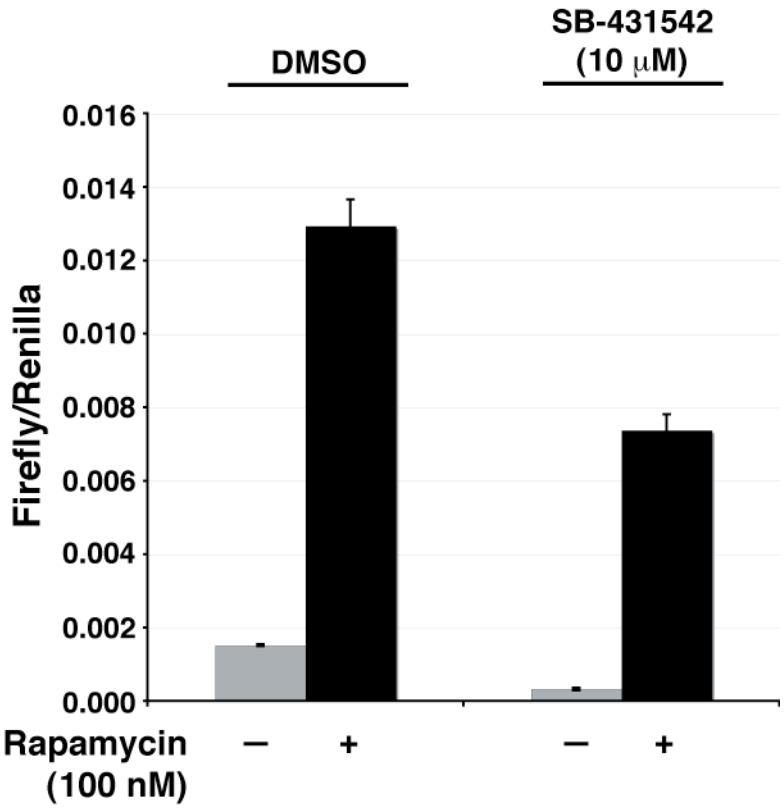


Figure 3-8 : Smad3E is a constitutive transcription factor. HeLa cells were pre-treated with the TβR inhibitor SB-43152 (10 μM) or DMSO vehicle for 16 h, and were subsequently transfected with a plasmid containing Smad3E or vector. After 24 h, luciferase production from a 3TP-luciferase plasmid was measured. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate.

In preliminary studies, we generated a constitutively active version of Smad3 in which the two phospho-serine sites were replaced by glutamic acids. Based on structural studies [103], this double mutant was expected to mimic the acidic bis-phosphate surface in the active Smad3. Indeed, this construct was shown to drive Smad-dependent transcription in cultured cells in a receptor-

independent manner (**Figure 3-8**). We then generated the SURF construct **7a** and its stabilization proline mutant **7b** (**Table 3-1**). Treatment of HeLa cells expressing **1** and **7a** or **7b** with rapamycin (100 nM) resulted in escape from destruction for both the cleavage and stabilization constructs (**Figure 3-9A**). Rescue of the cleavage construct resulted in a ten-fold increase in transcriptional output, as measured by the production of a luciferase reporter driven by the 3TP-binding element, when compared to cells treated with DMSO vehicle. Isolation of Smad3E activity from endogenous Smads, by pre-treatment with SB43152 (10 μM), an inhibitor of the TGFβ receptor kinase[104], resulted in a fifteen to twenty fold induction upon rapamycin treatment (**Figure 3-10**). This increase is consistent with an almost complete recovery of Smad3E activity as compared to cells transfected with Smad3E alone as a positive control (**Figure 3-9B**). In contrast to the cleavage construct, stabilization of the SURF proline mutant, **7b**, yielded no increase in luciferase output, despite its cellular presence as judge by western blotting.

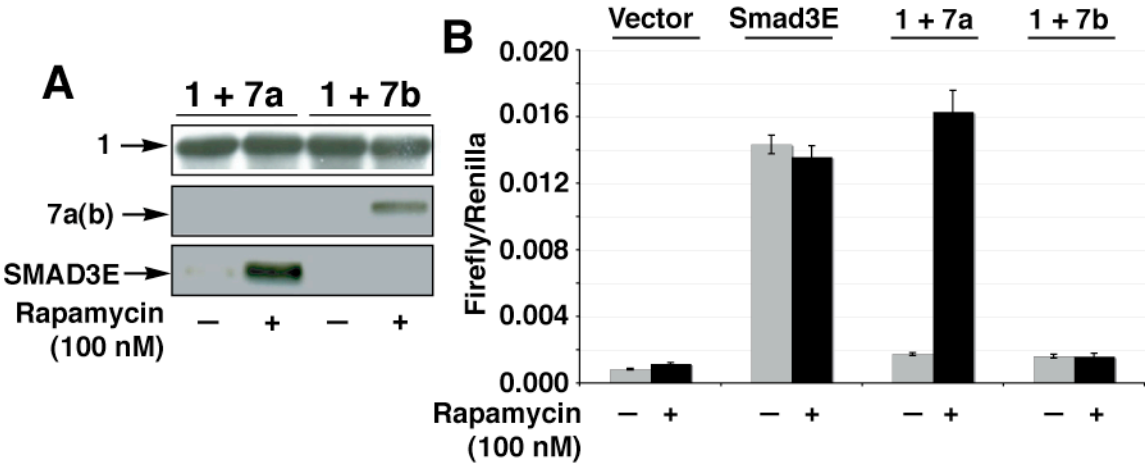


Figure 3-9 : SURF can rescue SMAD3E activity. HeLa cells were transfected with **1** and either a cleavage competent SURF construct **7a** or incompetent stabilization mutant **7b**. Rapamycin stimulated rescue was then examined by western blotting (**A**, anti-HA) and luciferase production (**B**) from a 3TP-luciferase-reporter plasmid. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate.

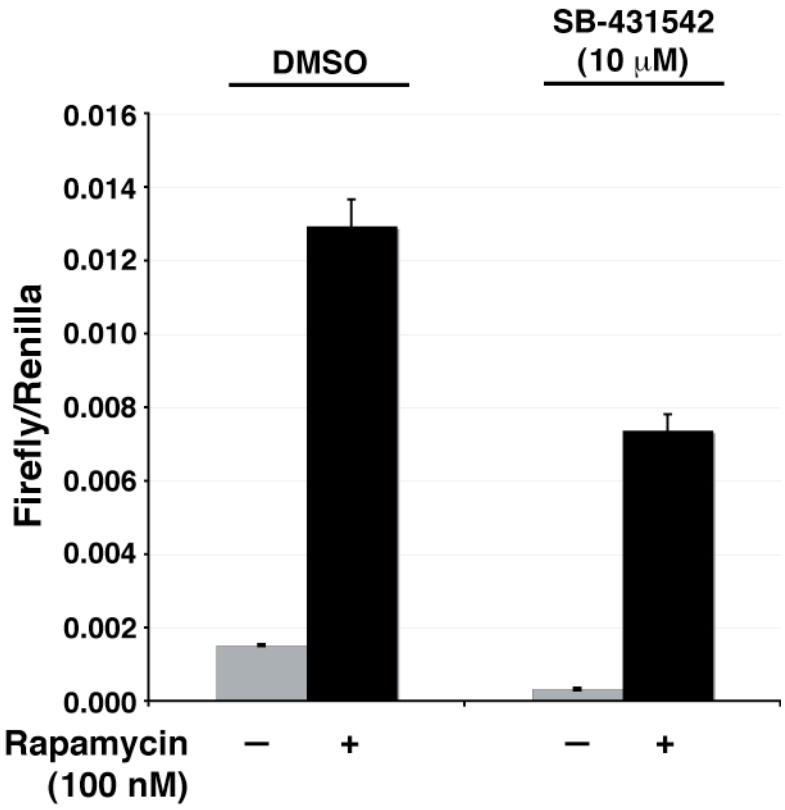


Figure 3-10 : Rescue of Smad3E in isolation from the TGFβ receptor. HeLa cells were pre-treated with SB-43152 (10 μM) or DMSO vehicle for 16 h, and were subsequently transfected with a plasmids **1** and **7a**. Luciferase production from a 3TP-luciferase plasmid was measured after treatment with rapamycin (100 nM) or DMSO vehicle for 16 h. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate.

Section 3.4 : SURF alleles can reveal mechanistic details

It has been known for almost two decades that TGFβ1 has the ability to activate its own translation and secretion [105]. In principle, SURF controlled Smad3 provides a window into the role of the canonical pathway in this feed forward mechanism since it should be possible to compare the kinetics of Smad-dependent transcription with and without autocrine stimulation of the receptor. Accordingly, HeLa cells transfected with **1** and **7a** were either left untreated or pre-treated with SB43152 (10 μM) for 16 h and then induced with rapamycin (100 nM) or rapamycin (100 nM) and SB43152 (10 μM) respectively. The extent of luciferase production from the Smad3 driven 3TP-luciferase reporter plasmid was then ascertained at different times (**Figure 3-11A**).

In cells treated with the kinase inhibitor an almost linear increase in luciferase activity is seen throughout the entire time course. Non-inhibited cells display the same linear increase at early time points, however, the rate of this increase is approximately double that of inhibited cells. One possibility for this is that Smad3 activity may affect parallel signaling pathways, thereby increasing the signal resulting from TGFβ already present in the growth media. Alternatively the difference may be an effect of treatment with rapamycin itself. It has been shown that FKBP12 binds the TGFβ type I receptors blocking their phosphorylation by TβR-II [106]. Rapamycin disrupts this interaction by blocking the binding surface on FKBP12 [107]. Both a TβR-I mutant defective in FKBP12 binding and

treatment of wild-type receptors with FKBP agonists generates modest increases in basal activity (see mock transfected cells **Figure 3-9**).

In contrast to the early time points, cells that were not treated with the kinase inhibitor exhibited a dramatic increase in luciferase activity that commenced between 18 and 24 h. This increase was not observed in cells treated with SB43152. We believe this is the result of increased secretion of TGFβ leading to autocrine receptor activation [104, 105], thereby revealing a role for Smad3 in the feed forward loop. This data agrees well with the published kinetics of TGFβ induced expression of TGFβ mRNA (~6 h) and secreted TGFβ (~16 h) [105].

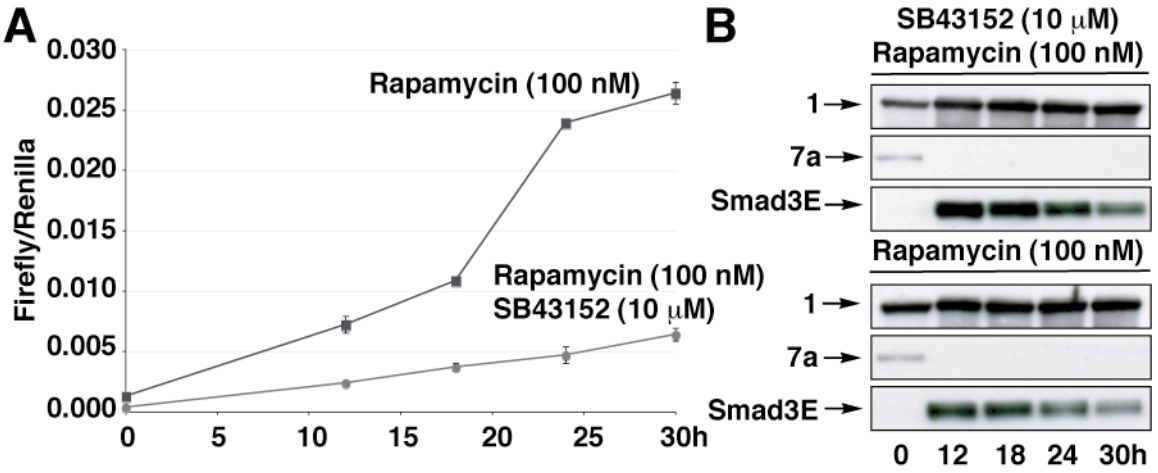


Figure 3-11 : Kinetic analysis of Smad3E transcriptional activation and rescue. HeLa cells were transfected with **1** and **7a** and treated with rapamycin or a combination of rapamycin and SB43152. (**A**) Smad3E activity was determined by luciferase activity at different timepoints. All luciferase activity was normalized to an internal renilla luciferase transfection control and performed in triplicate. (**B**) Smad3E rescue was determined by Western blotting (anti-HA) at different time points.

To further characterize this time course, samples from each time point were analyzed by Western Blotting (**Figure 3-11B**). In contrast to the build-up of luciferase (**Figure 3-4**) over the same timeframe, the amount of cleaved Smad3E decreased. This phenomenon may contribute to the linear kinetics of the transcriptional luciferase reporter (**Figure 3-11A**) and could be the result of two possible scenarios. The efficiency of Smad3E rescue could decrease at the later time points. This scenario is unlikely, however, when taken in context with the luciferase-SURF data (**Figure 3-4A**) and that fresh media containing rapamycin was added every 6 h to avoid possible rapamycin metabolism. Assuming that Smad3E rescue is constant at all time points, this decrease could be explained if Smad3E positively regulates its own destruction, perhaps by the ubiquitin-proteasome pathway [108].

Section 3.5 : SURF in S2 cells

Encouraged by the success of SURF in mammalian cells, we next attempted to use SURF to generate conditional alleles in S2 cells, with the intention of generating SURF controlled conditional alleles of DBT and SGG in *Drosophila*. As mentioned in chapter 2, these proteins are required for developmental viability, so a conditional allele would allow the generation of adult nulls that could illuminate the genes' roles in the maintainance of circadian rhythms.

We first transferred our SURF allele of luciferase into a *Drosophila* expression vector (pAc5). As expected, luciferase activity dramatically upon treatment with rapamycin (**Figure 3-12**). However, we could not definitively state that this was the result of the rescue of protein stability as we could not detect the rescued protein by western blot. However, based on the mammalian results, we concluded that rescue was functioning as intended and moved on to the generation of conditional alleles of DBT and SGG.

A SURF construct of DBT was generated and tested in S2 cells. To our surprise, SURF-DBT was robustly expressed, even in the absence of rapamycin. Additionally, although some rapamycin induced cleavage could be seen, it was inefficient, especially in comparison to the SURF alleles used in mammalian cells (**Figure 3-13**).

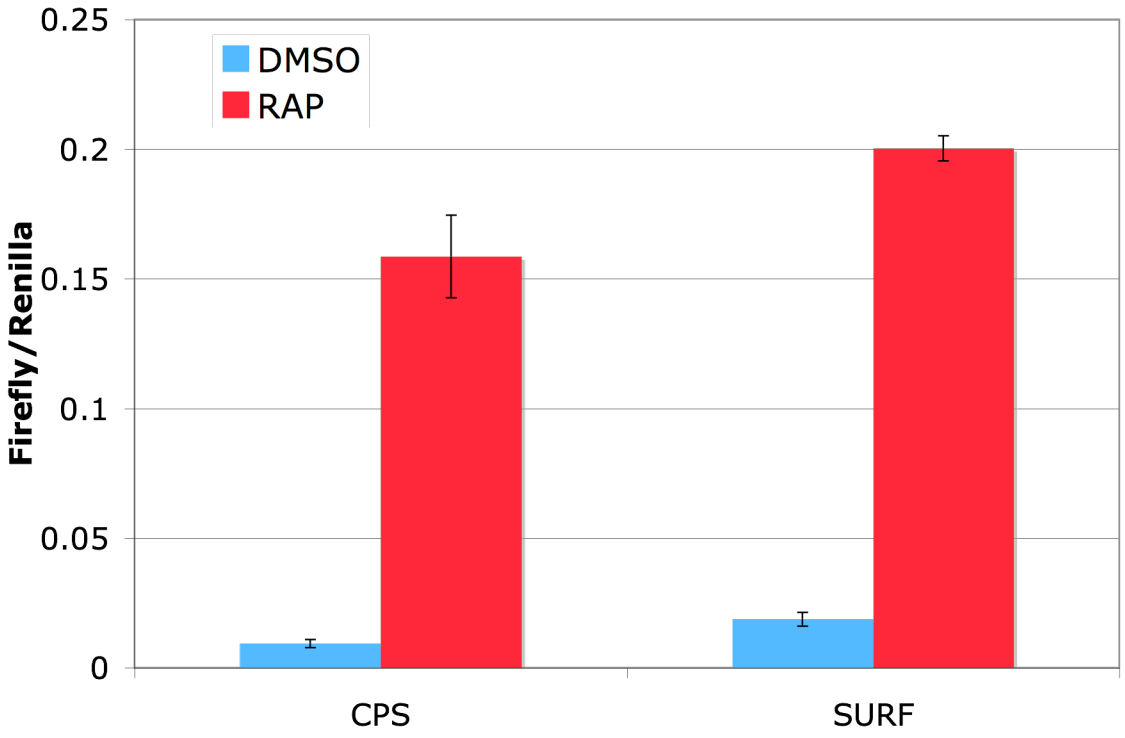


Figure 3-12 : SURF activation of luciferase in S2 cells. Constructs 1 and **4** were transferred to the pAc5 *Drosophila* expression vector. S2 cells were transfected either with N-luc and C-luc (CPS, **Figure 2-3B**) or with Construct **1** and Construct **4** (SURF). Cells were allowed to recover and express protein overnight and then were treated with rapamycin (100 nM) for 24 h. Luciferase activity was measured using renilla luciferase as a transfection control. Error bars represent standard deviation (n=3).

We reasoned that, since the SURF-DBT construct was generated in pUAS and driven with heat shock induced GAL4, the failure to degrade SURF-DBT may be the result of an upregulation of chaperones such as HSP-70. Thus, we instead expressed GAL4 under the control of the actin promoter. Unfortunately, heat shock made no difference in the expression level of SURF-DBT, leading to the conclusion that the cells were unable to recognize SURF-DBT as destabilized. Additionally, the failure to induce C-terminal ubiquitin cleavage suggests that either the ubiquitin hydrolases are unable to effectively recognize ubiquitin or that ubiquitin cannot complement in this context.

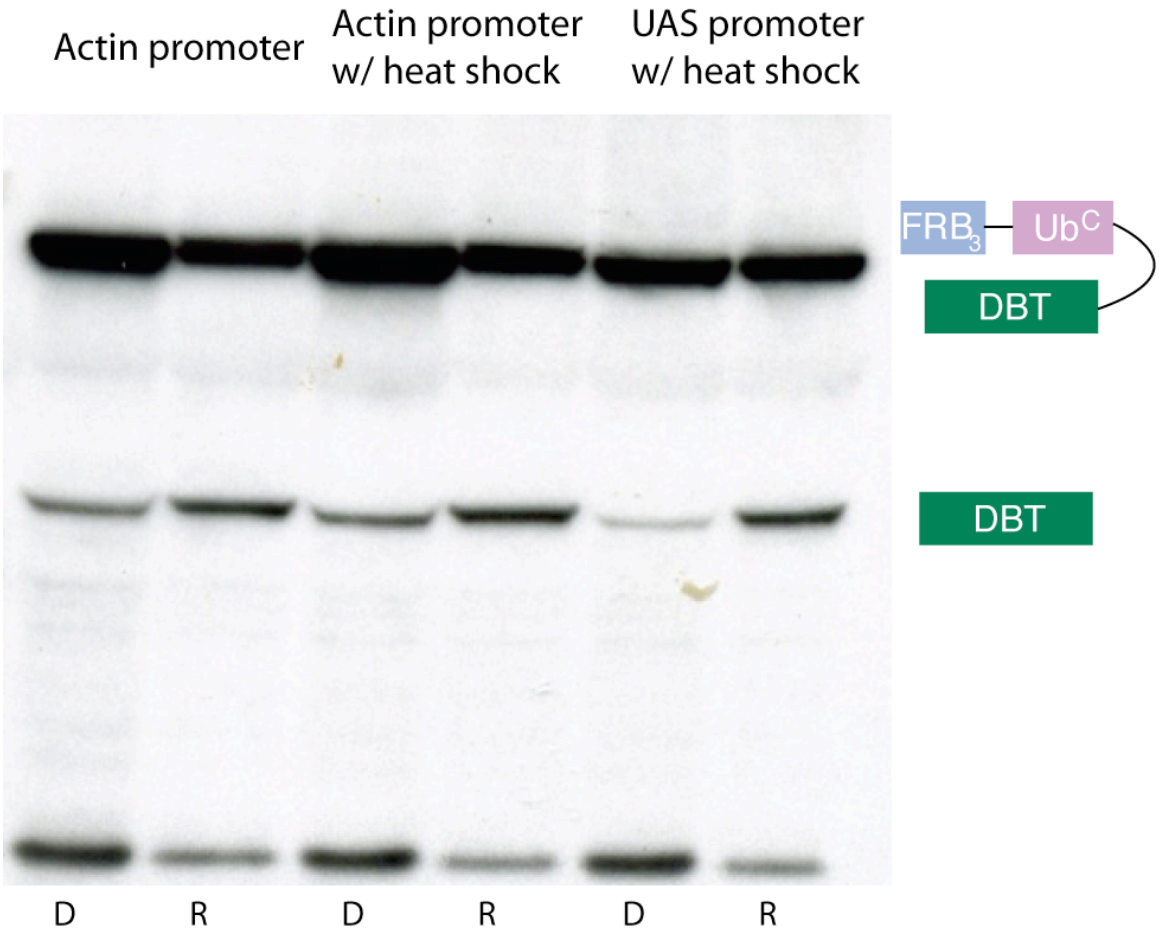


Figure 3-13 : A SURF allele of Doubletime. S2 cells were transfected with a SURF-DBT construct and with Construct **1** either in the pAc5 vector in the the pUAS vector. After overnight expression, indicated samples were incubated in a 37 °C water bath for 15 minutes. All samples were then treated with rapamycin (R, 100 nM) or DMSO (D, 0.1%) for 24 h. SURF-DBT expression and cleavage was assayed by Western blotting (anti-HA).

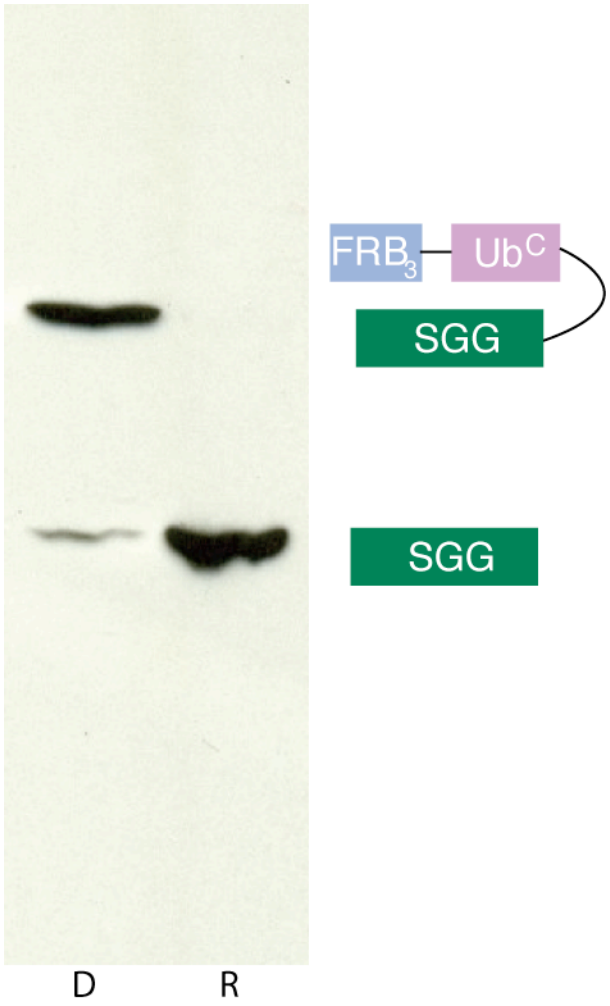


Figure 3-14 : A SURF allele of SGG. S2 cells were transfected with a SURF-SGG construct and with Construct **1** in the pUAS vector along with Gal4 in the pAc5 vector. After overnight expression, samples were treated with rapamycin (R, 100 nM) or DMSO (D, 0.1%) for 16 h. SURF-SGG expression and cleavage was assayed by Western blotting (anti-HA).

As in the case of our CPS-DBT alleles, we reasoned that perhaps the strange domain structure of DBT was responsible for the failure to generate a conditional allele. As such, we turned instead to SGG. A SURF-SGG construct was generated and tested. As in the case of SURF-DBT, the SURF-SGG construct was robustly expressed even in the absence of rapamycin (**Figure 3-14**). However, unlike SURF-DBT, SURF-SGG displayed almost complete rapamycin induced cleavage. Thus, it appears that the failure of a SURF construct to be directed to the proteasome is unrelated to the ability of ubiquitin to complement and be cleaved.

Further examination of SGG reveals that it has much the same domain structure as DBT – a well conserved kinase domain followed by a unique and likely unstructured tail. How this domain structure contributes to the difficulty in applying SURF (if indeed, that is the reason for the difficulty) remains a mystery though. A reasonable possibility may be that the C-terminal domains of SGG and DBT are recruiting other proteins, forming a complex that may be stabilized against degradation. In some cases, this complex may also interfere with protein complementation *in trans* – as in CPS-DBT and SURF-DBT.

Another possibility is simply that SURF is not general with respect to model systems. The problem may not be the proteins but the S2 cells. It is possible that FRB_3 is more stable at 25 °C than at 37 °C. It is also possible that ubiquitin complementation is not as efficient at 25 °C. Although SURF-luc worked from an enzymatic activity standpoint in S2 cells, it was never formally

shown that this was due to stabilization of a destabilized luciferase. However, given the experience with C-lucFRB$_3$ (**Figure 2-10**), it seems likely that FRB$_3$ is able to function as a degron in S2 cells. Additionally, the rapamycin induced cleavage of SURF-SGG was highly efficient, showing that ubiquitin complementation is quite efficient at 25 °C. Thus, it seems likely that SGG and DBT are the exceptions to SURF's generality rather than the S2 cells. To formally test this, however, more proteins that were successful in mammalian cells should be transferred to the S2 system. If they behave similarly in S2 cells as they do in mammalian cells, one can conclude that the limits to SURF's generality are at a protein level rather than at an organism level.

Section 3.6 : Conclusions

Section 3.6.1 : SURF can generate conditional alleles

In summary, we have developed a new posttranslational technology for controlling protein function, SURF (Spilt Ubiquitin for the Rescue of Function). We were able to observe ubiquitin complementation and cleavage of protein fusions when the FRB* and FRB degrons were employed, but not with the strong SopE degron. In contrast to SopE, both FRB* and FRB can be stabilized by interactions with rapamycin bound to FKBP. It is likely that this stabilized intermediate is essential, as it allows time for ubiquitin complementation and

hydrolysis of the protein fusion. Luciferase was used as a model protein to examine the kinetics and tunability of SURF. Like other degradation technologies [32, 109], the kinetics of SURF are limited to the rate of protein synthesis in the cell. SURF is highly responsive to the concentration of the small molecule activator rapamycin, and complementation and cleavage can be abrogated by the addition of another small molecule ascomycin. SURF should therefore allow for highly tunable protein rescue by simple adjustment in rapamycin and ascomycin concentrations. Unlike other degradation technologies, the protein of interest is released from the degron in SURF. Therefore its native structure and stability is restored, which is attractive for the interrogation of protein function.

We used SURF to control the levels and activity of members of three important classes of protein: a protease, a kinase, and a transcription factor. Rescue of caspase-3, a cysteine protease involved in apoptosis and important in many types of cancer, was observed and the cleaved native caspase-3 was apparently processed by upstream caspases during staurosporine-induced apoptosis. The tyrosine kinase activity of v-Src was also efficiently recovered. SURF was also able to rescue Smad3E, effectively restoring all activity upon the addition of rapamycin. In the midst of the complicated TGFβ signaling pathway, we were able to isolate Smad3 driven transcription by using two pharmacological agents, rapamycin and SB43152. Kinetic evaluation of Smad3E regulated transcription allowed us to take a first step towards identifying it as a participant in a TFGβ feed forward loop. Further analysis of this system with mutant TGFβ

receptors may allow us to define the effects of Smad3, as well as, rapamycin treatment on this important signaling pathway.

In the case of the v-Src, Smad3E, and to a lesser extent caspase-3 stabilization mutants, which rescued their proteins through stabilization and not cleavage, limited or altered activity was rescued even though protein levels did increase. This highlights the complimentary utility of SURF to existing degradation technologies that require permanent fusion of degron domains to proteins of interest. SURF represents an attractive alternative when the protein of interest is known to take part in protein-protein interactions or multimeric complexes as rescue is inextricably linked to generation of the native protein. In the future this cleavage dependent rescue technology could be expanded to include other methods of protein regulation such as mislocalization or autoinhibition [84].

Section 3.6.2 : SURF vs. CPS

We turned to SURF as a means of generating conditional alleles since CPS proved to be difficult to apply to novel proteins. In contrast, SURF controlled conditional alleles of novel proteins could be easily generated, at least in mammalian cells. Unfortunately, an open question is whether the failure to generate conditional alleles of SGG and DBT was because of the proteins selected or the cells used. Thus, although SURF appears to be more general

than CPS, there are limits, whether these turn out to be in terms of targetable proteins or in terms of accessible model organisms.

In its current incarnation, SURF is a considerably slower method of generating conditional alleles than CPS. SURF, since it relies on protein synthesis to generate activity requires hours to function rather than minutes, as is the case for CPS. If another N-terminal tag was used to inactivate the protein, such as one that mislocalizes the protein or one that directly inhibits the protein, much faster activation kinetics could potentially be obtained through SURF. So, just as there is room to improve CPS as a means of generating conditional alleles through the use of libraries to screen for splicing junctions, there is room to improve SURF through changing the way in which activity is inhibited.

Section 3.6.3 : The future of CPS and SURF

Conditional alleles generated by CPS are extraordinarily powerful. They result in rapid, tunable and wild-type activity when successful but, as we found, generating and optimizing a conditional allele of a new target protein is neither straightforward nor guaranteed to be successful. What is needed is an improvement in the generality of CPS. This could take a number of forms. The most straightforward would be a systematic variation of splicing sites, perhaps starting from our luciferase allele. From this it may be possible to determine the rules or at least trends governing successful splicing sites.

Alternatively, it may be possible to alter the intein itself to increase splicing promiscuity. Work is already under way in the lab to evolve the naturally split Synechocystis dnaE intein to expand the range of allowed splice junctions. The dnaE intein is not conditional, but as a naturally split intein it is more stable and at least appears to be more promiscuous than the artificially split VMA intein that we used. Thus, using a similar screen to that used by Liu et al. [36] it may be possible to develop a conditional dnaE intein, which would likely be a more general splicer than conditional VMA.

Finally, as mentioned earlier, a means to generate and screen a library of splice sites would streamline the process of applying CPS to new protein targets. However, as it currently stands, application of CPS to new proteins is cripplingly time consuming and difficult. Thus, any further work with CPS should be on improving the generality of the splicing rather than on immediately attempting to apply CPS to a new target.

SURF, in contrast to CPS is a fairly general technique. We have come across some exceptions to this, such as SGG and DBT, but these do appear to be just that – exceptions to an otherwise general technique. Although SURF is not as fast as CPS in generating protein activity, it is still quite rapid. Thus, given its generality, the generation of a new conditional allele should start with SURF. Only if temporal control on the timescale of minutes is required should CPS be attempted.

Chapter 4 : Regulation of the Bacterial Sigma Factor

This work was performed in collaboration with Seth Darst of the Rockefeller University and Alex Shekhtman and David Cowburn of the New York Structural Biology Center. This work is not related to the previous chapters but was nonetheless performed as part of my thesis work.

Section 4.1 : Introduction

The core bacterial RNA Polymerase (RNAP, subunit composition $\alpha_2\beta\beta'\omega$) is catalytically competent, but is unable to initiate promoter specific transcription [110-113]. Promoter specific transcription initiation requires the recruitment of a σ factor to form the RNAP holoenzyme [114, 115] (**Figure 4-1**). The σ factor confers upon the RNAP holoenzyme the ability to recognize and bind promoter DNA and also aids in dsDNA melting to allow transcription initiation [38, 112, 113, 116].

Bacterial species have anywhere from a single σ factor to over 60 which, together with various transcription factors regulate different transcriptional responses and programs [38, 112, 113]. σ factors fall into two broad classes, the σ-70 family and the σ-54 family [38]. The σ-70 family is more widespread, with multiple representatives in most bacterial species whereas some species do not even have a single member of the σ-54 family. The σ-70 family is further subdivided into 4 groups. The group 1 σ factors are responsible for the bulk of

transcription during log phase growth while group 2-4 σ factors or alternative σ factors fulfill more specialized functions such as stress responses [38].

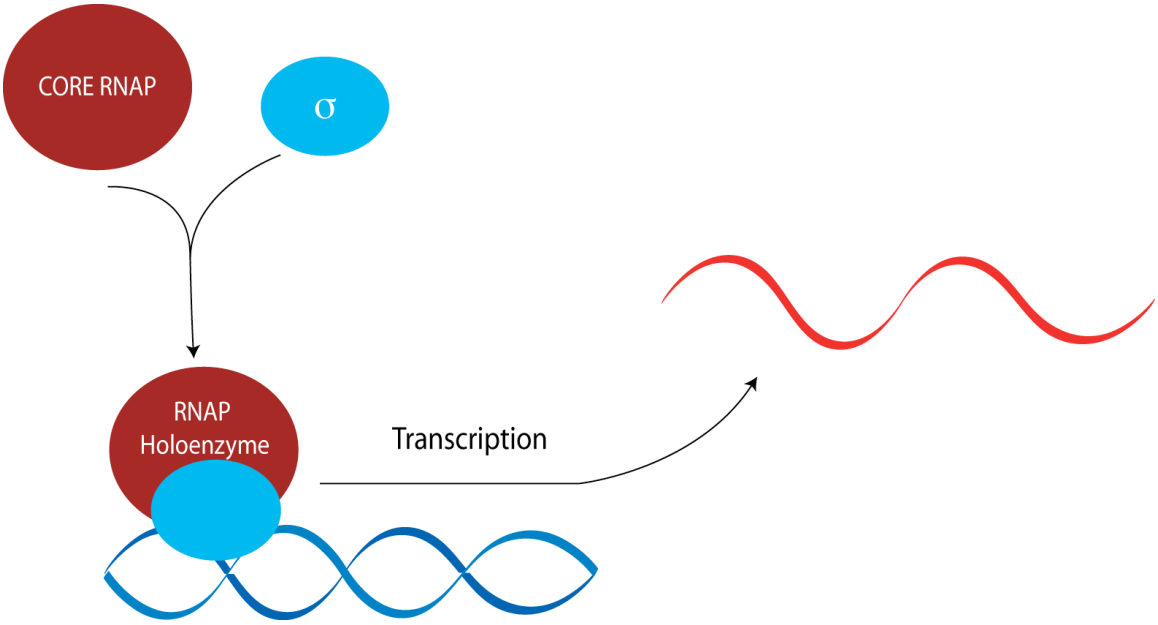


Figure 4-1 : Prokaryotic transcription initiation. In order to initiate promoter specific transcription initiation, the bacterial core RNAP ($\alpha\alpha\beta\beta^{1}\omega$) must form a holoenzyme complex with a σ factor. The σ factor confers the ability to recognize and bind promoter DNA upon the RNAP holoenzyme.

As DNA binding proteins, σ factors need to be regulated in order to prevent DNA binding at inappropriate times. Unlike eukaryotes, prokaryotes cannot regulate their DNA binding proteins simply by excluding them from the nucleus, so alternative methods are needed. The σ factors are regulated in two main ways (**Figure 4-2**). Specialized or alternative σ factors are regulated through binding to anti-σ's that prevent them from contacting DNA [38, 117, 118].

By contrast, the group 1 σ factors are self-regulatory in that they do not bind DNA prior to the formation of the RNAP holoenzyme [39, 40, 119].

Group 1 σ factors are made of 4 conserved domains, known as σ-1.1, σ-2, σ-3 and σ-4 connected by flexible linkers. A wealth of structural and biochemical information is available on σ-2, σ-3 and σ-4 [41-43, 117, 118, 120-123] but the structure of σ-1.1 has never been reported. σ-2 is involved in binding the –10 element of the promoter as well as aiding in the melting of the double stranded DNA molecule [124]. σ-3 is also involved in recognizing the –10 element of the promoter, particularly in the extended –10 promoter [125]. σ–4 is responsible for recognizing and binding the –35 region of the promoter [126, 127]. All three domains are also responsible for forming contacts with the core enzyme. The alternative σ factors contain σ-2, σ-4 and sometimes σ-3, but σ-1.1 is unique to the group 1 σ factors. Although there is no structural information available on σ-1.1, biochemical information shows that it plays a role in the transition from a closed to an open holoenzyme complex [128] and is responsible for preventing DNA binding prior to formation of the holoenzyme [39, 40, 119].

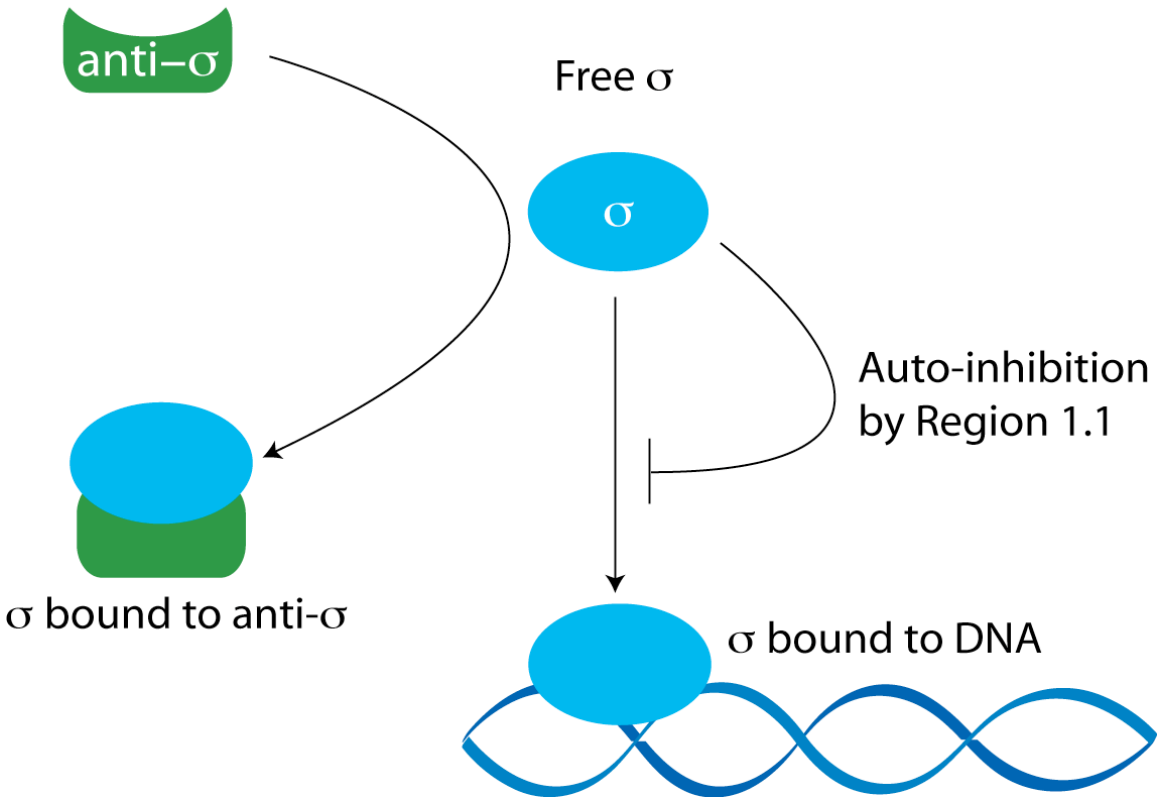


Figure 4-2 : Regulation of σ factors. σ factors are prevented from binding DNA at inappropriate times in two ways. The alternative σ factors are bound by anti-σ factors to preclude DNA binding. Group 1 σ factors, on the other hand, are regulated by an N-terminal domain known as region 1.1 that is responsible for preventing DNA binding prior to the formation of an RNAP holoenzyme complex.

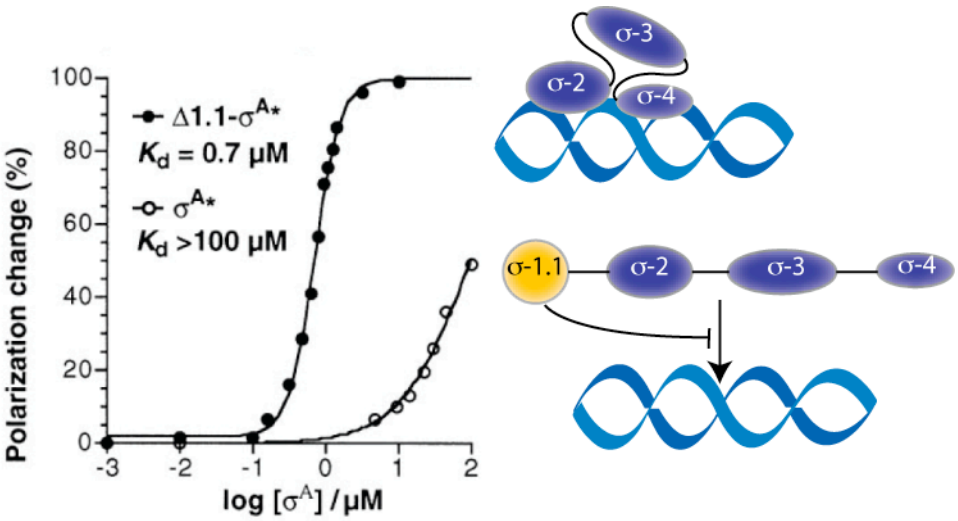


Figure 4-3 : Region 1.1 prevents group 1 σ factors from binding promoter DNA. 5'-fluorescein labeled -35 promoter element DNA was incubated with σ-A (open circles) and with σ-A lacking region 1.1 (closed circles). Protein-DNA binding was measured as a change in fluorescence polarization. Modified from reference [119].

The mechanism by which region 1.1 prevents DNA binding by the group 1 σ factors is still unknown. σ constructs lacking region 1.1 are able to bind promoter DNA while full length group 1 σ factors cannot [39, 40, 119] (**Figure 4-3**). Additionally, region 1.1 is able to interfere with DNA binding by region 4 in *trans* [40] (**Figure 4-4**). This led to the proposal that region 1.1 binds to region 4 to prevent DNA binding.

Camarero *et al.* investigated the possibility of a direct interaction between region 1.1 and region 4 through segmental isotopic labeling [119]. The last 50 residues of region 4, known as region 4.2, form the binding surface for the -35 promoter element. Region 4.2 was segmentally labeled with ^{15}N and ^{13}C in

constructs containing and lacking region 1.1. In this way, direct interactions between region 1.1 and region 4.2 would manifest themselves as differences in the NMR spectra of region 4.2 in the presence and absence of region 1.1. However, the spectra of the two constructs were largely similar (**Figure 4-5**), leading to the conclusion that there is no strong interaction between region 1.1 and region 4.2. As such, the authors proposed that region 1.1 was either binding N-terminal to region 4.2, inhibiting DNA binding through an allosteric mechanism or region 1.1 was not interacting with other domains of the σ factor and DNA binding was being prevented simply by electrostatic repulsion between the negatively charged region 1.1 and negatively charged DNA.

We wished to determine the structure of σ-1.1 as well as the mechanism of autoinhibition of DNA binding. A solution structure of σ-1.1 was determined by multidimensional NMR spectroscopy. DNA binding autoinhibition was first investigated using a similar segmental labeling approach to the one taken by Camarero *et al* [119]. However, the spectra generated from the segmental labeled protein were not of sufficient quality to make residue assignments. The only conclusion that could be drawn is that region 1.1 was likely making some contacts with other portions of the protein. Instead, an intramolecular crosslinking approach was taken which revealed that multiple interdomain contacts are made between a negative face of region 1.1 and the DNA binding domains of the sigma factor.

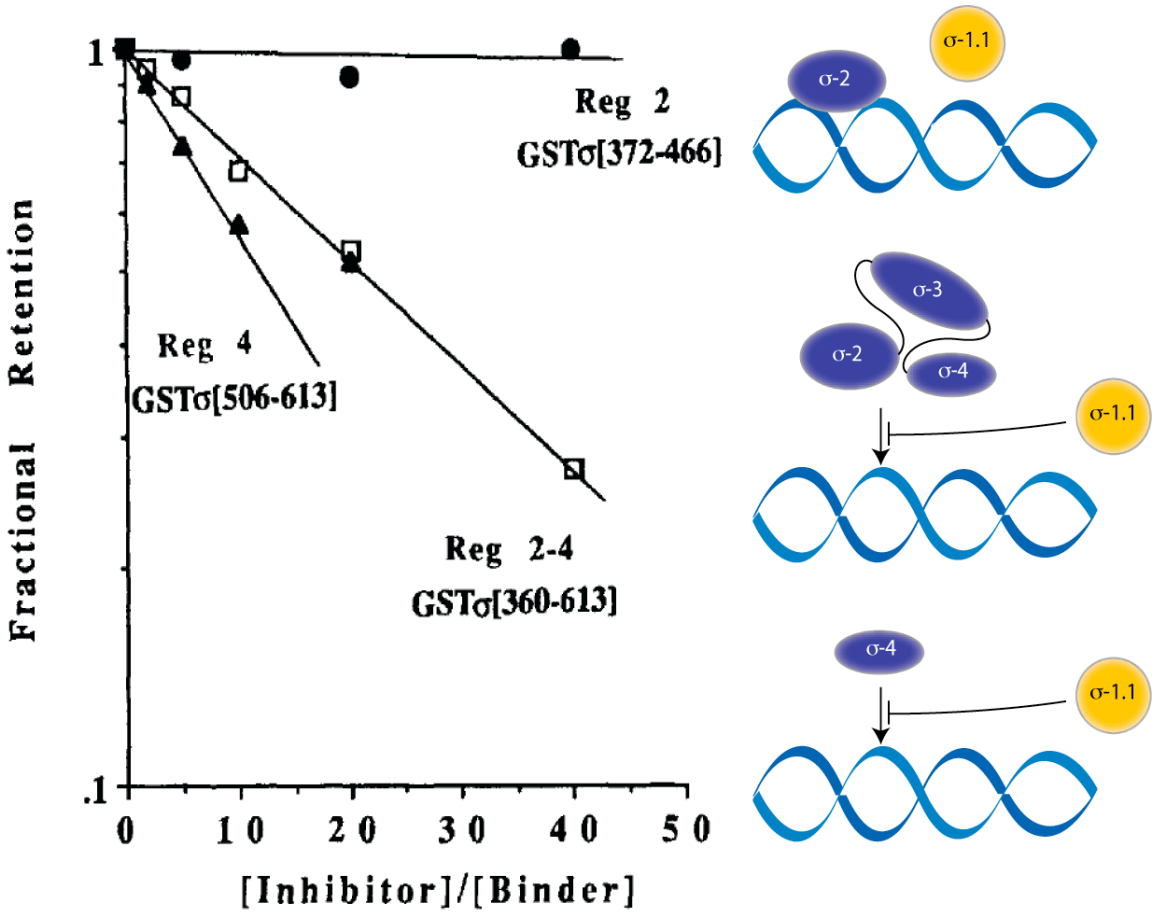


Figure 4-4 : DNA binding inhibition by region 1.1 in *trans*. End-labeled ptac promoter DNA was incubated with σ-70 constructs containing the indicated residues ("Binder") as well as σ-70 region 1.1 ("Inhibitor"). Protein-DNA complexes were captured on nitrocellulose filters. DNA-binding inhibition by region 1.1 was measured as a decrease in the retention of end-labeled DNA on the filters. Modified from reference [40].

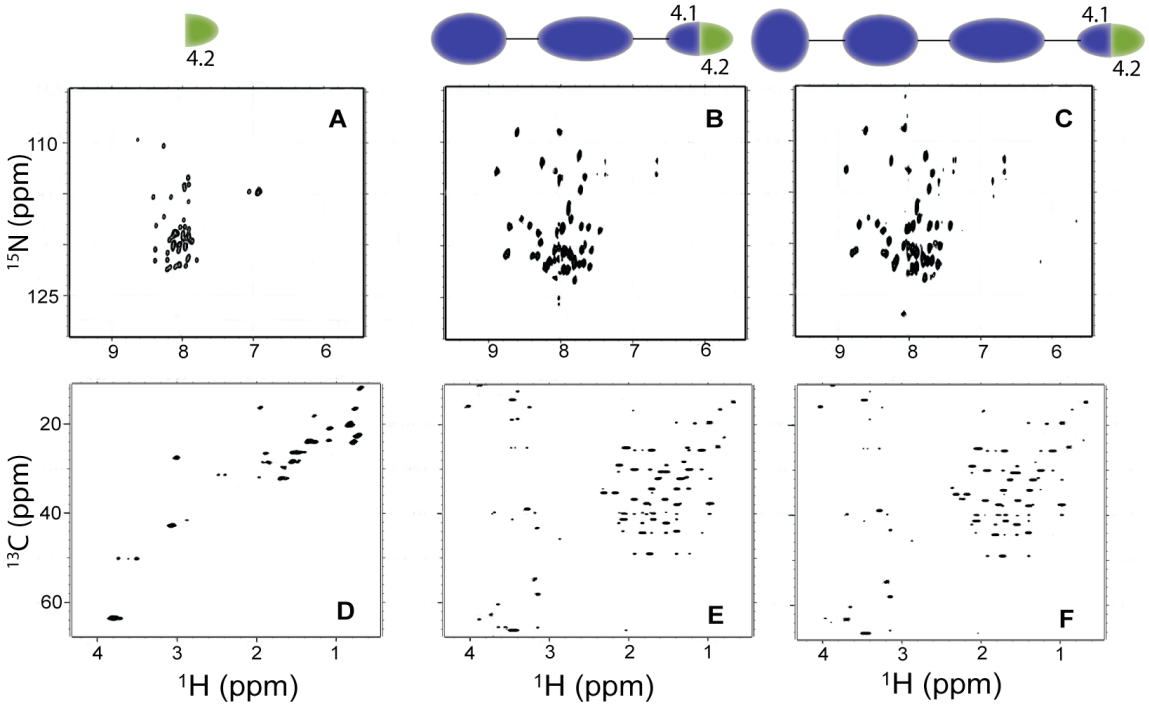


Figure 4-5 : Segmental labeling of region 4.2. (**A**-**C**) ^{1}H[^{15}N] HSQC-TROSY spectra of region 4.2 in isolation (**A**), in a σ-A construct lacking region 1.1 (**B**) and in full length σ-A (**C**). (**D**-**F**) ^{1}H[^{13}C] HSQC-TROSY spectra of of region 4.2 in isolation (**D**), in a σ-A construct lacking region 1.1 (**E**) and in full length σ-A (**F**). The presence (**C, F**) or absence (**B, E**) of region 1.1 does not appear to affect the spectra of region 4.2 significantly. Modified from reference [119].

Section 4.2 : Results

Section 4.2.1 : Determination of the structure of Region 1.1

We chose to work on the *T. maritima* σ-A protein rather than the *E. coli* σ-70 protein. *T. maritima* σ-A region 1.1 is well conserved in comparison to *E. coli* σ-70 (**Figure 4-6A**). In fact, *T. maritima* σ-A is homologous enough to *E. coli* σ-70 to be capable of forming a functional holoenzyme complex with *E. Coli* RNAP (**Figure 4-6B**). Thus, since *T. maritima* σ-A is functionally similar if not identical to *E. coli* σ-70 but is both considerably smaller (46.5 kDa vs. 70 kDa) and is thermostable, we reasoned it would be a good choice for our structural studies.

A list of constructs used in the structure determination and segmental labeling work is shown in **Table 4-1**. Preliminary work revealed that the first 25-30 residues of *T. maritima* σ-A are unstructured, as are residues ~95-116. These unstructured regions made determination of a structure difficult. Unfortunately, although the first 25-30 residues are unstructured, they are required for expression. As such, the construct we used for structure determination encoded residues 1-95 with mutations to install a thrombin protease site to remove the unstructured N-terminal region (**Figure 4-7**). This construct expressed well and could be efficiently processed to generate the desired construct (**Figure 4-8**). Spectra from this construct (**Figure 4-9**) were of sufficient quality for structure determination.

A

```
Ecoli        1-MEQNPQSQLKLLVTRGKEQGYLTYAEVNDHLPEDIV--DSDQIEDIIQMINDMGIQVMEE-58
Tmaritima   28--PPQIERRIKKLISLGKKKGYITYEDIDKAFPPDFEGFDTNLIERIHEELEKHGINIVEN-86
                 : : ::* *:: **::**:** :::. :* *:   *:: ** * : ::. **:::*:

Ecoli       59-APDADDLMLAENTADEDAAEAAAQVLSSVESEIGRTTDPVRMYMREMGTVELLTREGEID-118
Tmaritima   87-EPEEEEISASS---DEQELEELLEKESPEIHDSSNVRDSIKMYLKEIGKIPLLTP------138
              *: :::  :.   **:  *   :  *.   : ... *.::**::*:*.: ***
```

B

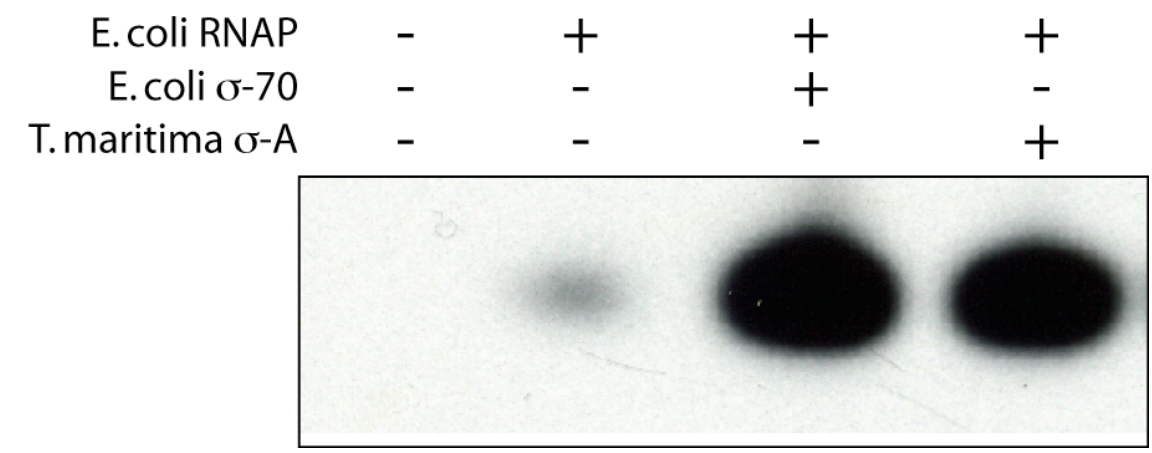


Figure 4-6 : *T. maritima* σ-A is similar to *E. coli* σ-70. (**A**) An alignment of *E. coli* σ-70 and *T. maritima* σ-A region 1.1. Identical residues are indicated by '*', similar residues are indicated by ':' and dissimilar residues are indicated by '.'. (25% identity, 58% similarity) (**B**) Abortive transcription initiation reactions on a T7A1 promoter in the presence of *E. coli* RNAP core and the indicated σ factor. Reaction products (CpA^{32}pU trinucleotide) were resolved by denaturing SDS-PAGE and analyzed by autoradiography.

Table 4 – 1 : Constructs used for NMR analysis of σ-A

Name	Sequence*	Notes
σ-A(29-95)	gshm-(29-95)-g	N-terminal unstructured region removed by thrombin cleavage
σ-A(1-95)	1-95-g	
σ-A(29-115)	gshm-(29-115)-g	N-terminal unstructured region removed by thrombin cleavage
σ-A(117-399)	c-(118-399)	Regions 2-4
σ-A(117-318)	c-(118-318)	Regions 2-3
σ-A(EPL)	gshm-(29-399) S116G, S117C	Generated by ligation of σ(29-115) and σ(117-399).
σ-A(Δ4)	gshm-(29-318) S116G, S117C	Generated by ligation of σ(29-115) and σ(117-318).
σ-A(97-399)	c-(98-399)	
σ-A(EPL2)	1-399 S96G,S97C	Generated by ligation of σ(1-95) and σ(97-318).

*Numbers refer to *T. maritima* σ-A residue numbers

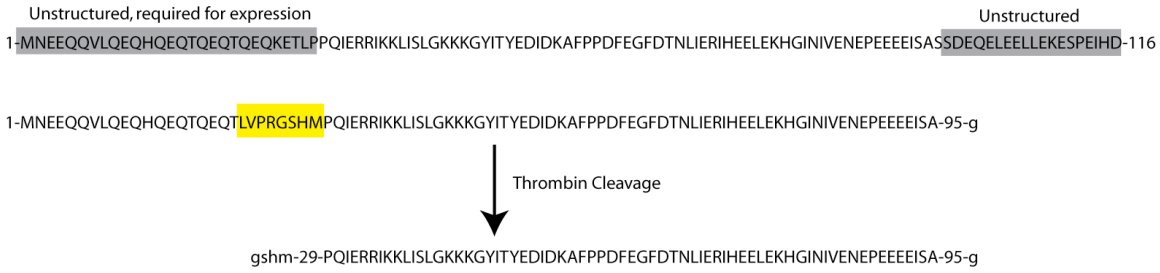


Figure 4-7 : Design of the σ-A(29-95) construct used for structure determination. Preliminary results revealed that the N- and C-termini of σ-1.1 were unstructured (top, highlighted in gray) but that the N-terminus was required for expression. A new construct was generated encoding residues 1-95 (middle). A thrombin protease recognition site (middle, highlighted in yellow) was incorporated by PCR mutagenesis for the removal of the unstructured N-terminus. The final, thrombin processed construct used for structure determination contained residues 29-95 (bottom) with short N- and C-terminal tails that were artifacts of cloning (bottom, lower-case).

gshm-29-PQIERRIKKLISLGKKKGYITYEDIDKAFPPDFEGFDTNLIERIHEELEKHGINIVENEPEEEEISA-95-g

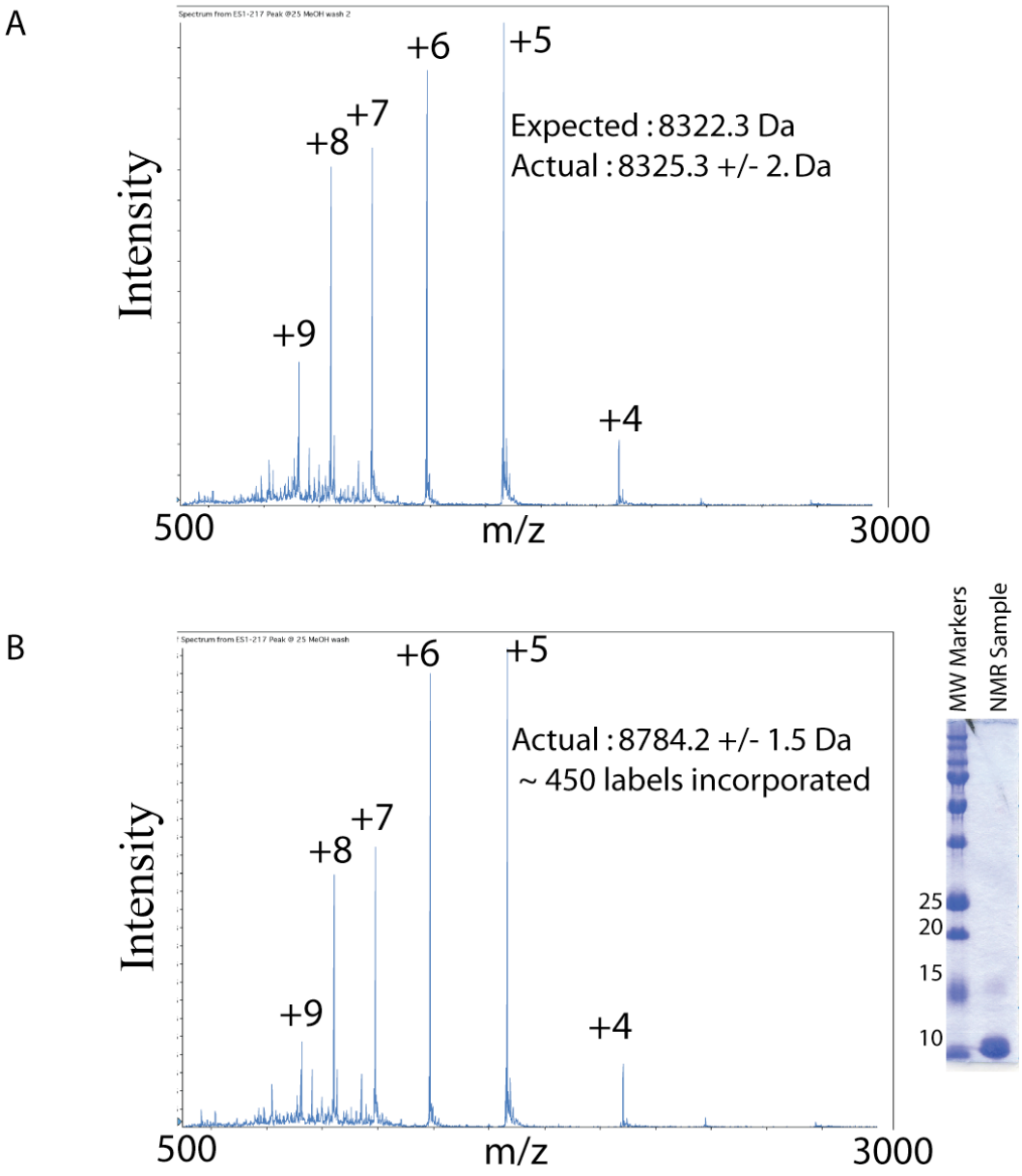


Figure 4-8 : Expression of σ-A(29-95). (**A**) ESI-MS of unlabeled σ-A(29-95) demonstrates the correct construct is expressed. (**B**) (left) ESI-MS of ^{15}N, ^{13}C labeled σ-A(29-95) shows that approximately 450 spin labels are incorporated. (right) SDS-PAGE of the NMR sample used to generate the spectrum shown in **Figure 4-9**. The amino acid sequence of the construct is shown at the top of the figure.

gshm-29-PQIERRIKKLISLGKKKGYITYEDIDKAFPPDFEGFDTNLIERIHEELEKHGINIVENEPEEEEISA-95-g

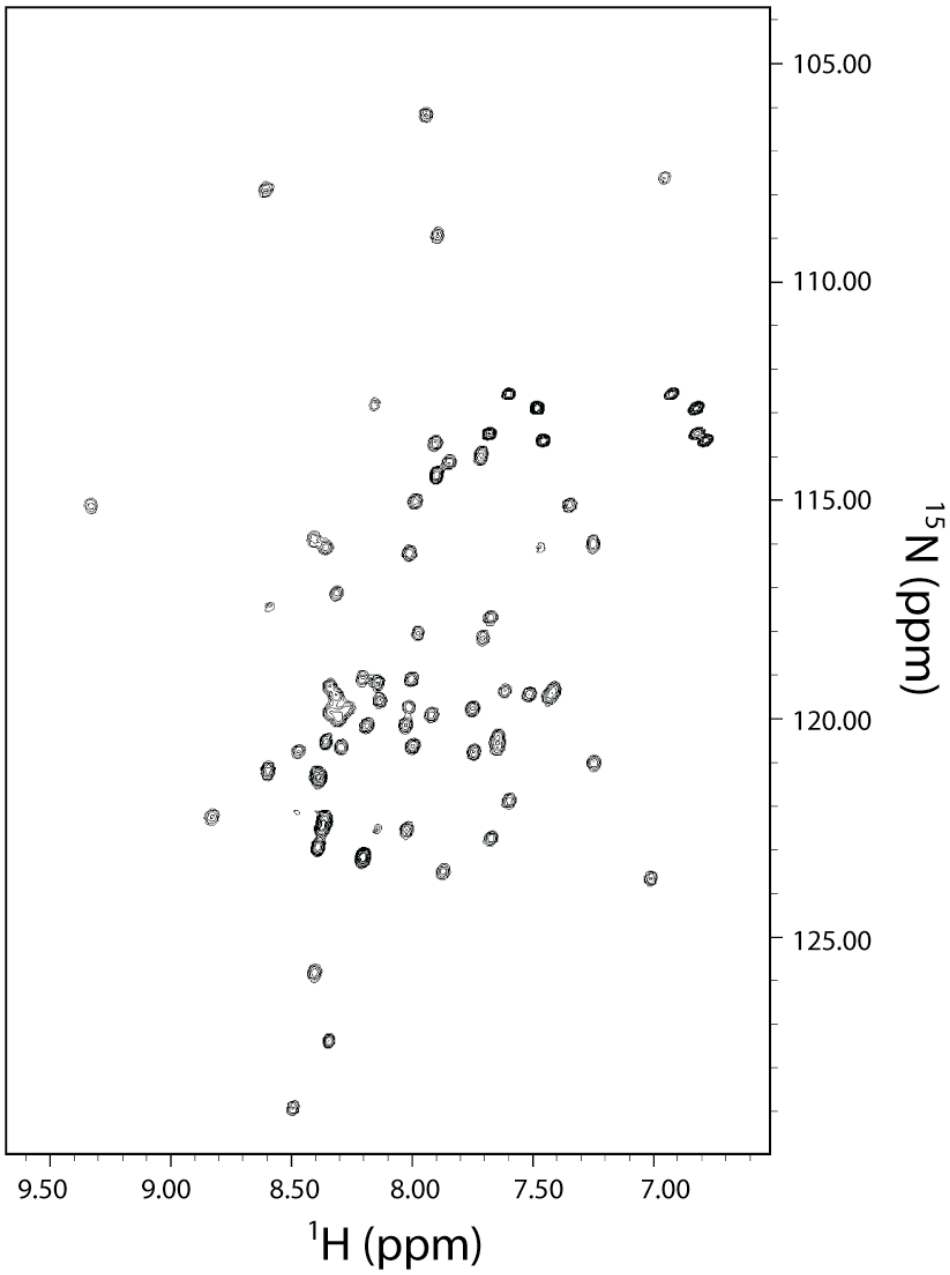


Figure 4-9 : ^{1}H[^{15}N] HSQC spectrum of σ-A(29-95). ^{1}H[^{15}N] spectrum of ^{15}N, ^{13}C labeled σ-A(29-95) at 298 K (protein concentration 370 μM in 10 mM NaPi pH 6.5, 100 mM NaCl, 1mM EDTA). The amino acid sequence of the construct is shown at the top of the figure.

Uniformly ^{15}N and ^{13}C labeled σ-A(29-95) was used to obtain the main- and side-chain assignments of the ^{1}H, ^{15}N, and ^{13}C resonances using the standard set of triple-resonance 3D experiments [129]. An initial assessment of the σ-1.1 secondary structure was made using the deviations of assigned chemical shifts from random coil [130]. This indicated that the domain contains three α helices. The solution structure of region 1.1 was solved using restraints generated from standard multi-dimensional NMR experiments by Alex Shekhtman, Kaushik Dutta and David Cowburn (NYSBC). Briefly, structure calculations were performed using 27 hydrogen bond restraints inferred from slow H/D exchange, 105 dihedral (φ and ψ) angle restraints and 919 distance constraints from aromatic, aliphatic and amide NOEs. The 20 lowest energy structures (**Figure 4-10**) have an average pairwise root-mean-square deviation (rmsd) of 0.45 ± 0.11 Å for backbone residues and 1.12 ± 0.08 Å over all atoms.

Figure 4-10 : The 20 lowest energy structures of region 1.1. Shown is a superimposition of the backbone of the 20 lowest energy structures of region 1.1 that were obtained from 3D NMR based structural restraints. Each structure is individually colored. The N-termini are in the upper right and the C-termini are in the bottom half of the figure. Images generated using MacPYMOL (Delano Scientific).

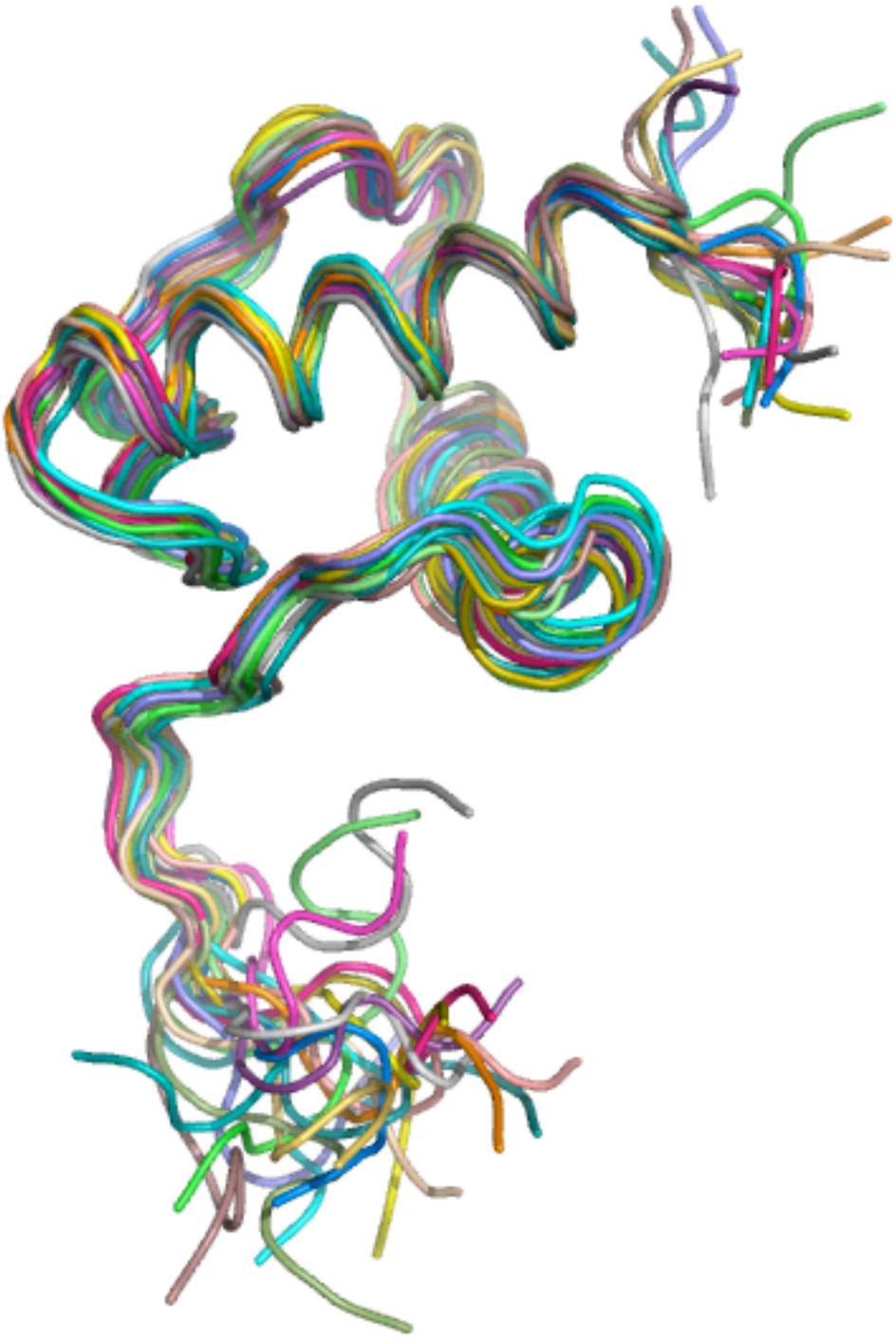


Figure 4-10 : The 20 lowest energy structures of σ-1.1.

Region 1.1 contains three helices (**Figure 4-11**) and has limited homology to any structure in the protein data bank (DALI search, [131]). Helix 1 (H1) is the longest helix and contains residues Q30-K45. H1 is connected by a four residue linker to the smallest helix (H2) comprising residues Y50-A56. H2 is connected by a 9 residue loop to the final helix (H3) which stretches from N67-K78. H2 and H3 are roughly anti-parallel to one another and pack perpendicularly against H1. A hydrophobic core is formed by residues from all three helices and part of the C-terminal tail (**Figure 4-12**). H1 contributes I35, L38, and I39. H2 contributes Y50, I53 and A56. H3 contributes L68 and I72. I69 and L76 appear to be partially buried as well. Finally, the side-chains of I81 and I83, from the C-terminal tail of σ-1.1 project directly into the hydrophobic core.

The relatively small structured portion of the region, less than 70 residues, may explain the difficulty in obtaining a crystal structure of region 1.1. The N- and C-terminal unstructured regions likely interfere with crystal packing. Additionally, the C-terminal unstructured region probably forms a flexible linker to σ-2. In the crystal structure that included region 1.1 but could not resolve it [43], region 1.1 was likely mobile in the crystals, resulting in a diffuse electron density that appeared unstructured.

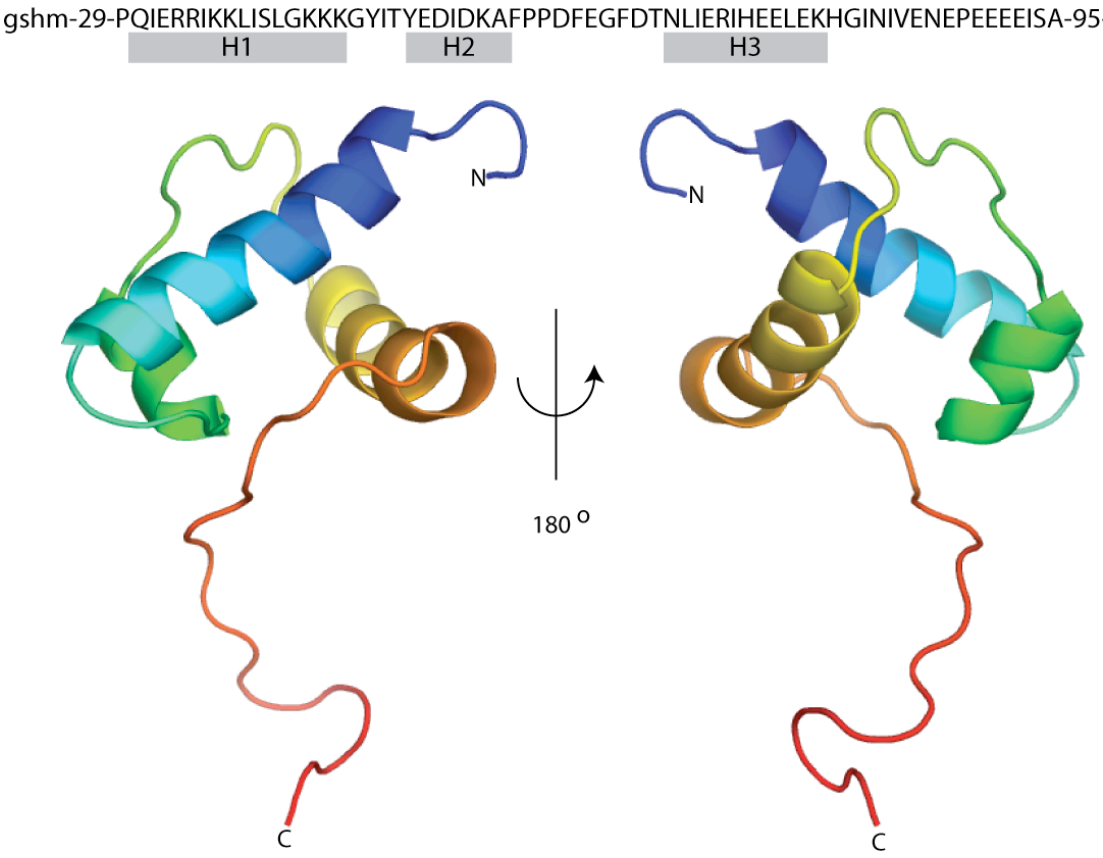


Figure 4-11 : Solution structure of region 1.1. Shown here is a cartoon representation of the secondary structure elements of σ-1.1. The structure is rainbow spectrum colored from N-terminus (blue) to C-terminus (red). Images were generated using MacPYMOL (Delano Scientific). The amino acid sequence of the construct is shown at the top of the figure with secondary structure elements (Helix 1 – Helix 3, H1-H3) indicated.

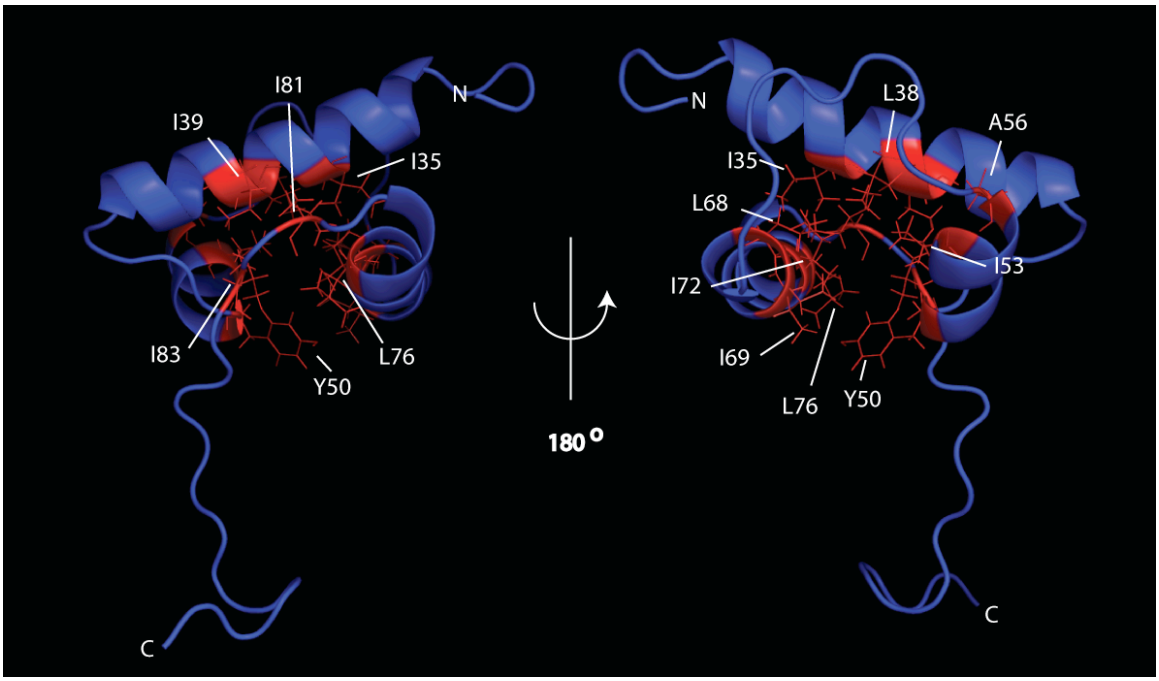


Figure 4-12 : The hydrophobic core of region 1.1. Shown is a cartoon representation of region 1.1. Residues that form the hydrophobic core are shown in red with line representations of the side-chains. Images generated using MacPYMOL (Delano Scientific).

It has been shown that I53 of σ-70 plays an important role in the function of region 1.1 as an I53A mutation is defective in promoter DNA binding by the RNAP holoenzyme [132]. The authors of the study propose that the defect is due to difficulty in displacing σ-1.1 from the DNA binding region of the RNA polymerase holoenzyme. Examining the structure of region 1.1, I53 (I81 in *T. maritima*) is located immediately after H3, with its sidechain projecting into and forming a portion of the hydrophobic core (**Figure 4-13**). Thus, it is likely that the I53A mutation weakens a hydrophobic interaction between the C-terminal tail of

region 1.1 and the helical core. This is consistent with the fact that I53V and I53F mutations in σ-70 behave roughly as wild-type [132]. The disruption in the fold of region 1.1 caused by the I53A mutation likely interferes with the conformational changes required in the σ factor to allow DNA binding by the RNAP holoenzyme.

The electrostatic surface of region 1.1 is mostly negative, as expected from the amino acid composition (**Figure 4-14**). However, strikingly, H1 of the structure is neutral if not slightly positive, resulting in a slightly positive patch on an otherwise strongly negative surface. The possible functional importance of this will be discussed in later sections.

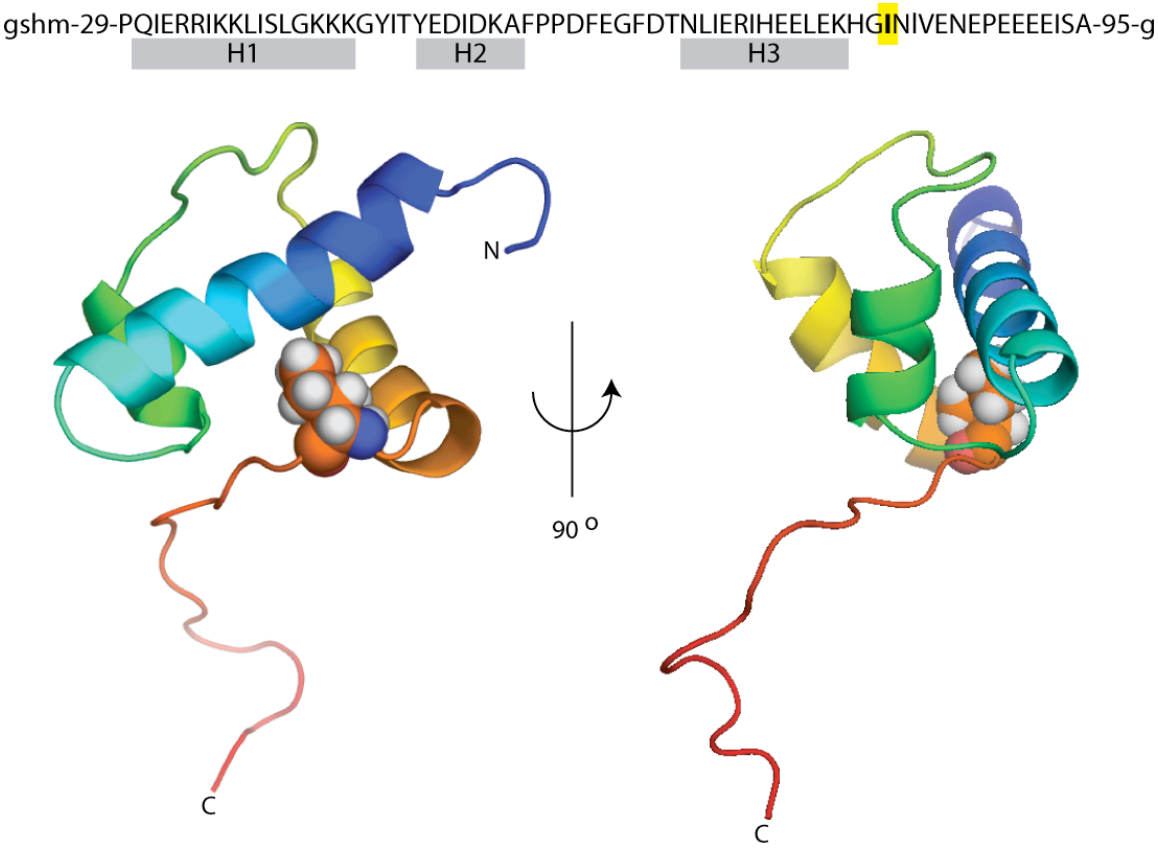


Figure 4-13 : The I53A mutant explained structurally. I53 (I81 in *T. maritima*) is required for efficient binding of RNAP holoenzyme to promoter DNA. Mutation to alanine disrupts RNAP : DNA binding. The side chain of I81 (spheres) projects into the hydrophobic core of σ-1.1, suggesting that the I53A deficiency is due to disruption of the σ-1.1 fold. Images were generated using MacPYMOL (Delano Scientific). The amino acid sequence of the construct is shown at the top of the figure with secondary structure elements (Helix 1 – Helix 3, H1-H3) indicated. I81 is highlighted in yellow.

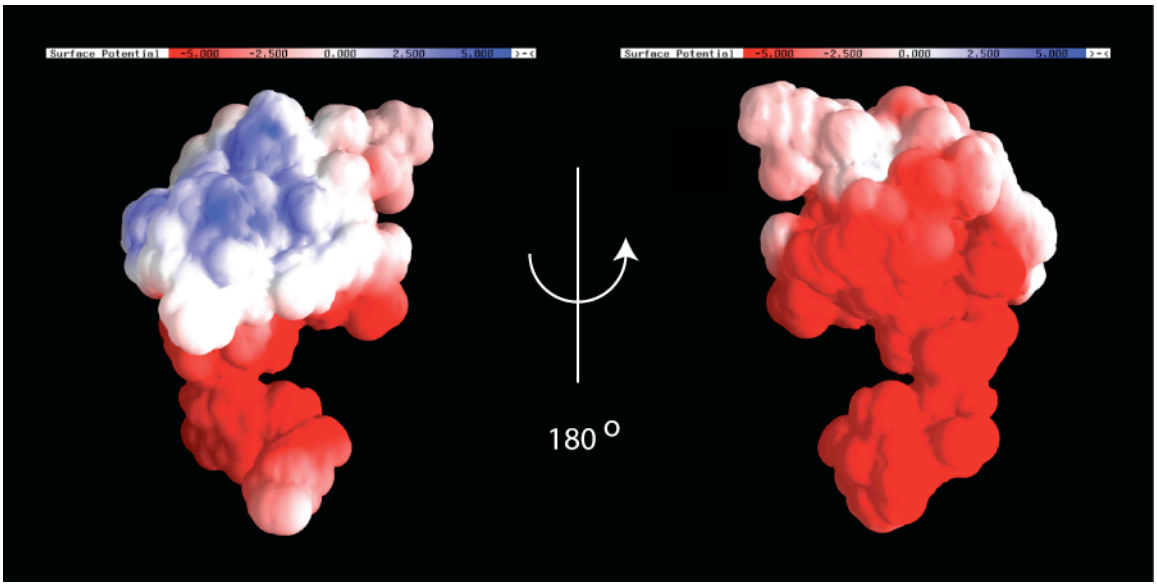


Figure 4-14 : Surface charge of region 1.1. Most of surface of region 1.1 is negatively charged (red), but the surface of helix 1 is neutral (white) with some positively charged (blue) patches. Structure is shown in a similar orientation to that in **Figure 4-11**. Images generated using the GRASP program [133].

Section 4.2.2 : Segmental labeling of *T. maritima* σ-A

In order to determine the mechanism by which region 1.1 inhibits DNA binding, we decided to use Expressed Protein Ligation (EPL) to make full length and truncated σ factors in which only region 1.1 was labeled with ^{15}N and ^{13}C. By comparing the NMR spectra from these samples to that of free region 1.1, we would be able to determine what, if any, part of σ interacts with region 1.1 as well as which residues of region 1.1are involved.

Some mechanisms by which region 1.1 could interfere with DNA binding are shown in **Figure 4-15**. Since region 1.1 is very negatively charged (calculated pI = 4.52) it is possible that region 1.1 prevents DNA binding by simple electrostatic repulsion (**Figure 4-15A**). This would require no interactions between region 1.1 and the rest of the protein. As such, the spectra of region 1.1 in isolation would be expected to be nearly identical to region 1.1 in both full length σ-A (σ-A(EPL)) and in σ-A lacking region 4 σ-A(Δ4)) (**Figure 4-16**). Such a model would be consistent with NMR data from an earlier study in which no strong interaction between region 1.1 and region 4.2 could be detected [119]. If region 1.1 acts by binding to region 4 (**Figure 4-15B**), then the spectra of region 1.1 should be very similar in isolation and in σ-A(Δ4). However, the spectra should be different in σ-A(EPL), reflecting the perturbations caused by contact between region 1.1 and region 4. An interaction between region 1.1 and region 4 is predicted by region 4 DNA binding inhibition in *trans* by region 1.1 [40]. If region 1.1 is binding N-terminal to region 4 (**Figure 4-15C**), then the spectra of σ-A(EPL) and σ-A(Δ4) would be similar to one another, but different from region 1.1 on its own. In this case, further truncations or deletions could be made to determine the exact region of binding. The three models are not mutually exclusive; a combination of mechanisms could be involved.

Figure 4-15 : Possible mechanisms by which region 1.1 could autoinhibit DNA binding. (**A**) The negative charge of region 1.1 could repel negatively charged DNA. (**B**) Region 1.1 could bind to region 4 and directly prevent DNA binding. (**C**) Region 1.1 could bind N-terminal to region 4 and prevent DNA binding by an allosteric mechanism.

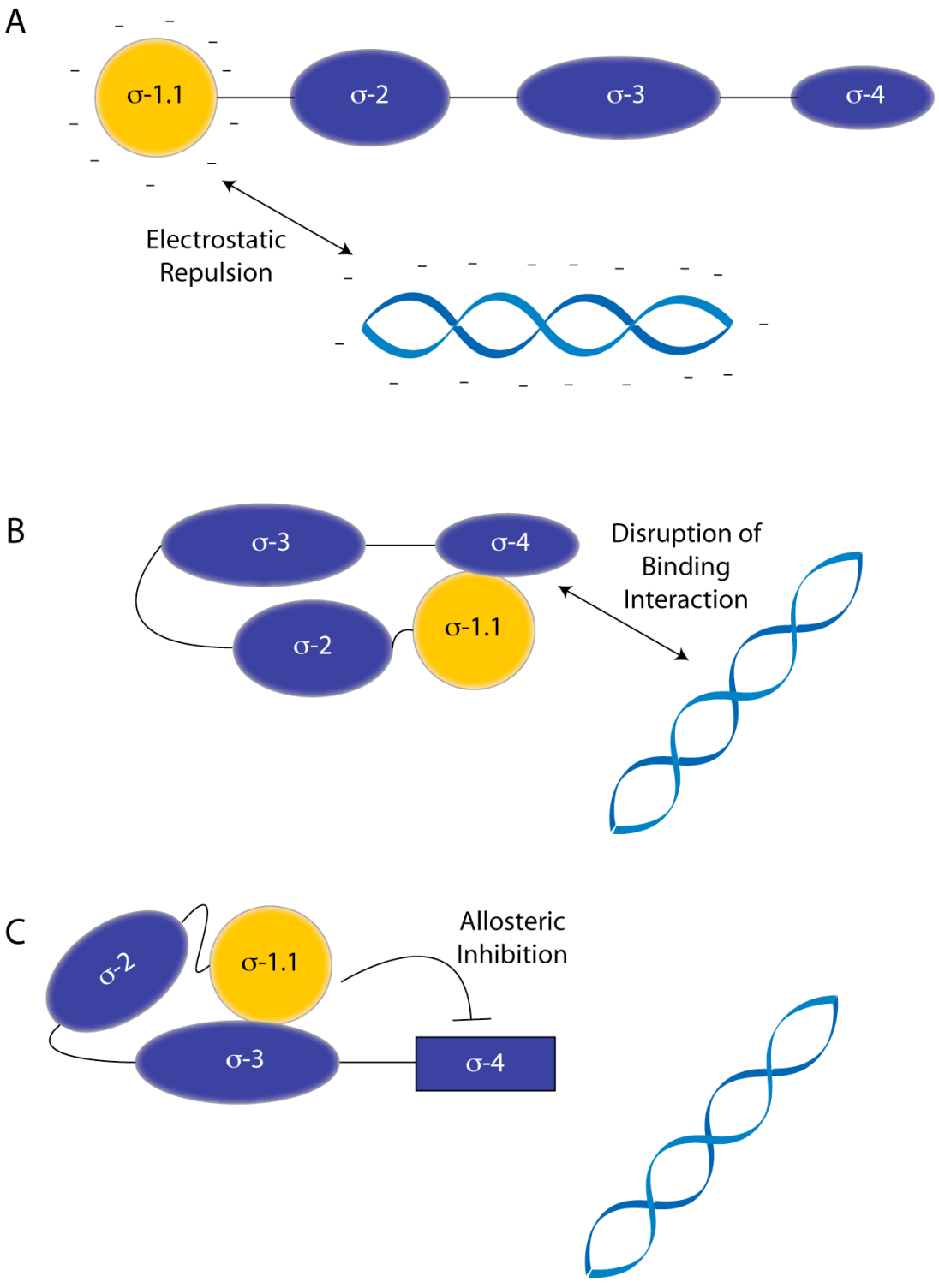


Figure 4-15 : Possible mechanisms by which region 1.1 could autoinhibit DNA binding.

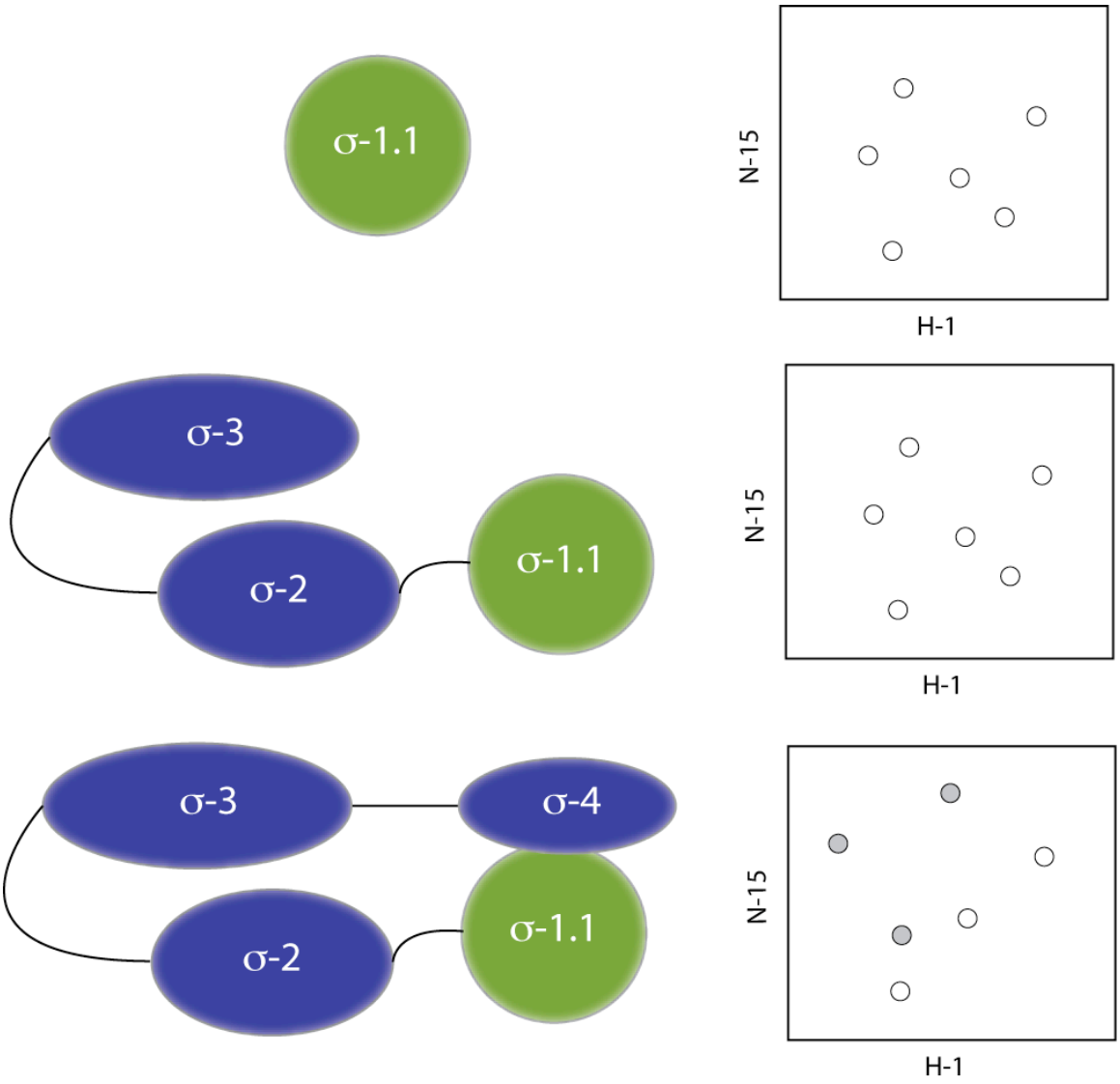


Figure 4-16 : Predicted NMR spectra of isotopic labeled region 1.1 if region 1.1 acts by direct binding to region 4. If σ-1.1 is directly binding to σ-4 (**Figure 4-14B**), then the spectrum of σ-1.1 in isolation (top) should be similar to the spectrum of σ-1.1 in σ-A lacking σ-4 (middle). However, interaction between σ-1.1 and σ-4 should be apparent in the spectrum of full length σ-A (bottom) as perturbations in the residues involved (filled circles).

Segmental isotopic labeling of σ-A was achieved through EPL between region 1.1 constructs containing C-terminal thioesters and σ-2-σ-4 constructs containing N-terminal cysteines (**Figure 4-17**). Isotopically labelled σ-1.1 was obtained by expression of a (σ-1.1)-(intein)-(chitin binding domain) fusion in minimal media made in D_2O and containing only ^{15}N and ^{13}C as nitrogen and carbon sources. The σ-1.1 thioester was generated by on column intein thiolysis (see Methods). C-terminal constructs were expressed in LB media. N-terminal cysteine was exposed by proteolytic processing with Factor Xa.

The segmental labeling work was initiated before it was determined that residues 95-115 of region 1.1 were unstructured, so a region 1.1 construct containing residues 1-116 (S116G as an artifact of cloning) that had a thrombin site to remove the unstructured N-terminal portion was used (**Figure 4-18**). The final processed construct contained residues 29-116 (S116G). We initially wished to make two segmental labeled constructs : full length σ-A (σ-A(EPL)) and a C-terminally truncated construct lacking region 4 (σ-A(Δ4)). We reasoned that with these two constructs we could determine if there were any contacts between σ-1.1 and σ-4 (**Figure 4-16**). The C-terminal constructs we used contained residues 117(S117C)-399 for the generation of σ-A(EPL) and residues 117(S117C)-318 for the generation of σ-A(Δ4).

Figure 4-17 : Semi-synthetic scheme used to generate segmental labeled σ-A. σ-A(29-115) was expressed in labeled (^{2}H, ^{15}N, ^{13}C) media as a fusion with the GyrA intein and a chitin binding domain affinity tag (top left). The sample was purified by immobilization on a chitin column and the unstructured N-terminal tail by thrombin proteolysis. σ-A(29-115) thioester was released by thiolysis. C-terminal constructs (σ-A(117-399) or σ-A(117-318)) were expressed with an N-terminal hexahistidine purification tag followed by a Factor Xa recognition site (IEGR). Following purification of the C-terminal constructs, the purification tag was removed by Factor Xa proteolysis, exposing an N-terminal cysteine. σ-A(EPL) or σ-A(Δ4) were generated by the native chemical ligation reaction between the thioester of σ-1.1 and the N-terminal cysteine of σ(117-399) or σ(117-318).

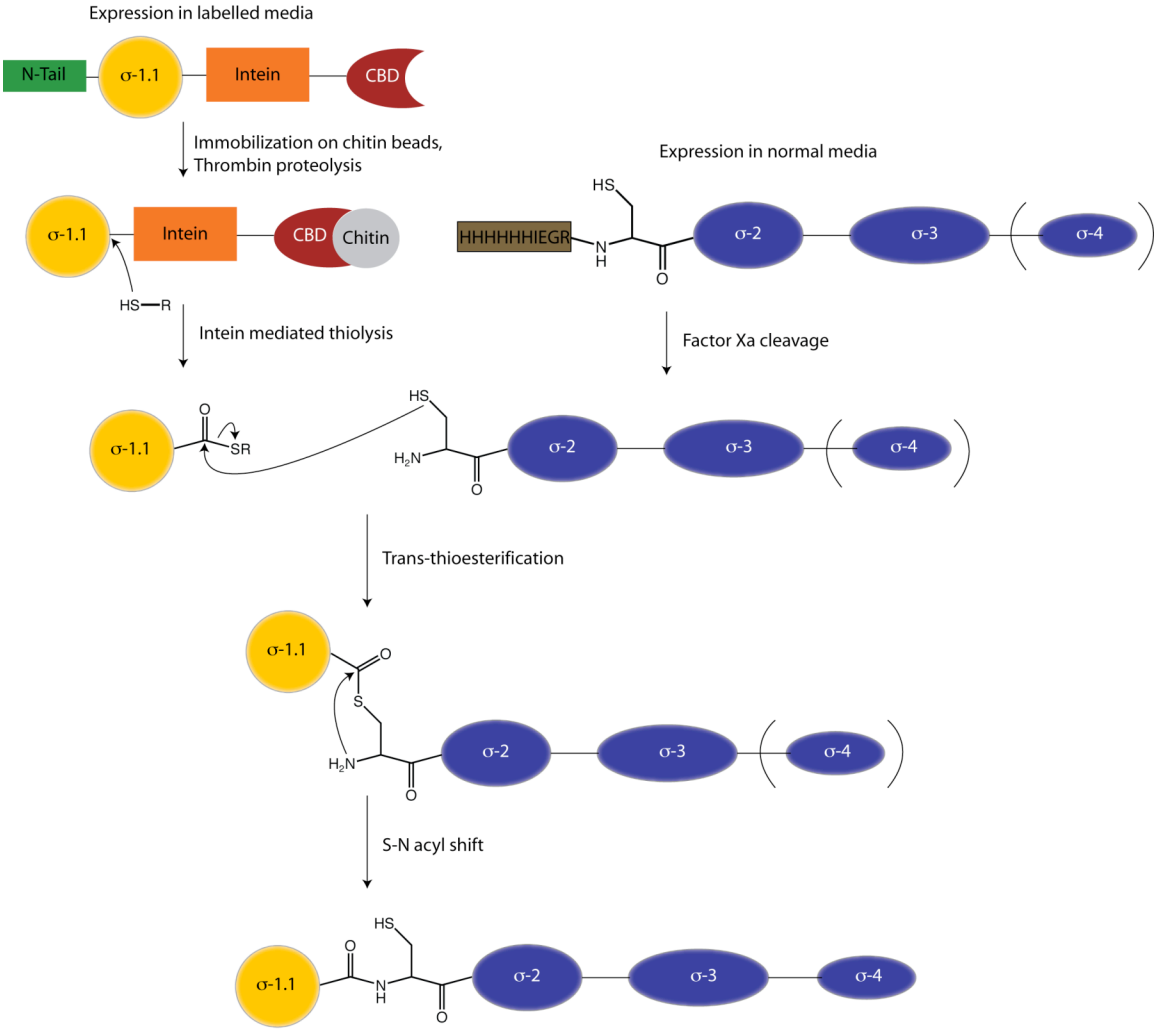


Figure 4-17 : Semi-synthetic scheme used to generate segmental labeled σ-A.

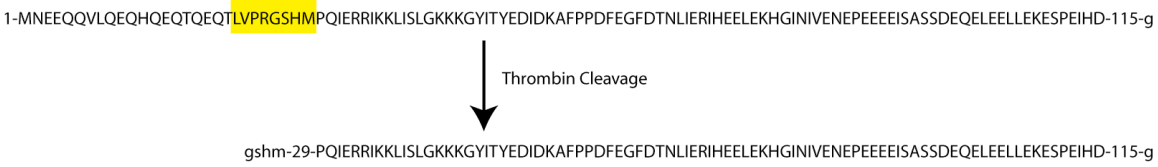


Figure 4-18 : σ-A(29-115) construct used in segmental labeling studies. The construct used for segmental labeling encoded residues 1-115 (top). A thrombin protease recognition site (top, highlighted in yellow) was incorporated by PCR mutagenesis for the removal of the unstructured N-terminus. The final, thrombin processed construct used for structure determination contained residues 29-115 (bottom) with short N- and C-terminal tails that were artifacts of cloning (bottom, lower-case).

The ^{2}H, ^{13}C, ^{15}N labeled region 1.1 thioester was generated with minimal difficulty and could be isolated in multimilligram amounts after chromatography purification (see methods). The C-terminal constructs were more difficult to express and purify. Despite extensive efforts, neither σ-A(117-399) nor σ-A(117-318) could be characterized by mass spectrometry, presumably due to poor ionization properties, so it was never clear what percentage of the N-terminal cysteine was exposed by Factor Xa cleavage. It was also determined that storage in glycerol, which contains glyceraldehyde contaminants was likely capping much of the N-terminal cysteines that were exposed (J. P. Pellois, unpublished data), so the protein had to be prepped immediately prior to ligation. σ-A(117-399) had the additional complication of poor solubility. Solubility was improved through the use of the detergent CHAPSO, but added new difficulties. The expense of CHAPSO prohibited buffer exchange by dialysis, necessitating

buffer exchange using a centrifugation concentration device. Not only was this process slow, but because it involved repeatedly concentrating the protein, it led to sample loss from precipitation. Additionally, as CHAPSO needed to be used at or near the CMC, sometimes the concentration of CHAPSO also rose to levels at which it interfered with gel or HPLC analysis as well as with the NMR spectra.

Also problematic was the purification of the ligation products. The reactions did not go to completion (**Figure 4-19**) so both reactants needed to be separated from the ligation product. Region 1.1 is strongly negatively charged (calculated pI = 4.52), the C-terminal fragments are strongly positively charged (σ-A(117-399) calculated pI = 9.39, σ-A(117-399) calculated pI = 9.28) and the ligation products weakly negative (σ-FL calculated pI = 6.52, σ-Δ4 pI = 5.45). We reasoned that these charge differences could be exploited using ion exchange chromatography to purify the ligation products. This was reasonably successful for σ-Δ4 reaction mixtures. Unfortunately, a portion of both region 1.1 and σ-A(117-399) co-eluted with σ-FL on ion exchange so only a small amount of the σ-FL containing fractions could be isolated. We interpret the co-migration of σ-FL, σ-1.1 and σ-A(117-399) as indicative of the formation of a complex between σ-A(117-399) and σ-1.1. This is possibly the *trans*-inhibited complex observed by Dombroski and Gross [40].

Despite these difficulties, σ-A(Δ4) and σ-A(EPL) were made, albeit in low quantities and of moderate purity (**Figure 4-19**). NMR spectra from both σ-A(Δ4) and σ-A(EPL) were obtained, but were of poor quality and could not be fully

interpreted. In fact, σ-A(EPL) was found to have degraded slightly over the course of the NMR experiment (**Figure 4-20**). The only thing that was clear in the segmentally labeled spectra is that in both σ-A(Δ4) and σ-A(EPL), the fingerprint region in the $^{1}H[^{15}N]$ HSQC of region 1.1 is perturbed relative to isolated region 1.1 (**Figure 4-21**). The spectra differ both in chemical shifts of the resonances, which suggests a change in a residues local environment and in broadening of peaks which suggests exchange processes on the NMR timescale that could indicate transient interactions. Thus, there are likely contacts between region 1.1 and other portions of the protein although, due to the quality of the spectra, it could not be determined which residues of region 1.1 were interacting and what they were interacting with.

The only major difference between the wild-type σ-A and the σ-A(EPL) construct is that σ-A(EPL) lacks the first 28 residues. These residues are unstructured and were removed out of a desire for a simplified NMR spectra. However, these residues might be required for protein stability and solubility. Earlier attempts to express region 1.1 without the N-terminal residues were unsuccessful, lending credence to this hypothesis.

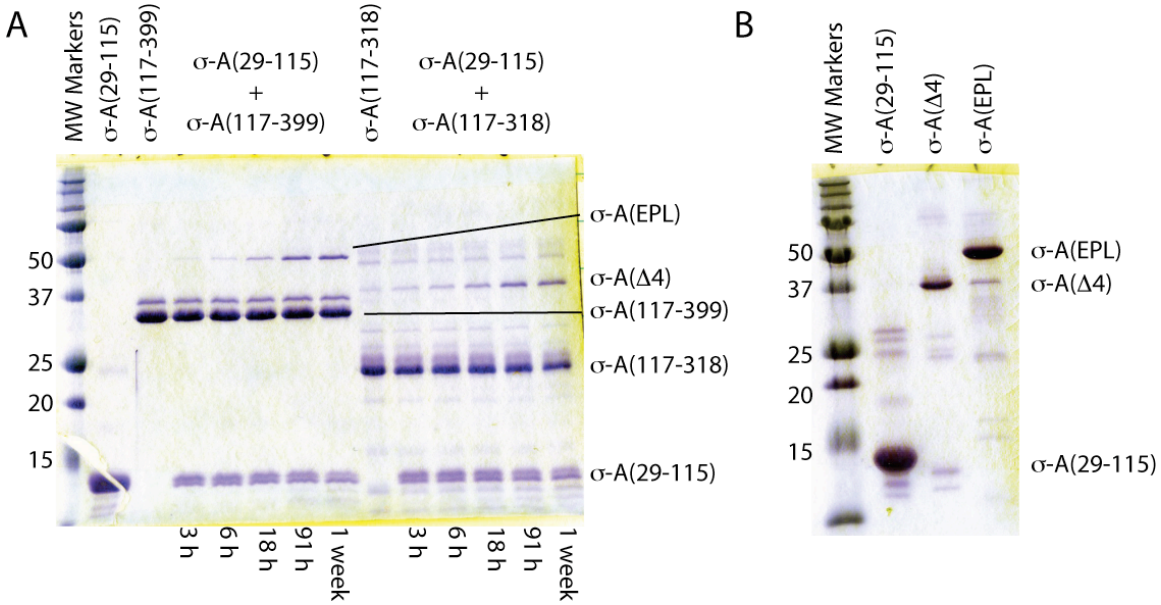


Figure 4-19 : Segmental labeling of σ-A(Δ4) and σ-A(EPL). (**A**) σ-A(EPL) (residues 29-399) and σ-A(Δ4) (residues 29-318) can be generated by EPL although the reaction is slow (1 week at room temperature) and inefficient. (**B**) Purified NMR samples used to generate the spectra shown in **Figure 4-21**. Protein samples were resolved by SDS-PAGE (15% acrylamide) and visualized by coomassie stain.

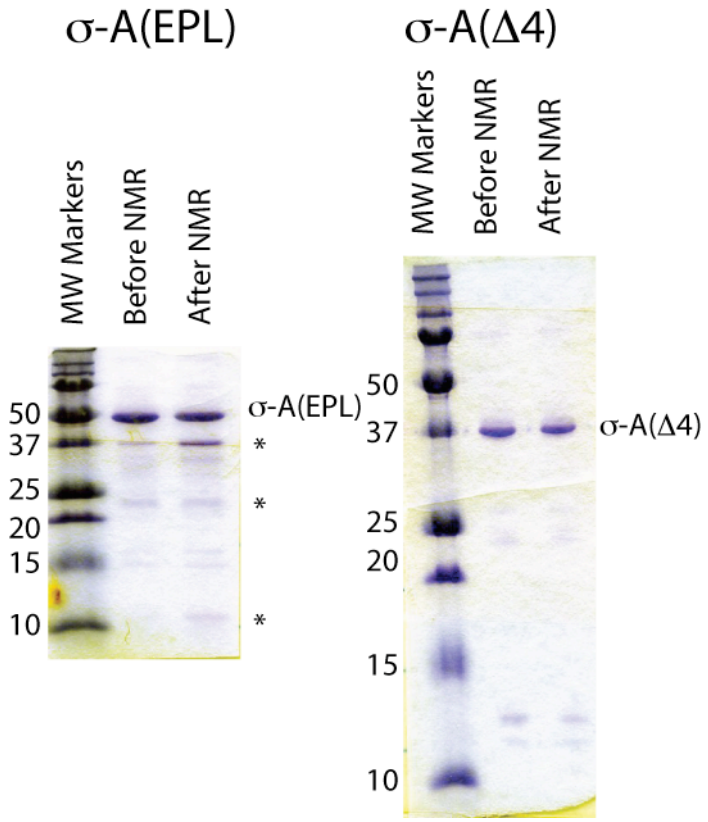


Figure 4-20 : Degradation of segmental labeled σ-A(EPL). NMR samples from **Figure 4-19** were analyzed before and after NMR spectra were acquired. In the case of σ-A(EPL), degradation products (marked by *) increased over the course of the experiment. Protein samples were resolved by SDS-PAGE (15% acrylamide) and visualized by coomassie stain.

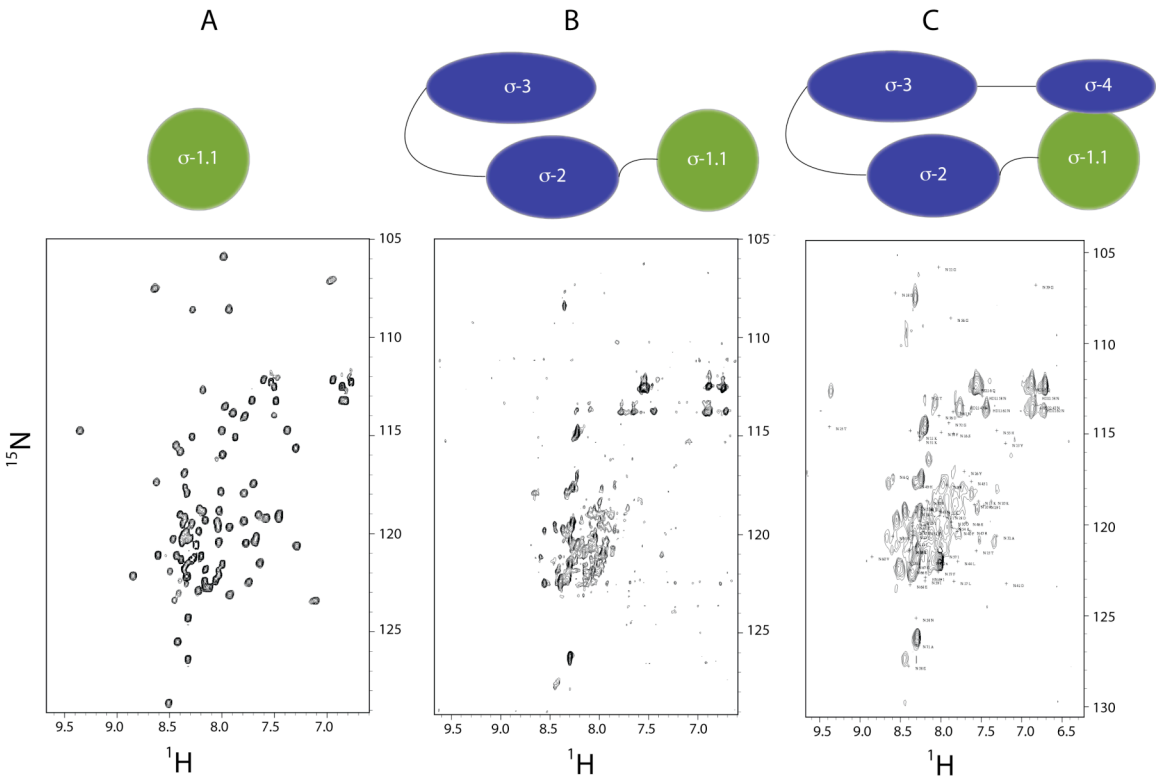


Figure 4-21 : ^{1}H[^{15}N] HSQC spectra of σ-A(29-115) in σ-A(EPL) and σ-A(Δ4). (**A**) ^{1}H[^{15}N] HSQC spectrum of region 1.1 (residues 29-115) in isolation. (**B**) ^{1}H[^{15}N] HSQC-TROSY spectrum of region 1.1 in σ-A(Δ4) (residues 29-318). (**C**) ^{1}H[^{15}N] HSQC-TROSY spectrum of region 1.1 in σ-A(EPL) (residues 29-399). All spectra were acquired at 298 K.

When the structure of region 1.1 was determined it was also found that residues from ~95-116 were also unstructured. We reasoned that by only labeling residues 1-95, we would be able to obtain spectra that were simple enough to interpret. By retaining residues 1-28, we hoped to obtain a stable, soluble protein. We decided to perform the ligation between residues 96 and 97. In this way, all of the structured residues would be labeled. Additionally, we

reasoned the S97C mutation that would be required for ligation was conservative enough to be unlikely to affect the σ function significantly.

To ensure that the spectra would be interpretable in the full length, ligated construct, we first generated ^{15}N, ^{13}C labeled σ-A(1-95) (**Figure 4-22**). The spectrum obtained was well dispersed, with sharp peaks, suggesting that meaningful conclusions could be drawn from the spectra in the full length protein (**Figure 4-23**). The full length protein could be obtained by EPL (**Figure 4-24A**) and behaved as wild type in abortive initiation assays (**Figure 4-24B**). However, the yield of the ligation reaction was low, and only small amounts of protein could be isolated after purification (data not shown). It is possible that with further optimization of the expression and purification of the protein building blocks, the ligation reaction conditions and especially the purification of the reaction products, we would have been able to generate an NMR sample of sufficient quality. However, the experience with ligation between residues 116 and 117 suggested that this optimization would be time consuming and difficult. Consequently, we decided to take a different approach involving crosslinking to probe potential intramolecular interactions.

1-MNEEQQVLQEQHQEQTQEQTQEQKETLPPQIERRIKKLISLGKKKGYITYEDIDKAFPPDFEGFDTNLIERIHEELEKHGINIVENEPEEEEISA-95-g

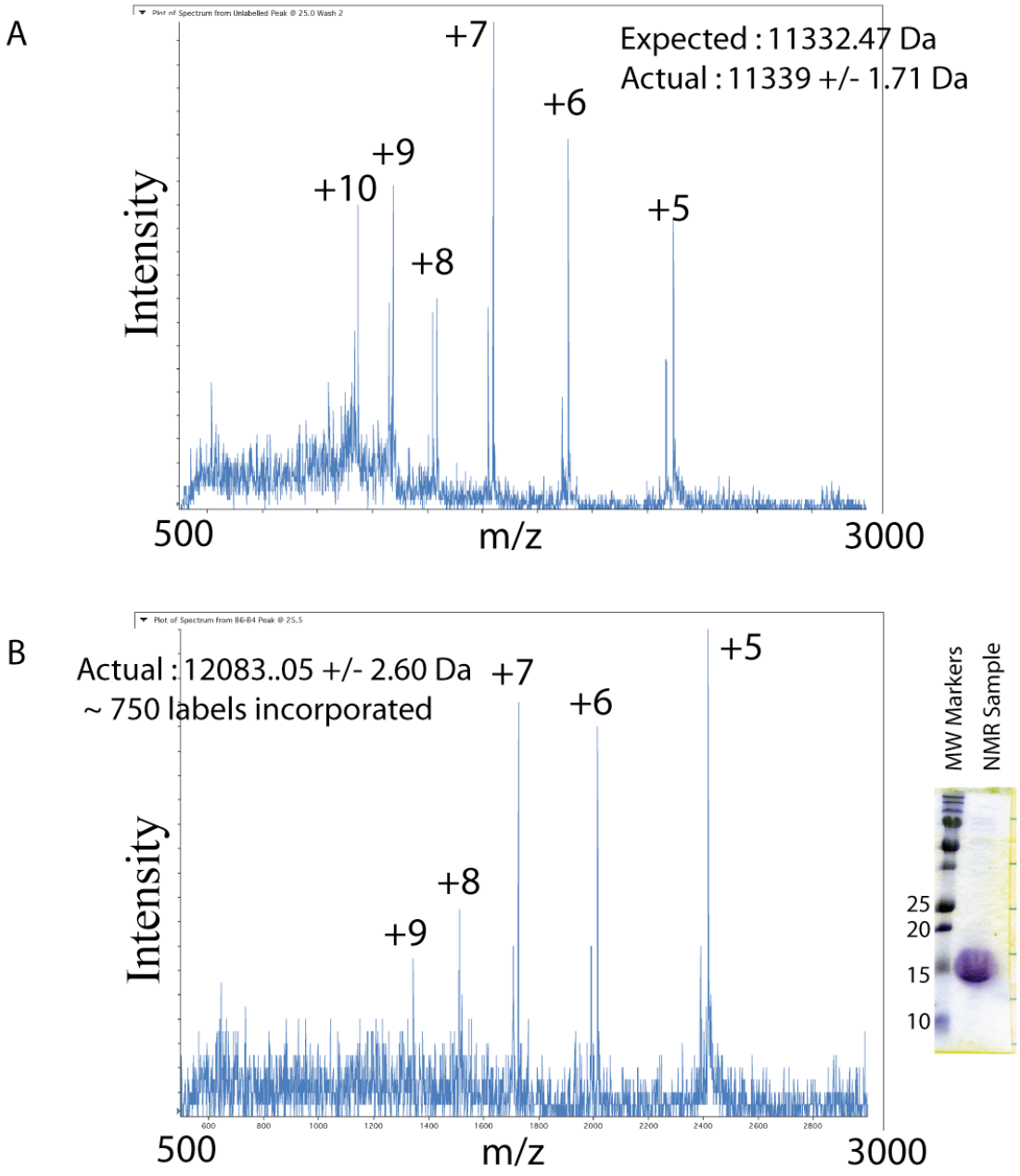


Figure 4-22 : Expression of σ-A(1-95). (**A**) ESI-MS of unlabeled σ-A(1-95) demonstrates the correct construct is expressed. (**B**) (left) ESI-MS of ^{15}N, ^{13}C labeled σ-A(1-95) shows that approximately 750 spin labels are incorporated. (right) SDS-PAGE of the NMR sample used to generate the spectrum shown in **Figure 4-23**. The amino acid sequence of the construct is shown at the top of the figure.

1-MNEEQQVLQEQHQEQTQEQTQEQKETLPPQIERRIKKLISLGKKKGYITYEDIDKAFPPDFEGFDTNLIERIHEELEKHGINIVENEPEEEEISA-95-g

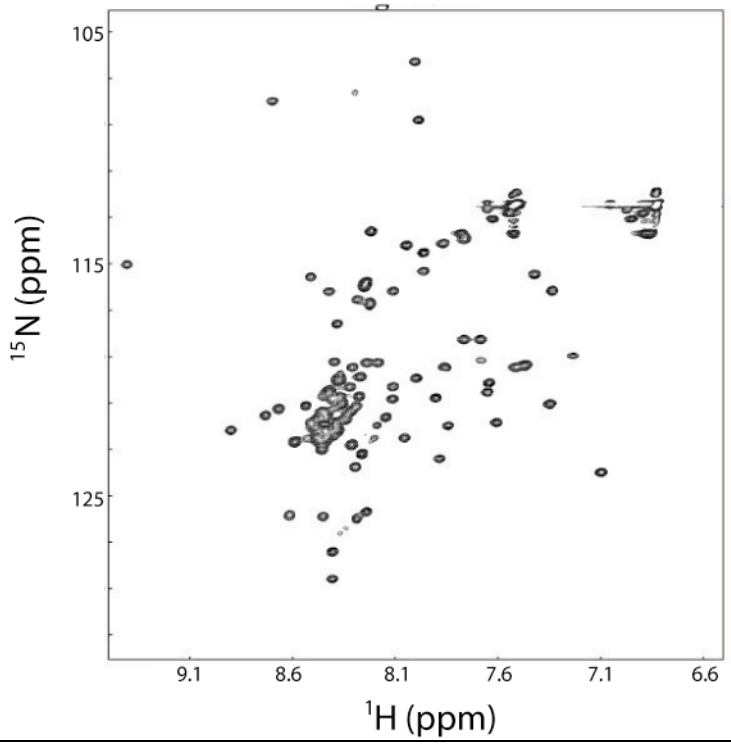


Figure 4-23 : ^{1}H[^{15}N] HSQC spectrum of σ-A(1-95). ^{1}H[^{15}N] spectrum of ^{15}N, ^{13}C labeled σ-A(1-95) at 298 K (protein concentration 300 μM in 10 mM NaPi pH 6.5, 100 mM NaCl, 1mM EDTA). The amino acid sequence of the construct is shown at the top of the figure.

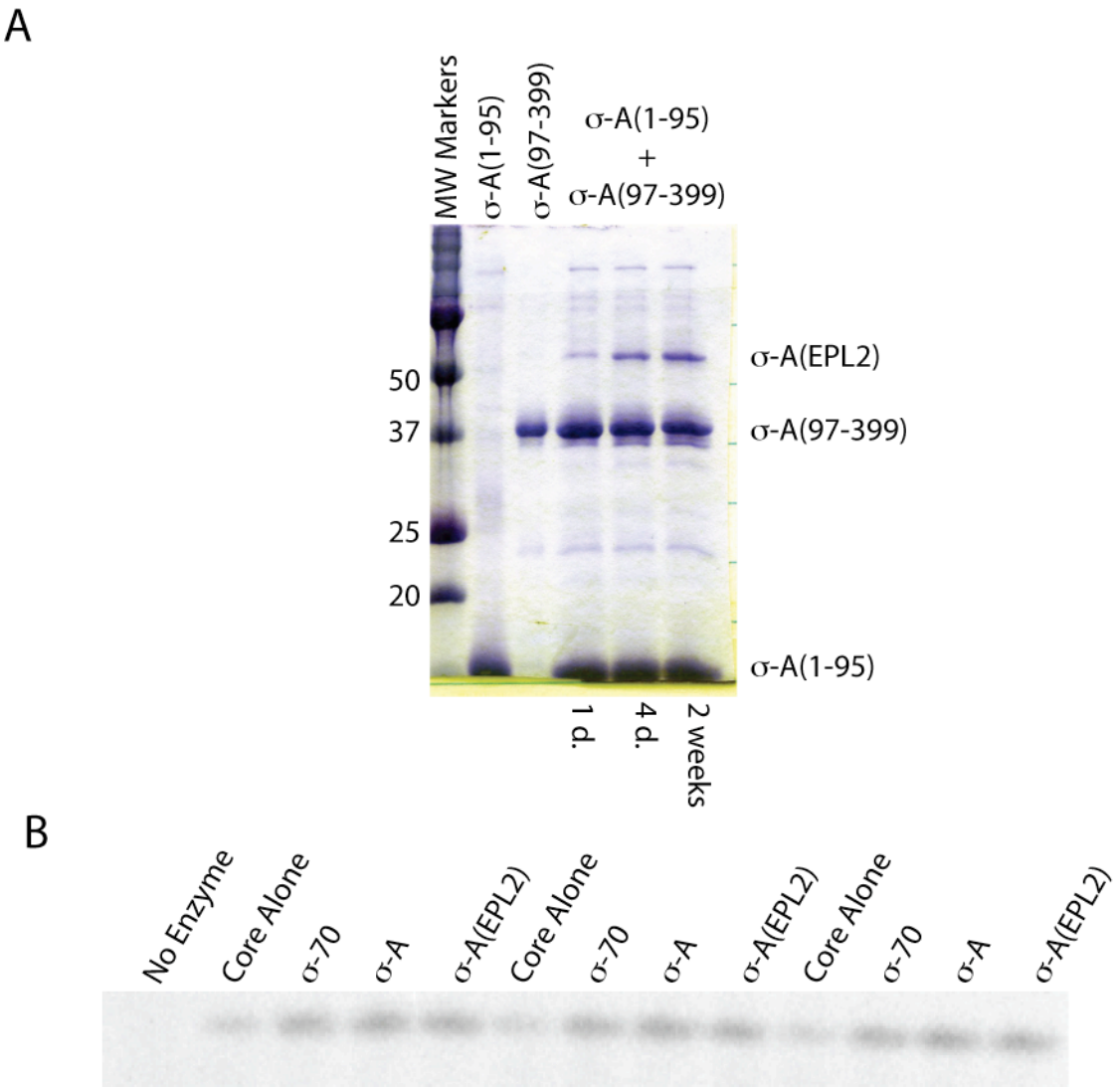


Figure 4-24 : Semi-synthesis of σ-A(EPL2). (**A**) σ-A(EPL2) can be generated by ligation of σ-A(1-95) thioester and σ-A(97-399) containing N-terminal cysteine although, as in the case of ligation at residue 116 the reaction is slow and inefficient. Protein samples were resolved by SDS-PAGE (15% acrylamide) and visualized by coomassie stain. (**B**) Purified σ-A(EPL2) behaves as wild type (σ-A) in abortive transcription initiation reactions on a T7A1 promoter in the presence of *E. coli* RNAP core and the indicated σ factor. Reaction products ($CpA^{32}pU$ trinucleotide) were resolved by denaturing SDS-PAGE and analyzed by autoradiography.

Section 4.2.3 : Probing autoinhibition by crosslinking

Since we were unable to characterize interdomain interactions with region 1.1 by NMR, we chose instead to investigate possible intramolecular interactions using a crosslinking strategy. The strategy we used is shown schematically in **Figure 4-25** and the crosslinker used is shown in **Figure 4-26**. We reasoned that transfer of a biotin tagged crosslinker from region 1.1 to another domain would indicate an interdomain interaction.

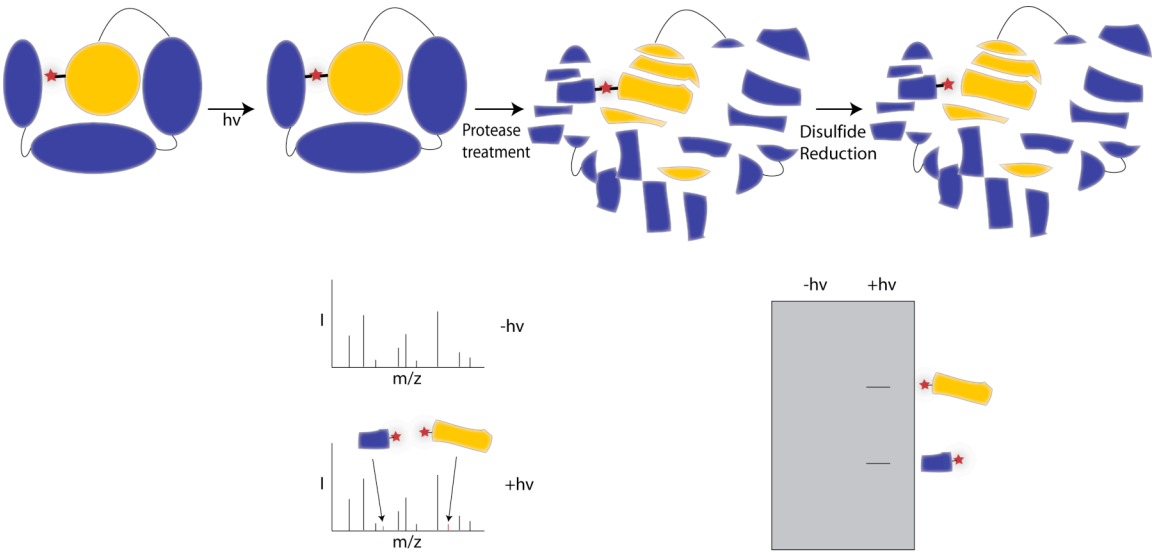


Figure 4-25 : Crosslinking strategy for detecting interdomain interactions. A tagged crosslinker (red star) is attached by disulfide exchange to a surface cysteine. Following crosslinking and digestion, the disulfide is reduced, transferring the tag to regions in close proximity to region 1.1. The tag transfer is detected by mass spectrometry or by gel based methods.

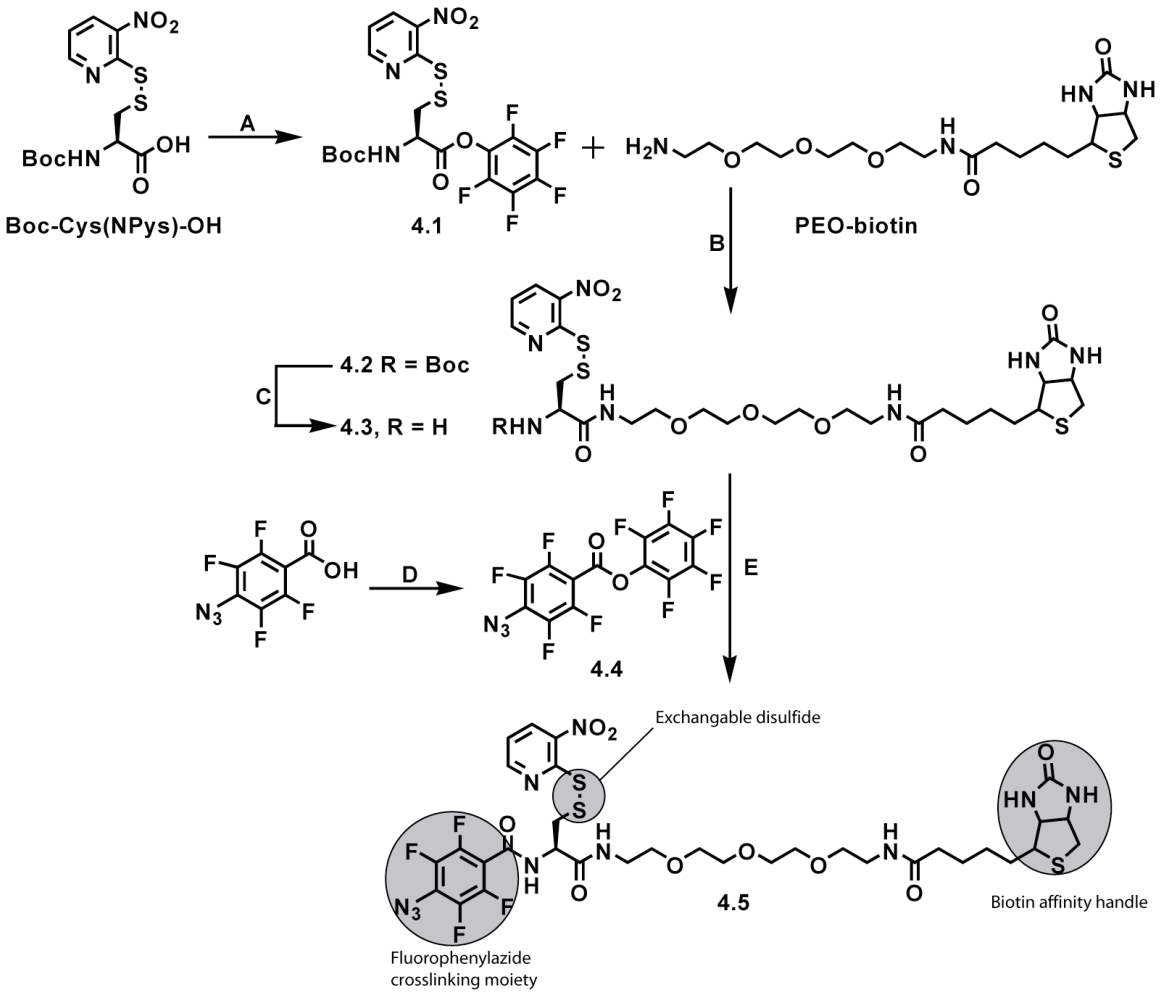


Figure 4-26 : Synthesis of crosslinker. A novel crosslinker was synthesized for use in these studies by Matthew Pratt (The Rockefeller University). (**A**) pentafluorophenyl trifluoroacetate, pyridine, DMF, 45 min, 85%; (**B**) diisopropylethylamine, DMF, 16 h, 83%; (**C**) trifluoroacetic acid, H_2O, 1h, 94%; (**D**) pentafluorophenyl trifluoroacetate, pyridine, DMF, 45 min, 80%; (**E**) diisopropylethylamine, DMF, 16 h, 74%. Three functional moieties were incorporated. A fluorophenylazide was included for crosslinking, an exchangeable disulfide was included for attachment to a surface cysteine and a biotin served as both an affinity tag for enrichment of crosslinked samples and as a marker for Western blotting.

Using the structure of region 1.1, we selected ten surface residues to mutate to cysteine for crosslinker attachment (**Figure 4-27, Table 4-2**). These were designed to give broad coverage of the surface of region 1.1. Since *T. maritima* σ-A contains no native cysteine, no additional mutations needed to be made. Each construct thus contained only a single cysteine, allowing site specific attachment of the crosslinker.

Table 4-2 : σ-A proteins used for crosslinking

Mutation	MW (Da) (Calculated)	MW (Da) (Actual)
Wild Type	46815.5	46808 ± 12
T26C	46817.5	46825 ± 10
E32C	46789.5	46794 ± 9
S40C	46831.6	46827 ± 11
E51C	46789.52	?
D60C	46803.6	46817 ± 10
F64C	46771.5	46789 ± 10
E70C	46789.5	46811 (reconstruct)
E77C	46789.5	?
V84C	46819.5	46793 (reconstruct)
S94C	46831.6	46837± 13
Δ1.1	32867.4	32877 ± 11

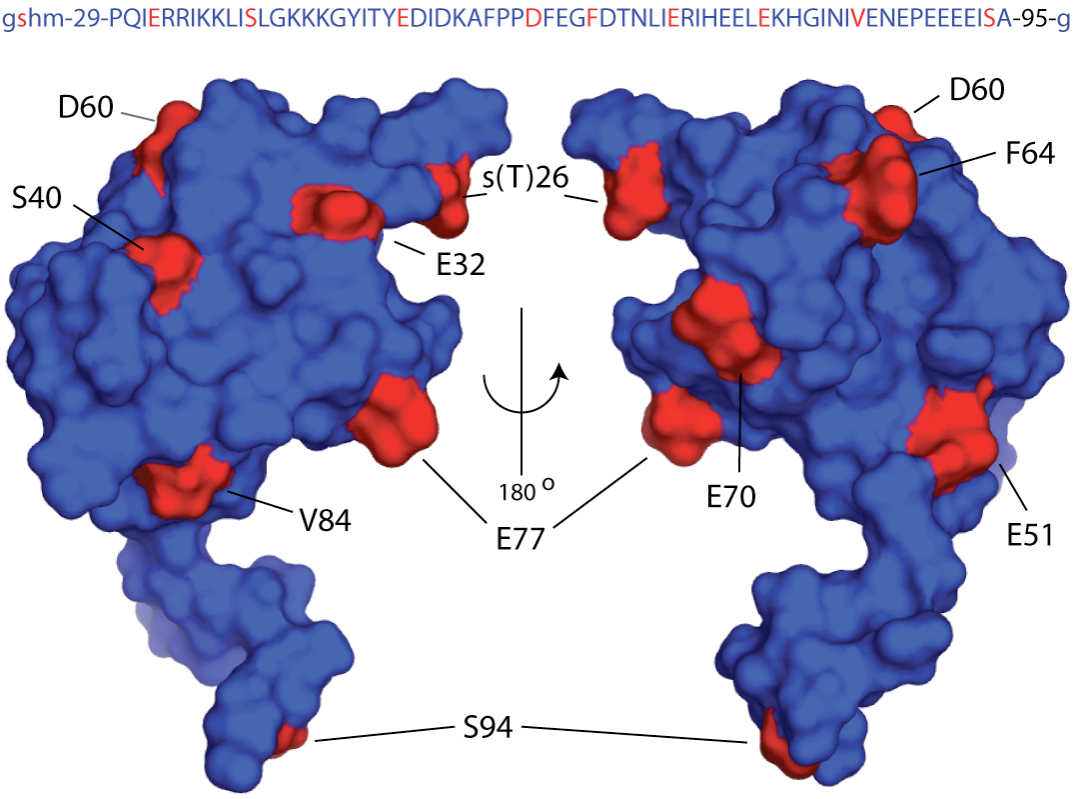


Figure 4-27 : Sites of crosslinker attachment. Surface rendering of region 1.1 shown in the same orientation as **Figure 4-11**. Shown in red are the residues that were mutated to cysteine for attachment of the crosslinker. T26 was mutated to serine in the construct used for structure determination and is thus indicated in the figure as "s(T)26". Images were generated using MacPYMOL (Delano Scientific). The amino acid sequence of the construct is shown at the top of the figure with residues that were mutated to cysteine in red, and those that were left unchanged in blue.

Of the ten point mutations made, eight could be expressed and purified (**Figure 4-28**). The failure to obtain two of the mutants is mysterious. DNA sequencing confirmed that the cloning was correct. The E51C construct appears to express in very low quantities and behaves strangely (**Figure 4-28**). The SDS-

PAGE mobility of E51C differs from the other mutants and from wild type. Additionally, a mass of the purified protein could not be obtained. No evidence of expression could be detected for the E77C protein (data not shown).

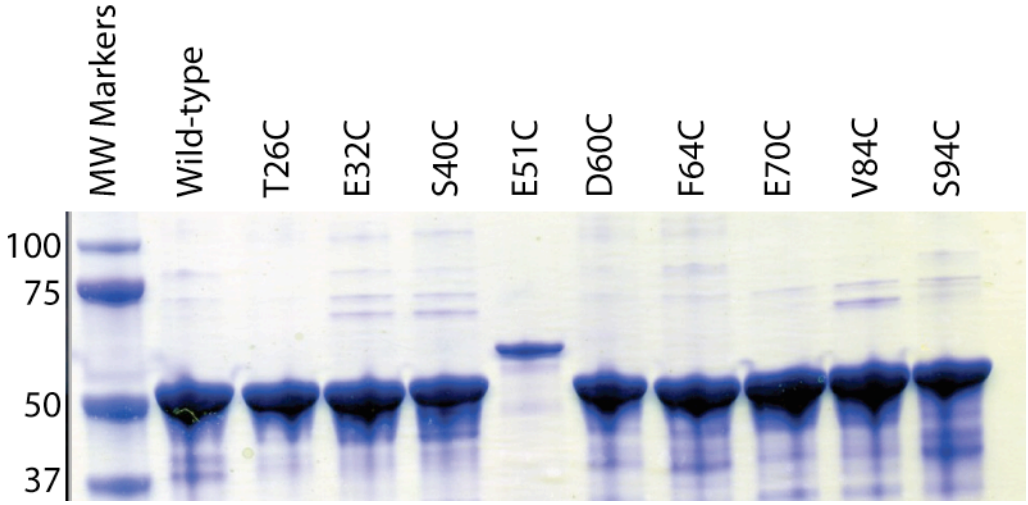


Figure 4-28 : Purification of σ-A cysteine mutants. The indicated purified σ-A cysteine mutants were resolved by SDS-PAGE (4-12% Acrylamide) and visualized by coomassie stain. When obtained, masses of the constructs are reported in **Table 4-2**.

The 8 mutants that were successfully generated could be labeled with the crosslinker **4.5** using a disulfide exchange process (**Figure 4-29**). In all cases, the crosslinker could be removed by reduction with DTT. However, after the labeled proteins were irradiated with a laser at 325 nm, the crosslinker could no longer be removed with DTT, suggesting that crosslinking to the protein had occurred. Labeled σ factors behaved as wild type in abortive trancription initation assays (**4-30A**). The presence of the label did not interfere with DNA binding autoinhibition by region 1.1 (**4-30B**).

Figure 4-29 : Crosslinker 4.5 can be attached through a disulfide. (**A**) Indicated σ-A mutants were labeled with crosslinker by disulfide exchange. Samples were then treated with laser light at 325 nm and reduced with DTT. (**B**) Samples were then resolved by SDS-PAGE and analyzed by Western blot (Streptavidin-HRP). DTT reduction removed the crosslinker (UV "–" samples) indicating that the attachment was through a disulfide bond to cysteine.

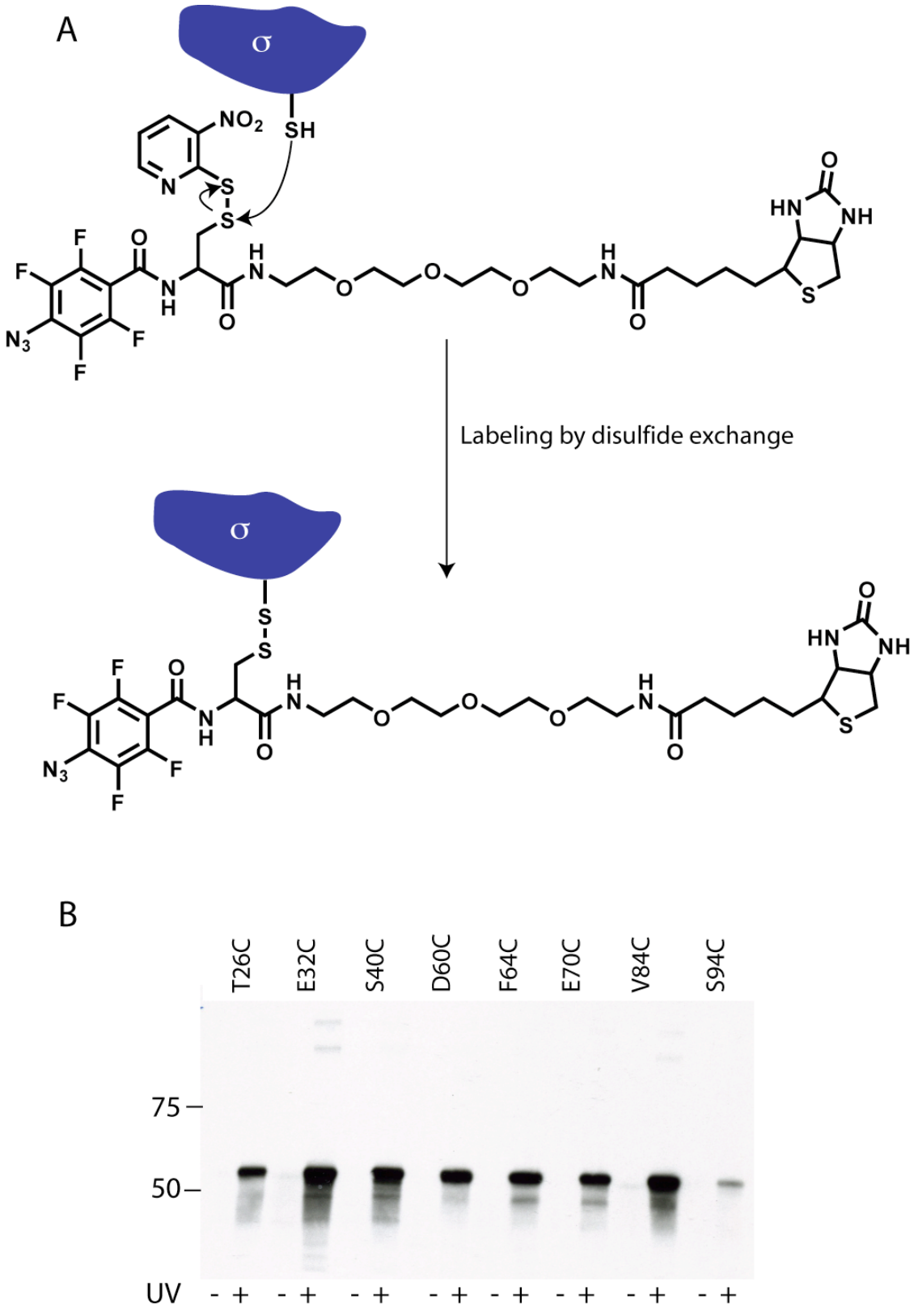


Figure 4-29 : Crosslinker 4.5 can be attached through a disulfide.

Figure 4-30 : Attachment of crosslinker does not affect σ-A function. Wild type σ-A and σ-A cysteine mutants were labeled with crosslinker and tested for function. (**A**) Three examples of abortive transcription initiation on a T7A1 promoter in the presence of *E. coli* RNAP core and the indicated σ factor. Reaction products ($CpA^{32}pU$ trinucleotide) were resolved by denaturing SDS-PAGE and analyzed by autoradiography. σ-70 was provided by Andrey Feklistov (Darst lab). It is believed that the low overall level of σ-A activity in these particular assays is because the protein used was frozen and not due to the incorporation of crosslinker since wild type σ-A (which does not bind crosslinker) displays similarly low activity. (**B**) DNA binding of σ-A constructs measured by gel shift. 5'-fluorescein labeled -35 promoter element DNA (5 μM) was incubated with the indicated σ-A (10 μM) then resolved on a 1% agarose gel and imaged by fluorescence. Protein:DNA binding was indicated as a shift in gel mobility (indicated by arrow). As expected, only σ-A that lacks region 1.1 (Δ-1.1) binds DNA in a manner stable enough to result in a gel shift.

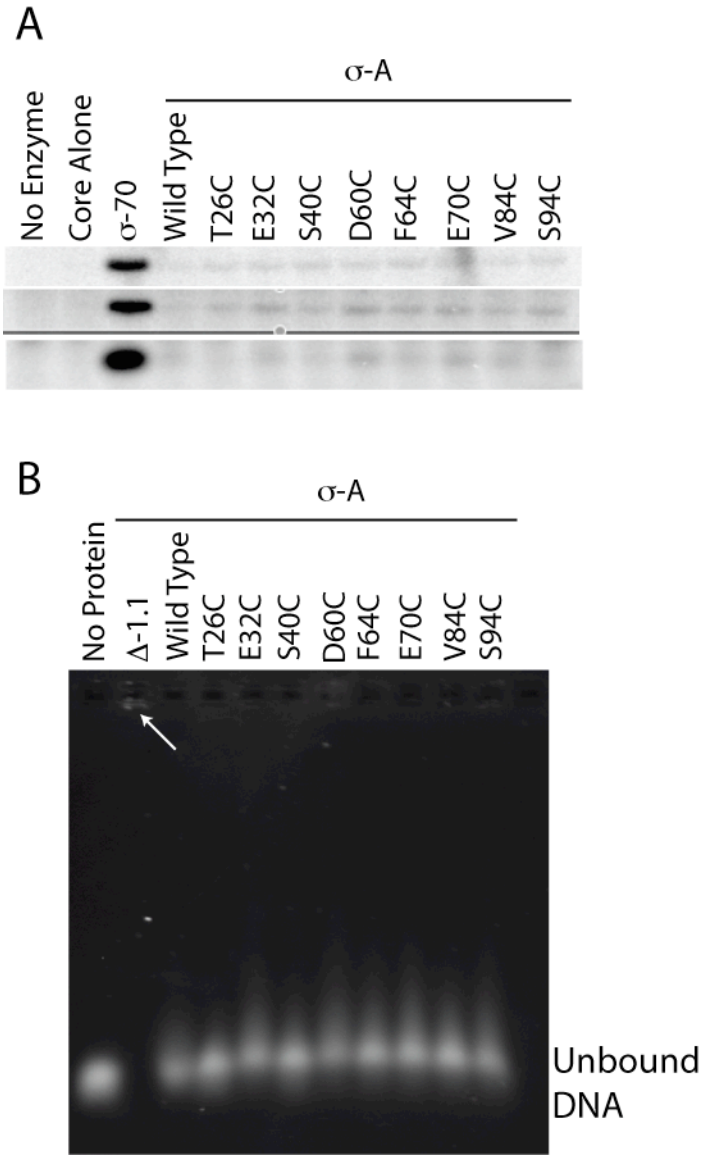


Figure 4-30 : Attachment of crosslinker does not affect σ-A function.

Having established that we could form crosslinks from region 1.1, we next set about determining what portions of the σ factor region 1.1 was contacting. We made numerous attempts to directly identify a crosslinked fragment by mass spectrometry. Crosslinked proteins were reduced with DTT or TCEP treatment and treated with iodoacetamide to cap the cysteine side chain. Samples were then resolved by SDS-PAGE. Silver stained bands were then digested by in gel trypsinolysis, extracted and analyzed by MALDI-TOF mass spectrometry. In no case could the unambiguous identification of a crosslinked peptide be made. Additionally, crosslinked samples were digested with trypsin in solution and analyzed by LS-MSMS and, once again, no crosslinked peptide could be identified.

We then turned to gel based methods to map the sites of crosslinking. We decided to use a cyanogen bromide (CNBr) digestion of crosslinked samples to identify interdomain crosslinks. CNBr cleaves after methionine residues, of which there are 13 in σ-A. The predicted CNBr digestion pattern is shown in **Figure 4-31**. The majority of crosslinking is predicted to occur locally, to nearby portions of region 1.1. Fortunately, region 1.1 is entirely contained within a single CNBr fragment that is more than twice as large as the next largest. Thus, region 1.1 can be easily resolved from region 2-4 CNBr fragments by SDS-PAGE, allowing one to distinguish between local and inter-domain crosslinking.

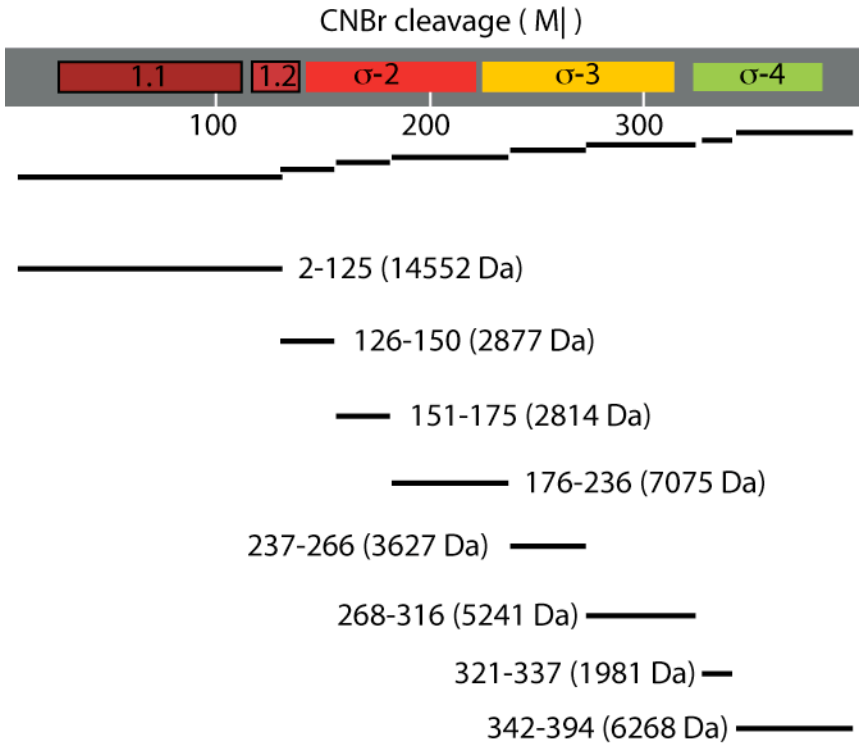


Figure 4-31 : Predicted CNBr digestion of σ-A. The domain structure of σ-A (top) is shown with the predicted CNBr digestion fragments below. Note that the fragment containing region 1.1 (2-125) is more than twice as large as the next largest fragment (176-236).

Crosslinker **4.5** was attached to a subset of the σ-A cysteine mutants and samples were irradiated with 325 nm laser light. Samples were then digested with CNBr, resolved by non-reducing SDS-PAGE and analyzed by Western blot with streptavidin-HRP (**Figure 4-32**). As the disulfide bond had not been reduced, the biotin tag remained attached to region 1.1. Four biotin containing bands were observed in all samples. We reasoned that the smallest represented the CNBr fragment 2-125. The gel mobility was more consistent with a 22 kDa protein than a 14.5 kDa protein. However, region 1.1 displays aberrant gel

mobility. For instance, see **Figure 4-22** in which σ-A(1-95) migrates as a 15 kDa protein even though its actual mass is 11.3 kDa. We interpret the larger fragments as partial digestion products containing region 1.1, likely 2-150, 2-175 and 2-236. Additionally, some higher molecular weight biotin containing species were observed in the UV irradiated D60C mutant. These most likely represent interdomain crosslinks.

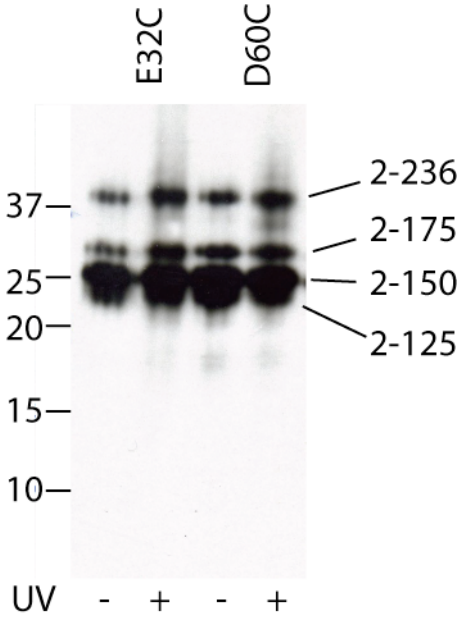


Figure 4-32 : Identification of region 1.1 containing CNBr fragments. Crosslinker was attached to the indicated σ-A constructs and samples were treated with UV (+) or left untreated (-). Samples were digested with CNBr and resolved by non-reducing SDS-PAGE (10-20% acrylamide, tris-tricine buffer system). Biotin containing bands were visualized by Western blot (Streptavidin-HRP). The smallest biotin containing band was interpreted to represent the smallest region 1.1 containing CNBr fragment. Identification of larger bands is based on the predicted size of partial CNBr digestion fragments.

In order to determine to what part of the protein crosslinking had occurred, the crosslinked samples were digested with CNBr followed by reduction to transfer the biotin tag. In all samples, the crosslinker was observed on the fragments that contain region 1.1 (**Figure 4-33**). This is consistent with local crosslinking to nearby residues. The presence of biotin in bands that were smaller than the region 1.1 containing bands were interpreted to indicate interdomain contacts. E32C and S40C did not display strong crosslinking to fragments from any other region. The other mutants all showed crosslinking to fragments that were smaller than region 1.1, and must therefore come from domains 2-4.

We reasoned that if the sample was enriched for biotin containing fragments following CNBr digestion, we would be able to identify the crosslinked bands by gel extraction and mass spectrometry. CNBr digested crosslinked samples of the E32C and D60C mutants (we chose to work with one mutant that formed inter-domain crosslinks (D60C) and one that did not (E32C)) were enriched by application to monomeric avidin beads. We were unable to elute crosslinked sample with excess biotin, but were able to by boiling in SDS loading buffer. The enriched sample revealed a more complex crosslinking pattern than seen before enrichment (**Figure 4-34A**). Unfortunately, the enriched protein was not obtained in quantities that could be detected by silver stain (**Figure 4-34B**). As such, we were unable to extract sufficient amounts of protein for identification by MS.

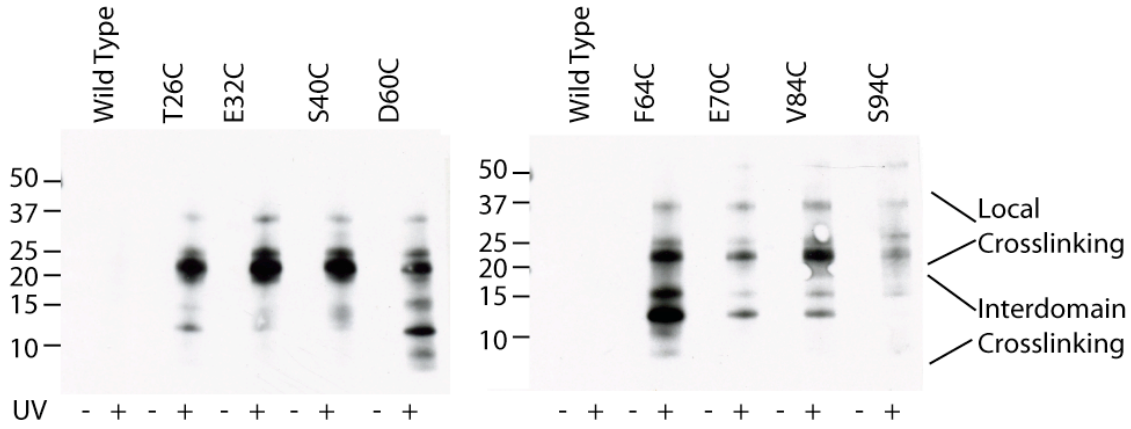


Figure 4-33 : Interdomain crosslinks from region 1.1. Crosslinker **4.5** was attached at the indicated residues and samples were irradiated with laser light. Samples were then digested with CNBr, separated by reducing SDS-PAGE (10-20% acrylamide, tris-tricine buffer system) and probed for biotin by Western blotting (HRP-streptavidin). Samples that were not irradiated retained no biotin. Biotin containing fragments with apparent gel mobility of 20 kDa or larger were interpreted to indicate local crosslinking to region 1.1. Biotin containing fragments that ran as if they were smaller than region 1.1 containing bands were interpreted to represent interdomain crosslinking.

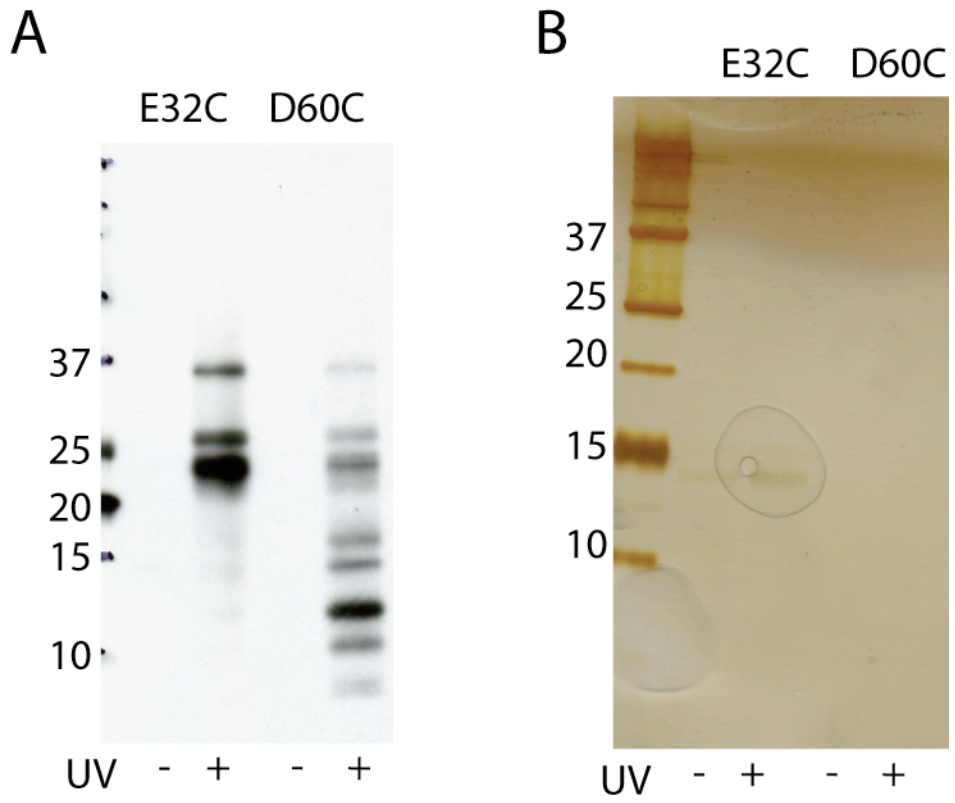


Figure 4-34 : Avidin enrichment of crosslinked samples. The indicated cysteine mutants of σ-A were crosslinked and digested with CNBr. Following reduction, samples were enriched using monomeric avidin beads. Protein was eluted by boiling in SDS loading buffer, resolved by SDS-PAGE (10-20% acrylamide Tris-Tricine gel) and analyzed by western blot (Streptavidin-HRP, **A**) and by silver stain (**B**).

We next attempted to identify CNBr digestion fragments in samples that were not enriched for biotin. We reasoned that biotin containing fragments could be identified by gel mobility. Crosslinked, reduced samples were digested and run on a 20 cm gel in duplicate. One of the replicates was blotted with HRP-streptavidin (**Figure 4-35A**). The other sample was silver stained (**Figure 4-35B**), digested with trypsin and peptide fragments were extracted and identified by mass spectrometry. Three fragments could be identified in this manner. The region 1.1 containing band 2-150 was identified with a gel mobility consistent with our earlier identification in **Figure 4-32**. The smallest crosslinking band was determined to be made up of two fragments, comprising residues 342-394 and 342-399 and thus represents region 4.2, with some of region 4.1 included. Unfortunately, all other bands were found to contain tryptic fragments from multiple regions of σ-A, likely indicating the co-migration of multiple digestion products.

Figure 4-35 : Crosslinking performed on large scale for MS identification of bands. Indicated σ-A mutants were labeled, exposed to UV light and digested with CNBr. Samples were then resolved by reducing SDS-PAGE (~300 μg total protein per lane, 20% acrylamide) and analyzed by Western blotting (**A**, Streptavidin-HRP) or by silver stain (**B**). Silver stained bands were excised, digested with trypsin and analyzed by MALDI-TOF and ESI mass spectrometry. Three bands could be identified unambiguously, as indicated. Other bands contained tryptic fragments that indicated that multiple species were present in each.

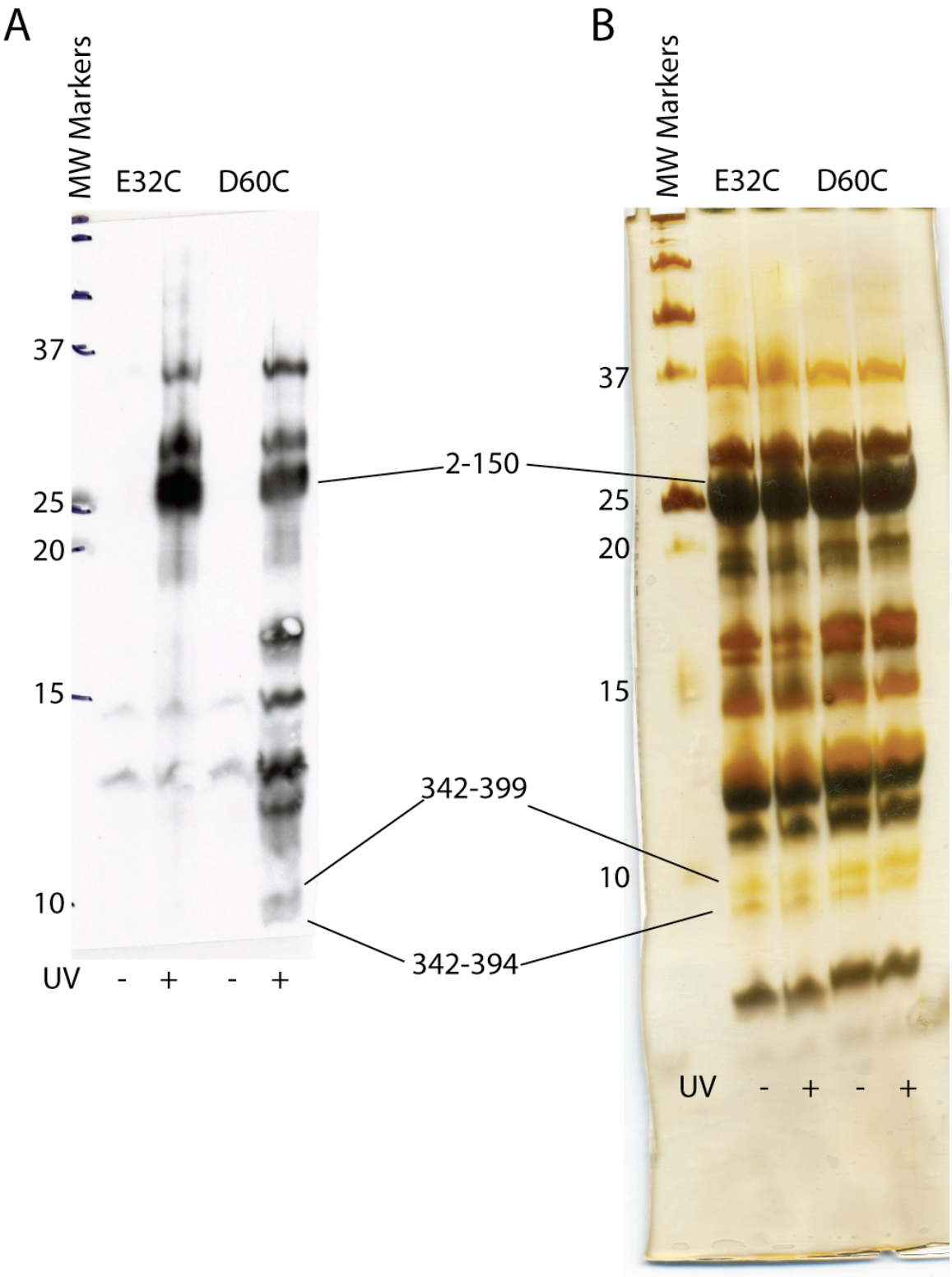


Figure 4-35 : Crosslinking performed on large scale for MS identification of bands.

Figure 4-36 : Two dimensional separation of σ-A CNBr digest by RP-HPLC and SDS-PAGE. (**A**) Wild-type σ-A was digested with CNBr and separated by reverse phase HPLC. Peaks were collected and identified by ESI-MS. (**B**) Fractions with the indicated retention time on the RP-HPLC trace from (**A**) were analyzed by SDS-PAGE (10-20% acrylamide, tris-tricine buffer system). A crude CNBr digest of σ-A (Full length) is included for comparison. Indicated to the right of the gel image is the domain structure of σ-A and identifications of fragments that could be made based on masses obtained from (**A**).

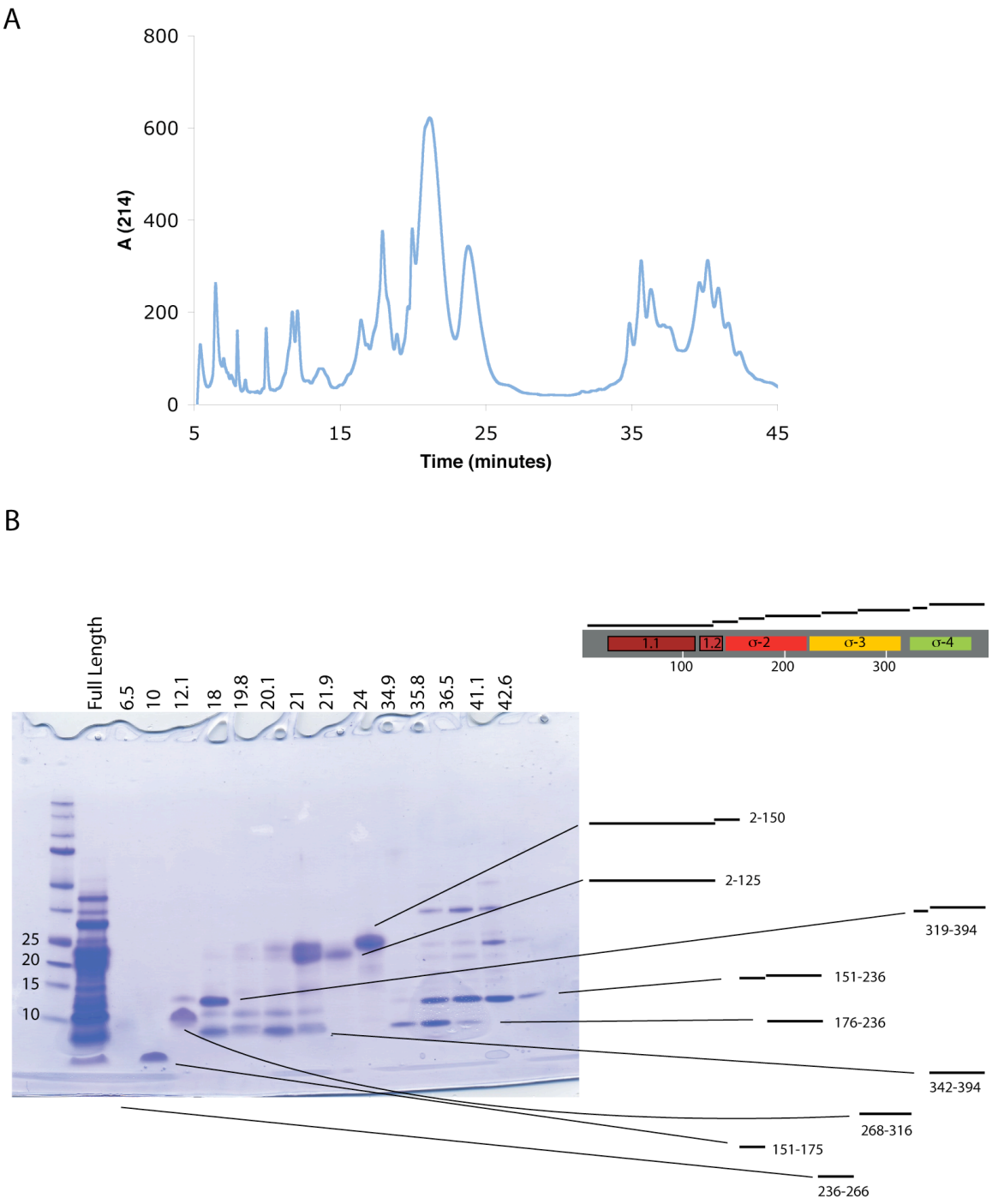


Figure 4-36 : Two dimensional separation of σ-A CNBr digest by RP-HPLC and SDS-PAGE.

In order to identify additional CNBr fragments, wild type σ–A was digested with CNBr and separated by RP-HPLC. Individual peaks were collected, analyzed by mass spectrometry then lyophilized and resolved by SDS-PAGE (**Figure 4-36**). The fragment corresponding to residues 342-394 was identified by mass and found to display the same gel mobility as the smallest crosslinked band. Region 1.1 was also identified by this method. It was also determined that, as suggested by earlier results, numerous digestion fragments co-migrate between 10 and 15 kDa. This made identification of crosslinked bands based on gel mobility in this region difficult to impossible.

In order to aid in the identification of crosslinked fragments, we turned to BNPS-skatole (2-(2'-Nitrophenylsulfenyl)-3-methyl-3-bromoinolenine) digestion, which cleaves following tryptophan residues. There are three tryptophan residues in σ-A : W201, W213 and W214. Thus, BNPS-skatole digestion effectively cuts σ-A in half (**Figure 4-37**). Crosslinked protein samples were digested with BNPS-skatole, separated by reducing SDS-PAGE and blotted with streptavidin-HRP. As expected, when crosslinker was attached to residues 32 or 40, very little crosslinking was seen on the band representing 202/215-399. All the other sites that were labeled clearly formed crosslinks with this fragment, indicating inter-domain crosslinking. All eight sites of attachment crosslinked to the fragment comprising residues 1-201/214 (**Figure 4-37**) which is consistent with the formation of local crosslinks to region 1.1.

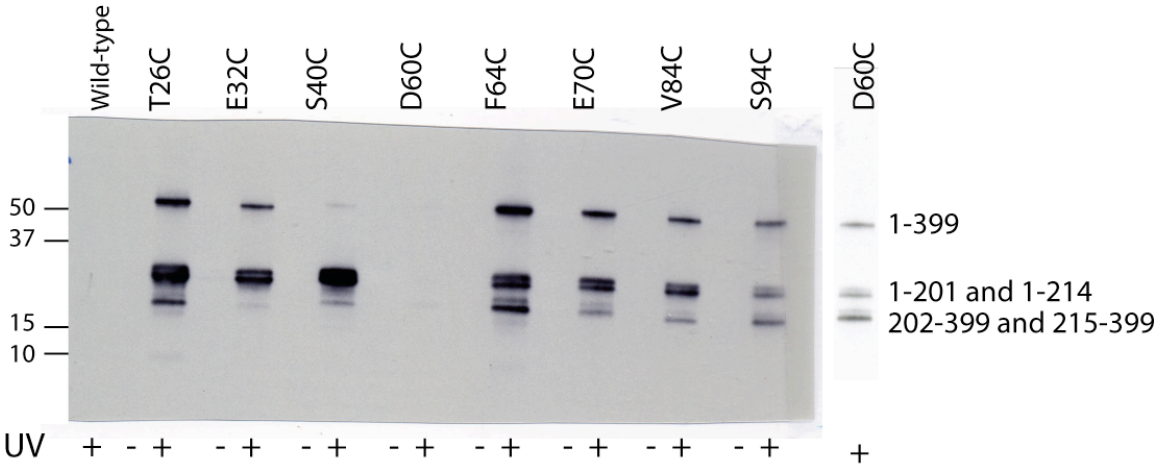


Figure 4-37 : BNPS-Skatole digestion of crosslinked σ-A. Labelled and crosslinked σ-A was digested with BNPS-skatole and analyzed by Western blot (Streptavidin-HRP, SDS-PAGE 4-12% acrylamide). The D60C sample did not label well, so a longer exposure is shown to the right. Strong crosslinking was detected for all samples in the fragments representing residues 1-201/214. Significant crosslinking to the fragments representing residues 202/215-399 was not detected in E32C or S40C but was seen in all other samples. Digestion fragments were identified by excision, digestion with trypsin and analysis by mass spectrometry (data not shown).

To further map the interactions, a subset of the crosslinked samples were subjected to BNPS-skatole digestion and analyzed both by HRP-streptavidin blotting and by coommassie stain. We chose E32C, F64C and E70C as representatives of what appears to be three distinct crosslinking patterns (**Figure 4-33**). The digested bands were cut out and extracted by boiling in ~0.5% SDS in 1N HCl followed by digestion with CNBr. The two stage digestion was analyzed by blotting with HRP-streptavidin (**Figure 4-38**). In the F64C and E70C mutants, the 215-399 fragment was further digested into smaller biotin containing

fragments. In the case of F64C, this includes what is likely the 342-394/9 fragment seen in the digestion of the full fragment (**Figure 4-35**). From this we conclude that region 1.1 is interacting with region 3 or 4. No biotin was detected in the extracted and digested 215-399 fragment from E32C, consistent with no inter-domain crosslinking occurring from this site.

The second digestion of the fragment representing 1-201 was more revealing. While E32C showed no crosslinking to smaller digestion fragments, F64C and E70C both displayed crosslinking to fragments which displayed greater gel mobility than region 1.1 (**Figure 4-38C**). These can only represent digestion fragments containing region 2. Thus we conclude that some sites on region 1.1 are making contacts with region 2.

Figure 4-38 : Mapping crosslinking pattern by sequential digestion. Crosslinked samples of the indicated cysteine mutants were digested with BNPS-skatole, separated by SDS-PAGE (4-12% acrylamide) and visualized by Western blot (HRP-streptavidin, **A**) and by coomasie stain (**B**). All three samples crosslinked to the band representing residues 1-201/215, but only F64C and E70C displayed significant crosslinking to the band representing residues 202/216-399. (**C**-**D**) Coomasie stained, UV treated BNPS-skatole digestion bands from (**B**) were excised and extracted then digested with CNBr and analyzed by SDS-PAGE (10-20% acrylamide, tris-tricine buffer system) and Western blot (HRP-streptavidin). (**C**) CNBr digestion of 1-201/215 fragment. E32 displayed crosslinking only to the parent band of 1-201/215 while F64C and E70C both displayed crosslinking to smaller digestion products. As these run smaller than region 1.1 containing bands, they must represent fragments that contain region 2 (residues ~126-225). (**D**) CNBr digestion of 202/216-399 fragment. E32C displays no crosslinking to either the parent band nor digestion products. F64C and E70C display crosslinking to both the parent band and region 3-4 fragments. Based on comparison to previous experiments (**Figure 4-34, Figure 4-35**), it is most likely that these are region 4 digestion fragments.

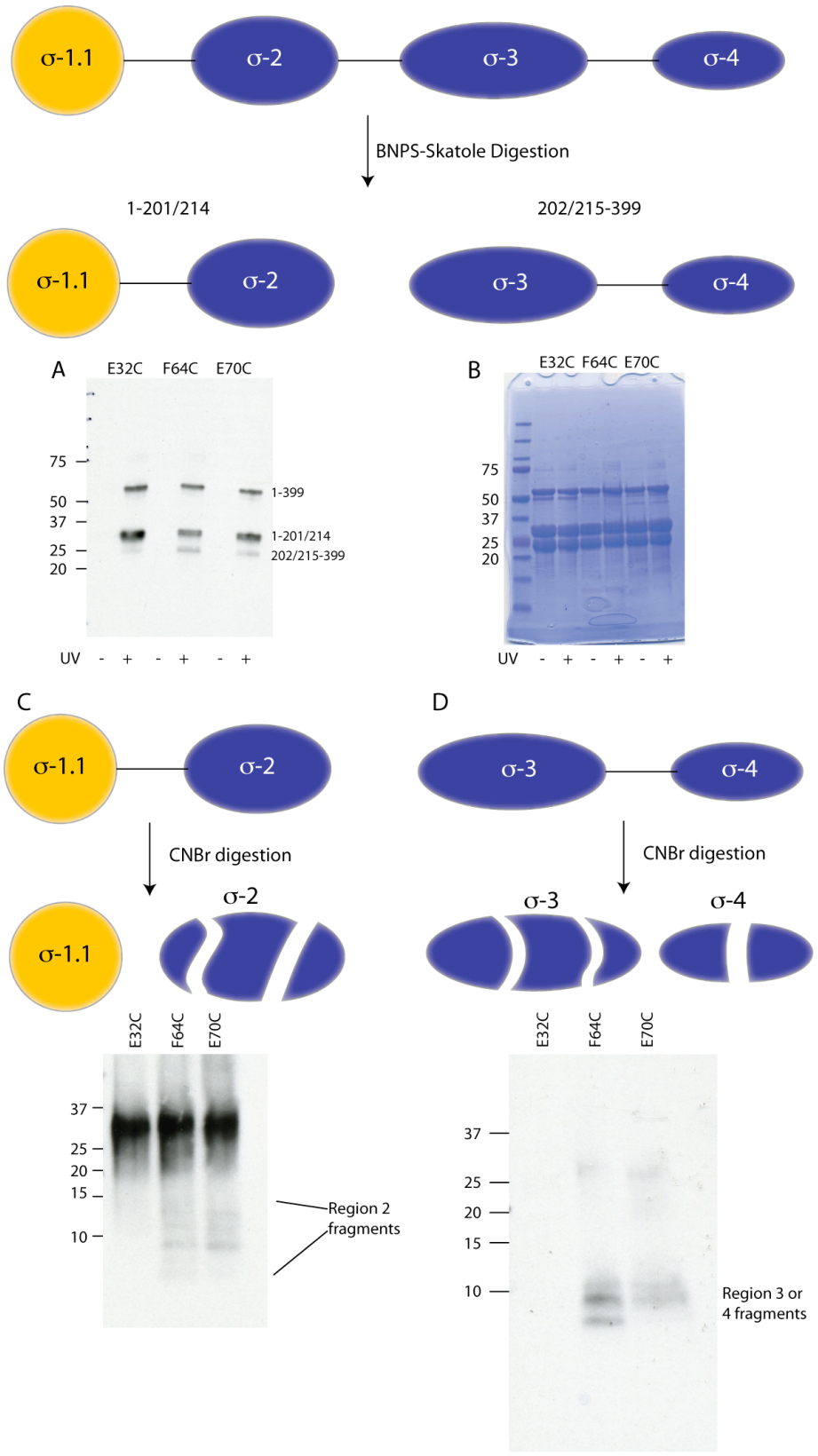


Figure 4-38 : Mapping crosslinking pattern by sequential digestion.

Section 4.3 : A model of group 1 σ factor regulation

Given the inability to determine a crystal structure of region 1.1, one possibility was that region 1.1 is essentially unstructured, acting mainly as an amorphous blob of negative charge. By using NMR, we have determined that this is not the case. In fact, region 1.1 does have a structured core, but this core is in between a flexible N-terminal tail and another flexible linker to region 2. This may provide an explanation for the difficulty in crystallizing σ factors containing region 1.1 as well as the inability to resolve region 1.1 in the crystals that do contain region 1.1 [43].

The residues of region 1.1 that make contacts with region 2 and region 4 turn out to be localized to a negatively charged area (**Figure 4-39**, compare to **Figure 4-14**). The only two residues tested that do not make interdomain contacts are both located on the neutral/slightly positive face of region 1.1. Region 2 and region 4, which are both DNA binding domains are positively charged (region 2 pI ~ 10.5; region 4 pI ~ 10). It is therefore possible that region 1.1 is interacting with region 2 and 4 through a simple electrostatic interaction.

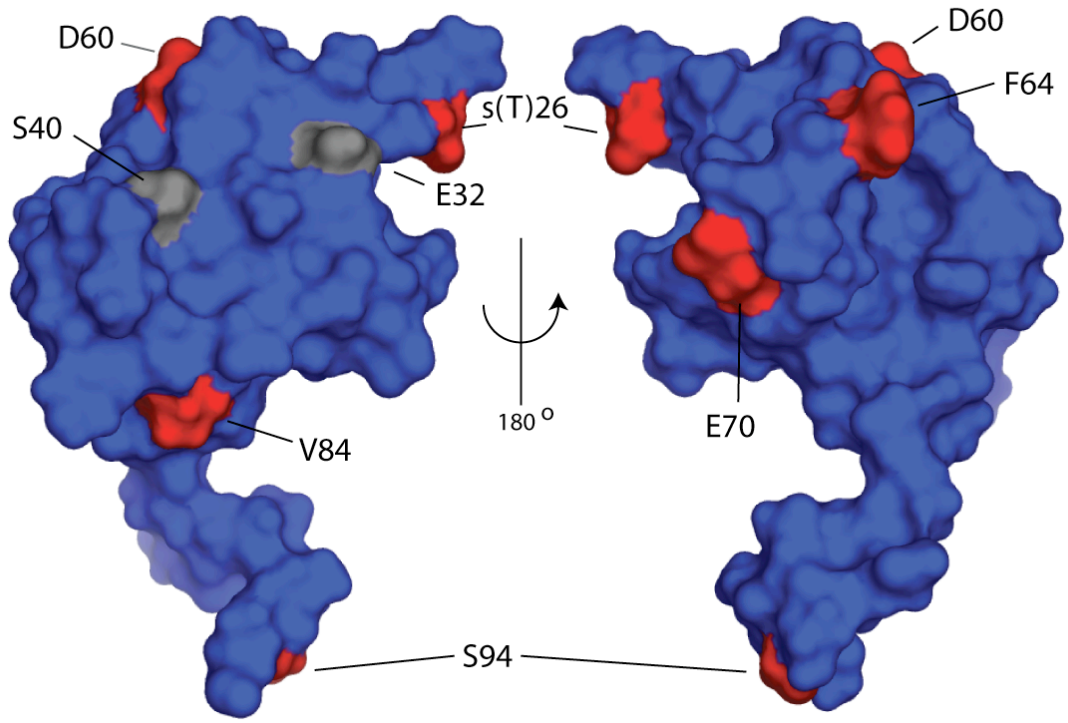


Figure 4-39 : Interacting sites of region 1.1. Surface rendering of region 1.1 shown in the same orientation as **Figure 4-11**. Shown in red are the residues that were able to form crosslinks to other domains of σ-A. Shown in gray are regions that only formed local crosslinks to region 1.1. Images were generated using MacPYMOL (Delano Scientific). The amino acid sequence of the construct is shown at the top of the figure with interdomain crosslinking residues in red and residues that display only local crosslinking in gray.

We could not detect crosslinks that could be definitively mapped to region 3, but we cannot fully rule them out as absence of evidence is not necessarily evidence of absence. However, based on the idea that region 1.1 interacts with regions 2 and 4 through an electrostatic interaction, it is less likely that a region

1.1 – region 3 interaction is occurring. Region 3 bears a negative charge (pI ~ 4.5) which would make it difficult to form contacts with the also negatively charged region 1.1. Although it is possible that region 3 could bind to the slightly positive face of region 1.1, we could not detect any crosslinking to this surface and conclude that no contacts are being made.

This leads us to propose that region 1.1 effects a compaction of the sigma factor in solution by binding to regions 2 and 4, as shown in **Figure 4-40**. Consistent with this model, LRET [134] and FRET [135] data has shown that interdomain distances in the σ factor are smaller in free σ than in the RNAP holoenzyme. The autoregulatory role of region 1.1 is therefore fulfilled by locking the DNA binding domains in a conformation that is incompatible with DNA binding. The contacts between region 1.1 and the DNA binding domains are likely broken and replaced by contacts between the core RNAP and σ2 - σ4, relieving the DNA binding inhibition and allowing transcription initiation. In hindsight, the relative insolubility and instability of the σ(117-399) and σ(117-318) constructs used in the segmental labeling may be the result of the loss of these, presumably stabilizing interactions.

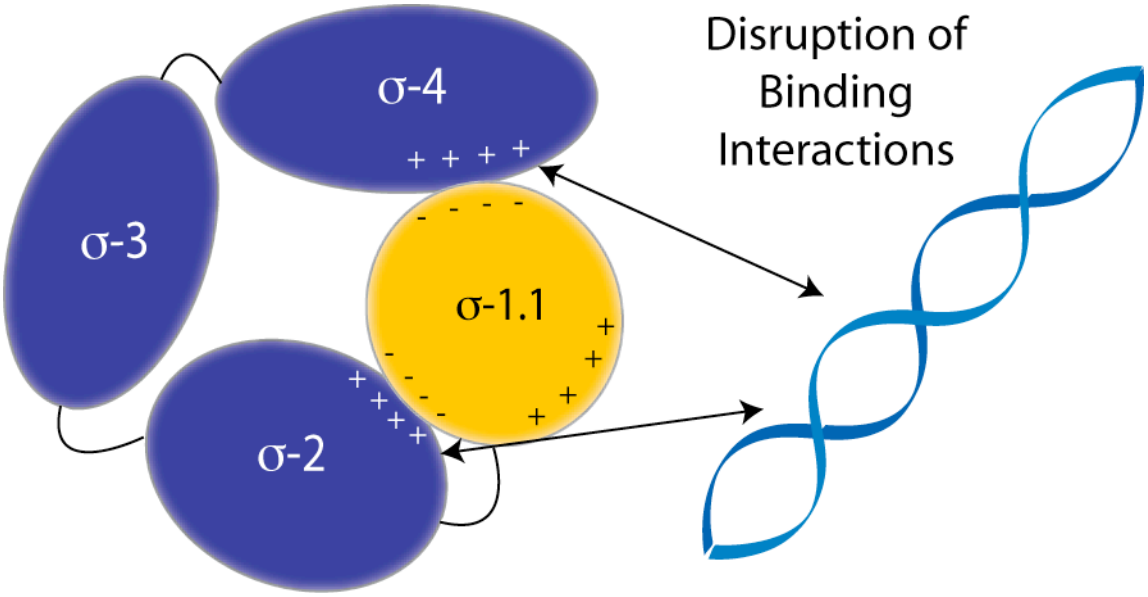


Figure 4-39 : Proposed model of DNA binding inhibition by region 1.1. The negative surface of σ-1.1 is capable of forming crosslinks to the DNA binding domains σ-2 and σ-4. It is thus likely that σ-1.1 organizes the σ factor into a compacted structure that is incapable of binding DNA.

We cannot, using the approaches we have, determine the strength of the interactions between region 1.1 and the DNA binding domains of the σ factor. We predict, however, that they are fairly weak, and likely relatively transient for a number of reasons. First, mechanistically, the interdomain interactions need to be broken in order to form the DNA binding competent RNAP holoenzyme. If the interactions between the sigma domains are of high affinity, it would be difficult and energetically expensive to form productive holoenzyme complexes due to effective molarity [136]. Second, while σ constructs containing region 4 could be inhibited by the addition of region 1.1 in *trans*, achieving a 50% reduction in DNA

binding required a 15-20 fold excess of region 1.1 over region 4 [40], suggesting that binding between the two domains is weak. Additionally, region 2 binding by region 1.1 could not be detected in this *trans* inhibition assay, once again implying weak binding. Finally, the NMR spectra of region 4.2 is not significantly perturbed by region 1.1 [119], despite the fact that our crosslinking data shows that region 1.1 and region 4.2 must be in close proximity. This points to an interaction that is both weak and somewhat fluid, which could be difficult to detect when using NMR based approaches. This highlights the sensitivity of the crosslinking approach as it can capture relatively weak and transient interactions.

Section 4.4 : Open questions and future experiments

The most obvious remaining questions are exactly what residues in σ–2 and σ–4 are region 1.1 contacting and whether or not region 1.1 interacts with σ-3. These questions could both be answered with a direct extension of the crosslinking approach used. When mapping crosslinking interactions from region 1.1, we were aided by the fact that, in a CNBr digest, region 1.1 is readily identified by gel mobility. Thus, if the crosslinker was instead attached to sites on σ-2, σ-3 or σ-4, crosslinking to region 1.1 could be easily identified on an HRP-streptavidin blot without the laborious digestion mapping we were forced to use in this study. Numerous structures of σ-2 - σ-4 are available [41-43, 117, 118, 120-123], which would aid in the design of cysteine mutations in these domains.

Of additional interest is the position of region 1.1 in the RNAP holoenzyme before and after transcription initiation. Ebright and coworkers have demonstrated, through the attachment of fluorescent FRET pairs that region 1.1 is located in the active-center cleft of the RNAP in the closed conformation, possibly functioning as a DNA mimic [135]. When the RNAP holoenzyme transitions to the open complex, region 1.1 is displaced by more than 50 angstroms. Once again, a direct extension of our use of crosslinking could also address this question. By performing the crosslinking experiment in RNAP holoenzyme with labeled σ-1.1 in the presence and absence of DNA, one could determine what subunits region 1.1 is contacting before and after transcription initiation.

Finally, left unanswered is the question of the strength of the interactions detected. As they are predicted to be very weak, accurately measuring them is likely to be difficult. Simply measuring a binding affinity between two domains by mixing them in *trans* would be unlikely to be informative since it has already been determined that the σ-1.1:σ–2 interaction cannot be formed in *trans*. Additionally, it is quite plausible that the σ-1.1:σ–2 and the σ-1.1:σ–4 interactions may function cooperatively. If so, then binary domain:domain affinity measurements would not be meaningful. The most accurate measure of the strength of interaction would probably be to measure and sum the denaturation energies of σ-1.1 and σ–Δ1.1 separately and then compare that to the denaturation energy of full length σ_A.

Any energetic difference found could be attributed to the compaction of the protein effected by σ-1.1.

Chapter 5 : Summary and Conclusions

An ideal conditional protein allele is one that is activated rapidly, and in a dosable manner and whose activity is indistinguishable from wild-type activity. An ideal method of generating conditional protein alleles should also be general in that the generation of a conditional allele of any protein target should require minimal optimization. Most of my thesis work was spent developing two novel methods for the generation of conditional protein alleles.

The first technique is an extension of CPS (**Figure 2-2**) which prior to this work had been used only to tag proteins [74, 83] and to relieve autoinhibition of an enzyme *in vitro* [84]. We demonstrated that an enzyme that was inactivated by splitting it into two fragments could be activated by splicing the inactive fragments together to generate full length protein. This technique proved effective (**Figure 2-8A** and **Figure 2-9**), fast (**Figure 2-8C**), dosable (**Figure 2-10**) and applicable to animal studies (**Figure 2-11**). Unfortunately, CPS was difficult to apply to new protein targets, due to the sequence and context dependence of splicing sites (**Figure 2-14** and **Figure 2-16**). However, the search for a splicing site in a target protein may be streamlined by the use of library screening methods such as those shown in **Figure 2-17** and **Figure 2-18**.

The second technique was developed in response to the difficulty in generating CPS alleles of new protein targets. We reasoned that if we could avoid the splicing reaction, we could avoid the sequence dependence that came

along with it. We turned to the idea of rescuing destabilized proteins, which Crabtree and Wandless have demonstrated is a general means of generating conditional alleles [30-32]. However, the techniques described by Crabtree and Wandless result in the rescue of an FRB or FKBP fusion protein rather than the wild type protein. In some cases, this may not be an issue, but for others, the fusion protein may not display wild type activity. As such we developed the SURF technology (**Figure 3-1**) which activates protein activity by removing a destabilizing sequence rather than by stabilizing it.

In contrast to the CPS system, it was relatively easy to generate SURF alleles of a variety of proteins in mammalian cells (**Figure 3-2**, **Figure 3-5, Figure 3-6, Figure 3-8**). In S2 cells, we were able to generate a SURF allele of luciferase (**Figure 3-12**) but were unable to generate SURF alleles of DBT or SGG (**Figure 3-13** and **Figure 3-14**). Although this does mean that SURF is not a completely general technology, we are inclined to believe that, given the domain structure of DBT and SGG, they represent uncommon exceptions to the generality of SURF. The only drawback to SURF alleles is that their activation is slow when compared to CPS since SURF alleles require new protein synthesis for activation.

In addition to my work on conditional protein alleles, I have also worked on the characterization of region 1.1 of bacterial σ-70 like factors. The function of this domain has long been mysterious, in large part due to the lack of a high resolution structure. We generated a high resolution structure based on NMR

spectroscopy and provided evidence by segmental NMR and interdomain crosslinking that inhibition of DNA binding is achieved by direct binding of region 1.1 to the DNA binding domains σ-2 and σ-4.

Chapter 6 : Materials and Methods

Section 6.1 : CPS

Plasmid Construction

Constructs used in chapter 2 are shown **Table 6-1**. Constructs were cloned by standard methods and confirmed by sequencing. The pCre-luc plasmid (Promega) was used as a source of firefly luciferase. Constructs **1** and **2** from reference [83] were used as sources of the split VMA intein, FKBP/FRB domains and MBP. Renilla luciferase transfection control was cloned from pRL-SV40 (Promega). N-luc and C-luc intein mutants were generated by mutating the first (C1) and last (N454) residues of the VMA intein to alanine using Quikchange (Stratagene) mutagenesis. Oligonucleotides were purchased from Integrated DNA Technologies. Sequencing was performed by GENEWIZ. Constructs were cloned into one or more of the following vectors : pcDNA3+ (Invitrogen), pUAST [137], or pAc5.1/V5-His (Invitrogen).

Table 6-1 : Constructs used in Chapter 2.

Construct	Sequence	Vector
Luc(1-436)	Flag tag – GGRVGK - Luciferase(1-436) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(437-551)	Flag tag - GGRSH - FRB - VMA (390-454) – Luciferase(437-551 R437C)	pAc5
MBP-I_N-FKBP	MBP - SSSNNNNNNNNNNLGIEGRISEFLKG - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
N-luc	Flag tag – GGRVGK - Luciferase(1-490 H489K) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5 (S2 cells) pUAS (flies) pcDNA (mammalian cells)

C-luc	Flag tag - GGRSH - FRB - VMA (390-454) – Luciferase(491-551 K491C)	pAc5 (cells) pUAS (flies) pcDNA (mammalian cells)
C-lucFRB_2	Flag tag - GGRSH - FRB - FRB - VMA (390-454) – Luciferase(491-551 K491C)	pAc5 (cells) pUAS (flies) pcDNA (mammalian cells)
C-lucFRB_3	Flag tag - GGRSH - FRB - FRB - FRB - VMA (390-454) - Luciferase(491-551 K491C)	pAc5
Luc(1-507)	Flag tag – GGRVGK - Luciferase(1-507 T507G) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(508-551)	Flag tag - GGRSH - FRB - VMA (390-454) – Luciferase(508-551 T508C)	pAc5
Luc(1-215)	Flag tag – GGRVGK - Luciferase(1-215) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(216-551)	Flag tag – GGRVGK - FRB - VMA (390-454) – Luciferase(216-551)	pAc5
Luc(1-184)	Flag tag – GGRVGK - Luciferase(1-184 E184G) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(185-551)	Flag tag – GGRVGK - FRB - VMA (390-454) – Luciferase(185-551 S185C)	pAc5
Luc(1-80)	Flag tag – GGRVGK - Luciferase(1-80 A79G) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(81-551)	Flag tag – GGRVGK - FRB - VMA (390-454) – Luciferase(81-551)	pAc5
Luc(1-297)	Flag tag – GGRVGK - Luciferase(1-297) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Luc(298-551)	Flag tag – GGRVGK - FRB - VMA (390-454) – Luciferase(298-551 S298C)	pAc5
Dbt(1-266)	MBP - SSSNNNNNNNNNNLGIEGRISEFLKG - Dbt(1-266) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(267-441)	MBP - SSSNNNNNNNNNNLGIEGRISEFLKG	pAc5

	- FRB - VMA (390-454) – Dbt(267-441)	
Dbt(1-95)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - Dbt(1-95) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(96-441)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - FRB - VMA (390-454) – Dbt(96-441)	pAc5
Dbt(1-281)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - Dbt(1-281) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(282-441)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - FRB - VMA (390-454) – Dbt(282-441 F282C)	pAc5
Dbt(1-310)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - Dbt(1-310) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(311-441)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - FRB - VMA (390-454) – Dbt(311-441 A311C)	pAc5
Dbt(1-351)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - Dbt(1-351) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(352-441)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - FRB - VMA (390-454) – Dbt(352-441 S352C)	pAc5
Dbt(1-366)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - Dbt(1-366) - VMA (1-182) - LYSRNNGNGNGTR - FKBP	pAc5
Dbt(367-441)	MBP - SSSNNNNNNNNNLGIEGRISEFLKG - FRB - VMA (390-454) – Dbt(367-441 S367C)	pAc5

S2 cell culture experiments

Drosophila Schneider's Line 2 (S2) cells were maintained in M3 media (Sigma) supplemented with Fetal Bovine Serum (Sigma, 15%) and Penicillin/Streptomycin (Invitrogen, 93U/mL & 93μg/mL) at 25 °C. Cells were transfected with 500 ng of each plasmid (500 ng of empty vector was added in the case of single transfections) and 100 ng of renilla luciferase control when appropriate using the Qiagen Effectene protocol. Following 16h of expression at 25 °C, cells were aliquoted into 1.5 mL microcentrifuge tubes and treated with DMSO (Sigma, 0.1% final concentration) or with rapamycin (Sigma, 100 μM in DMSO, 100 nM/0.1% DMSO final concentration) for 4 h unless otherwise stated. Cells were then pelleted by centrifugation, media was removed and cells were lysed with 1x Passive Lysis Buffer (PLB, Promega) supplemented with 1 mM $ZnCl_2$ to prevent splicing after lysis [83]. Luciferase readings were performed in a Berthold Sirius luminometer using the Promega Dual-luciferase assay kit. In the ascomycin competition experiment, cells were initially treated with rapamycin (5 μM in DMSO, 5 nM/0.1% DMSO final concentration) followed by ascomycin (Sigma) at the indicated time points without changing media (500 μM in DMSO, final concentrations: 5 nM rapamycin/500 nM ascomycin/0.2% DMSO).

Mammalian cell culture experiments

HeLa cells were cultured in Dulbeco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% FBS (Sigma) at 37 °C and 5% CO_2 according to standard procedures. Cells were transfected using FuGENE-6 (Roche Applied

Science). Plasmid DNA (250 ng of N-luc and C-luc constructs and 25 ng of the renilla control pRL-SV40 (Promega) per 1 mL, twelve-well plate) was mixed in a 1:3 ratio with the FuGENE reagent in DMEM and was then applied to cells growing in DMEM/10% FBS. Transfection was allowed to proceed 16 h before replacement with fresh media containing DMSO (Sigma, 0.1% final concentration) or with rapamycin (Sigma, 100 μM in DMSO, 100 nM/0.1% DMSO final concentration) for 4 h. After the indicated drug treatments, cells were washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 10mM Na_2HPO_4, 2 mM KH_2PO_4 (pH 7.4)) and lysed with 1x Passive Lysis Buffer (PLB, Promega) for 15 min at rt. Following centrifugation for 10 min at 10 *g*, luciferase readings were performed in a Berthold Sirius luminometer using the Promega dual-luciferase assay kit.

Western Blotting

Protein samples in SDS loading buffer (50 mM Tris·HCl, 2% (w/v) SDS , 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, 100 mM DTT) were separated by SDS-PAGE before being transferred to PVDF (Biorad) using standard western blotting procedures. The anti-luciferase antibody (Cambridge Antibodies) was used at 1:15000. The anti-MBP antibody (New England Biolabs) was used at 1:5000. Loading was normalized to transfection efficiency by loading an equal amount of Renilla luciferase activity. Western blots were blocked in TBST (10 mM Tris (pH 8.0), 150 mM NaCl, 0.1% Tween 20) containing 5% dry milk for 1 h,

followed by incubation with the appropriate primary antibody in blocking buffer for an additional 1 h at rt or overnight at 4 °C. After washing three times with TBST, the blots were incubated with HRP-conjugated secondary antibodies (Amersham Bioscience for anti-rabbit, Cambridge Antibodies for anti-goat) for 1 h in blocking buffer, washed three more times with TBST, and developed using ECL reagents (Perkin Elmer) and Kodak Biomax MR film.

Transgenic Flies

N-luc and C-luc were cloned into the pUAST [137] vector. 20μg of each was co-precipitated with 5μg of the wings-clipped plasmid [138] and resuspended in sterile water. Dechorionated yw embryos were injected with the plasmid mixture and covered with Halocarbon brand 27 oil at 18 °C. Upon emerging, larvae were placed on fresh fly food and grown to adulthood at room temperature. Adult flies were backcrossed to yw flies and w+ offspring were selected and balanced using standard methods. C-lucFRB2 flies were generated and balanced using standard methods (BestGene, Inc.).

N-luc insertions into the 2nd and 3rd chromosome were crossed with flies containing P-element GAL4 insertions under the armadillo and timeless promoters respectively. Third chromosome crossover products containing N-luc and armadillo-GAL4 and second chromosome crossover products containing N-luc and timeless-GAL4 were selected by screening for luciferase activity following breeding with C-luc insertions into the second and third chromosomes respectively.

Male flies (female flies had similar readings, data not shown) homozygous for N-luc, C-luc and a GAL4 driver as well as yw flies and flies expressing full length luciferase under the timeless promoter were deprived of food for 1 hour and then placed on fly food supplemented with 100 μM rapamycin (2% DMSO) and 200 μM luciferin (Sigma) and placed in a Hamamatsu Photon detection unit maintained at room temperature. Luciferase readings were taken in 1 minute bins every 12 minutes for up to 4 days.

For western blotting, flies were deprived of food for 1 hour and then placed on 3MM filter paper soaked in 40% glucose supplemented with 100μM rapamycin (2% DMSO) for 12h. Individual flies were homogenized in PLB with 1 mM $ZnCl_2$. Protein concentration in fly lysates was determined by the Bradford assay (Biorad). 5 μg of total protein was loaded from each fly lysate. A non-specific band served as a control for transfer efficiency.

Section 6.2 : SURF

Plasmid construction

SURFn and SURFc constructs were prepared using standard molecular cloning techniques. Constructs **1** and **2** from reference [74] were the sources of the FRB and FKBP domains, ubiquitin cDNA (Sigma) was the source of ubiquitin, pCre-luc plasmid (Promega) was the source of luciferase, pcDNA3-Casp3-myc (Addgene) was the source of caspase-3, pHA-Smad3 was the source of Smad3E, *Salmonella Ty*. genomic DNA was the source of SopE, and pLNCX

chick src (Addgene) was the source of v-Src. The final architectures of all constructs are shown in **Table 6-2**. FRB refers to the single mutant (T2098L) used in chapter 2, FRB* refers to the triple mutant (K2095P, T2098L, W2101F). All N-terminal ubiquitin constructs contain an I13A mutation. In our hands, split ubiquitin of wild-type sequence displayed constitutive complementation in mammalian cells, whereas the I13G mutant resulted in no inducible cleavage (data not shown). All constructs were characterized by DNA sequencing (Genewiz). Stabilization mutants were generated by mutating the residue immediately following ubiquitin in the C-terminal constructs to proline using Quikchange (Stratagene) mutagenesis.

Table 6-2 : Constructs used in Chapter 3

Construct	Sequence	Vector
MBP-UbN-FKBP (**1**)	MBP-EFEGGST-Ubiquitin (I13A, 1-37)-GGSTMAAA-FKBP	pcDNA (mammalian cells) pUAS, pAc5 (S2 cells)
SopE-FRB-UbC-Luc (**2**)	SopE (1-100)-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase	pcDNA (mammalian cells) pUAS, pAc5 (S2 cells)
FRB_3*-UbC-Luc (**3**)	FRB*-ASLEGSTCT-FRB*-ASLEGSTCT-FRB*-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase-HA tag	pcDNA
FRB_3-UbC-Luc (**4**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Luciferase-HA tag	pcDNA
FRB_3-UbC-Casp3 (**5a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-Caspase-3-HA tag	pcDNA
FRB_3-UbC-Casp3	FRB-ASLEGSTCT-FRB-	pcDNA

(**5b**)	ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-Caspase-3-HA tag	
FRB_3-UbC-vSrc (**6a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-vSrc-HA tag	pcDNA
FRB_3-UbC-vSrc (**6b**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-vSrc-HA tag	pcDNA
FRB_3-UbC-Smad3 (**7a**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-MHRSGIMAAA-HA tag-Smad3	pcDNA
FRB_3-UbC-Smad3 (**7b**)	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-HA tag-Smad3	pcDNA
SURF-DBT	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-HA tag-DBT	pUAS pAc5
SURF-SGG	FRB-ASLEGSTCT-FRB-ASLEGSTCT-FRB-GSLEGSTMSG-Ubiquitin (35-76)-PHRSGIMAAA-HA tag-DBT	pUAS

Mammalian Cell Culture Experiments

HeLa and SYF cell lines were cultured in Dulbeco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% FBS (Sigma) at 37 °C and 5% CO_2 according to standard procedures. Cells were transfected using FuGENE-6 (Roche Applied Science). Plasmid DNA (250 ng of SURFn and SURFc constructs and 25 ng of reporter construct and/or 25 ng of the renilla control pRL-SV40 (Promega) per 1 mL, twelve-well plate) was mixed in a 1:3 ratio with the FuGENE reagent in DMEM and was then applied to cells growing in DMEM/10%

FBS. In the time course experiments, transfection was allowed to proceed for 24 h before replacement with fresh media containing DMSO (Sigma, 0.1% final concentration) or with rapamycin (Sigma, 100 μM in DMSO, 100 nM/0.1% DMSO final concentration) for the indicated time points. For all other experiments the above transfection mixture was applied to cells growing in DMEM/10% FBS containing DMSO (0.1% final concentration), rapamycin (Sigma) and/or ascomycin (Sigma) in DMSO (0.1% - 0.2% final concentration) where appropriate. After 12 h, fresh medium for continued drug treatment was added. When indicated, Cells were treated with the proteasome inhibitor ZL_3VS [95] for 6 h.

Luciferase Measurements

After the indicated drug treatments, cells were washed twice with PBS and lysed with 1x Passive Lysis Buffer (PLB, Promega) for 15 min at rt. Following centrifugation for 10 min at 10 *g*, luciferase readings were performed in a Berthold Sirius luminometer using the Promega dual-luciferase assay kit.

MTT cell proliferation assay

The MTT cell proliferation assay was purchased from American Tissue Culture Collection (ATCC) and performed as per the manufacturer's instructions. Briefly, SYF cells were tranfected with appropriate plasmids as above. After 24 h to allow for protein expression, the cells were trypsinized, counted, and ~1 X 10^4 were plated in 96 well plates in DMEM containing 10% FBS and rapamycin (100

nM) or DMSO where appropriate. After an additional 24 h to allow for rescue, the MTT cell proliferation assay was performed.

S2 cell culture experiments

Cell line maintenance, transfection, drug treatment and harvesting was performed as described in Section 6.1.

Western Blotting

Mammalian cells were washed twice with PBS and lysates prepared using NP buffer (50 mM Tris [pH 7.4], 0.5% NP40, 5 mM $MgCl_2$, protease inhibitor cocktail [Roche Applied Science]). Total protein concentrations were normalized by Bradford assay (BioRad) and NP lysates were then mixed 3:1 with 4X SDS loading buffer and separated by SDS-PAGE before being transferred to PVDF (Biorad) using standard western blotting procedures. S2 cell westerns were performed as described in Section 6.1. The anti-HA antibody (Roche Applied Science) was used at 1:10000. The anti-MBP antibody (New England Biolabs) was used at 1:5000. Anti-avian Src (clone EC10, Upstate) was used at 1:1000. The anti-phosphotyrosine (Calbiochem) was used at 1:1000. Anti-phosphotryrosine western blots were blocked in TBST containing 5% BSA for 1 h. All other western blots were blocked in TBST containing 5% dry milk for 1 h, followed by incubation with the appropriate primary antibody in blocking buffer for an additional 1 h at rt or overnight at 4 °C. After washing three times with TBST,

the blots were incubated with HRP-conjugated secondary antibodies (Amersham Bioscience) for 1 h in blocking buffer, washed three more times with TBST, and developed using ECL reagents (Perkin Elmer) and Kodak Biomax MR film.

Section 6.3 : σ factor

NMR samples

T. maritima σ-A region 1.1 (residues 1-96 or 1-116, with mutations Q21L, E22V, Q23P, K24R, E25G, T26S, L27H, P28M to install a thrombin site and mutations S96G or S116G to facilitate intein thiolysis and expressed protein ligation) was cloned into the pTXB1 vector (New England Biolabs) to generate a fusion construct of σ-A region 1.1 – *Mxe* GyrA intein – chitin binding domain. BL-21(DE3) (Invitrogen) cells were transformed with this construct, grown overnight on LB plates (100 μg/mL ampicillin) then grown in 1L M9 minimal media (100 μg/mL ampicillin) with 1.2g $^{15}NH_4SO_4$ and 2g ^{13}C glucose as the only nitrogen and carbon sources. The culture was grown at 37 °C to an OD_{600} of ~0.6. Expression was induced with 0.8 mM IPTG and was carried out at 30 °C for 4.5 h. Bacteria were harvested by centrifugation, resuspended in lysis buffer (500 mM NaCl, 20 mM Tris pH 8.0, 1 mM EDTA) and lysed using a french press. The lysate was centrifuged and the supernatant applied to a chitin bead column (New England Biolabs) and washed extensively with lysis buffer. Intein cleavage was induced with 50 mM DTT in lysis buffer and allowed to proceed overnight at 4 °C with gentle rocking. Cleaved protein was eluted with lysis buffer and was exchanged

into 1x PBS by dialysis. Thrombin protease (Amersham) was added (50 U) and proteolysis was carried out overnight at room temperature with gentle rocking. The resulting protein was H_2N-GSHM-(σ-A 29-95/115)-G-COOH. The sample was further purified by ion exchange chromatography (Hiprep Q16/10 column, AKTA FPLC, Amersham) using a 0.1-1M NaCl gradient in 20 mM Tris·HCl (pH 8.0). Protein samples were characterized by electrospray mass spectrometry (ESMS, Sciex API-100 single quadropole electrospray mass spectrometer, see **Figure 4-8** and **Figure 4-22**). The protein sample was treated at 65 °C for 30 min to eliminate *E. coli* proteases and other contaminants and then concentrated in NMR buffer (10 mM NaPi pH 6.5, 100 mM NaCl, 1mM EDTA) to a final concentration of 300-800 μM depending on the sample (sample shown in **Figure 4-9** is 370 μM, sample shown in **Figure 4-23** is 300 μM). Protein concentrations were determined using Bradford assay (BioRad) using a BSA (New England Biolabs) standard curve.

NMR

All NMR experiments were recorded at 298 K using 600, 700 and 800 MHz Bruker Avance NMR spectrometers equipped with cryoprobes. All NMR data were processed using Topspin 1.1 (Bruker) and analyzed by NMRView [139]. ^{1}H chemical shifts were referenced to water at 4.75 ppm (at 25 °C) whereas the $\gamma^{13}C/\gamma^1H$ and $\gamma^{15}N/\gamma^1H$ ratios were used for indirect referencing of the ^{13}C and ^{15}N chemical shifts. ^{1}H-^{15}N-HSQC, ^{1}H-^{13}C-HSQC, 3D-^{15}N-NOESY-HSQC,

HNCO, HNCACB, CBCA(CO)NH experiments were used to obtain the sequential assignment of the backbone [140]. Side-chain carbon and proton assignments were obtained using C(CO)NH, HC(CO)NH, HBHA(CO)NH and HCCH-TOCSY experiments. For assigning the aromatic resonances, ^{1}H-^{13}C HSQC, (Hβ)Cβ(CγCδ)Hδ and (Hβ)Cβ(CγCδCε)Hε [141], and 3D ^{13}C NOESY-HSQC data was utilized.

Structure Calculations

The backbone dihedral angles (ϕ and ψ) were calculated by analyzing the $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and ^{15}N chemical shifts with the TALOS program that predicts the backbone torsion angles from the amino acid sequence and chemical shift information [142]. Hydrogen bond restraints were inferred from slow H/D exchange of backbone amides. Distance restraints were derived from 3D ^{13}C-edited NOESY and 3D ^{15}N-edited NOESY experiments. Using NMRView, the NOESY cross peak volumes/intensities were obtained and converted into distance restraints using the symmetry ambiguous distance restraints (ADR) protocol within the ARIA program [143, 144]. Structure calculations were performed using the Cartesian dynamics simulated anneling protocol within ARIA/CNS. Lowest energy structures were further refined in water using the protocol described by Linge *et al* [145]. The NMR structures were analyzed using PROCHECK [146]. The electrostatic potentials were calculated using GRASP [133]. Structures were displayed using MACPymol (Delano Scientific).

Expressed Protein Ligation

Region 1.1 (residues 1-116, with mutations described above) was expressed in IL M9 minimal media (100 μg/mL ampicillin) in $^{2}H_2O$ with 1.2 g $^{15}NH_4SO_4$ and 2 g ^{13}C glucose as the only nitrogen and carbon sources, respectively. Region 1.1 (residues 1-96, with mutations described above) was expressed in LB media (100 μg/mL ampicillin) using the construct described in the NMR sample section above. Protein expression was induced and samples were immobilized on chitin beads as described in the NMR samples section. Thrombin protease (Amersham) was added (50U) and proteolysis was carried out on column overnight at room temperature with gentle rocking. Intein cleavage was induced with 3% (vol/vol) EtSH in lysis buffer and allowed to proceed overnight at room temperature. Region 1.1 thioester was eluted with 3% (vol/vol) EtSH in lysis buffer and concentrated to 100-250 μM (concentrations determined using Bradford assay).

C-terminal fragments (residues 117-399, S117C; 117-318, S117C and 97-399, S97C) were cloned into the pET15B vector (Novagen). Final sequences were MGSS - his tag – SSGLVPRGSHM – Factor Xa site (IEGR) followed by the indicated σ-A residues. BL-21(DE3) cells were transformed with these constructs and grown in LB broth (100 μg/mL ampicillin) at 37 °C to an OD_{600} of ~0.6. Protein expression was induced with 0.8 mM IPTG at 37 °C for 4-6 h. Bacteria

were harvested by centrifugation, treated at 65 °C for 20 minutes lysed using a french press and purified using the Ni-NTA system (Novagen). A buffer exchange to 1x PBS was performed using centrifugal filter devices (Amicon, 5,000 MWCO). N-terminal cysteine was exposed by treatment with Factor Xa protease (Amersham) in 1x PBS supplemented with 1 mM DTT overnight at room temperature. Factor Xa digestions were purified by ion exchange chromatography (Hiprep SP16/10 column, AKTA FPLC, Amersham) using a 0.1-1M NaCl gradient in 20 mM Tris·HCl (pH 6.5). Samples were buffer exchanged into 10 mM CHAPSO (Sigma) in 1x PBS using a centrifugal filter device (Amicon, 10,000 MWCO) and concentrated to ~20-100 μM. The presence of CHAPSO in the samples seemed to result in unreliable readings using the Bradford assay, so protein concentrations can only be considered approximate. When possible, protein concentration was also estimated by examination of coomassie stained acrylamide gels.

EPL reactions were performed in 1x PBS with 100 mM MESNa, 1 mM EDTA, 3% (vol/vol) EtSH at room temperature with gentle rocking. It was empirically determined that a 1:1 molar ratio between the thioester containing fragment and the N-terminal cysteine containing fragment resulted in the greatest yield. The two fragments were present at the highest concentration possible, generally 20-25 μM. In the reaction shown in **Figure 4-19** the concentrations of the fragments is ~20 μM. In the reaction shown in **Figure 4-24** the concentrations of the fragments is ~25 μM. Reaction mixtures were purified by

ion exchange chromatography (Hiprep SP16/10 column, AKTA FPLC, Amersham) using a 0.1-1M NaCl gradient in 20 mM Tris·HCl (pH 6.5) and concentrated in NMR buffer containing 5 mM CHAPSO.

SDS-PAGE analysis

SDS-PAGE analysis of constructs used for all NMR studies was performed on 12 or 15% acrylamide (1% SDS), Tris-glycine buffered gels. SDS-PAGE analysis in the crosslinking section was performed using the Criterion system (Biorad) except in the case of the experiment shown in **Figure 4-35**. Criterion gels used were 4-12% acrylamide Bis-Tris buffered gels and 10-20% acrylamide Tris-tricine buffered gels, as indicated. The experiment shown in **Figure 4-35** was performed on a 20% acrylamide (1% SDS) Tris-glycine buffered gel. Samples for western blotting were separated by SDS-PAGE before being transferred to PVDF (Biorad) using standard western blotting procedures. The membrane was blocked in 5% milk in TBST for 1 hour at room temperature then washed 3 times with TBST and incubated with Streptavidin-HRP (GE Health Sciences, 1:5000 in TBST) for 1 hour. The membrane was washed 3 more times with TBST and developed using ECL reagents (Perkin Elmer) and Kodak Biomax MR film.

Abortive Initiation Reactions

Reactions were performed in 10 μL of transcription buffer (20 mM Tris-HCl (pH 8.0); 50 nM NaCl; 5 mM $MgCl_2$). *E. coli* core RNAP (0.1 μM) was incubated with σ factor (0.5 μM) for 10 min. at 37 °C. T7A1 promoter fragment (0.1 μM) was then added and the mixture was incubated at 37 °C for a further 5 min. Reactions were initated by the addition of 0.5 mM CpA initiating dinucleotide and 1 μL of α-^{32}P-UTP. The reactions proceeded for 15 min. at 37 °C then were terminated by the addition of an equal volume of gel loading buffer (7 M urea; 90 mM Tris; 64.6 mM boric acid; 2.5 mM EDTA (pH 8.3)). Reaction products were separated by polyacrylamide gel electrophoresis on a 25% gel in 7 M urea and visualized by autoradiography.

Crosslinker Synthesis

All chemical reagents were purchased from Sigma or Aldrich and used without further purification. Boc-Cys(NPys)-OH was purchased from Bachem and 4-azido-2,3,5,6-tetra-fluoro-benzoic acid and PEO-biotin were purchased from Pierce. Thin-layer chromatography was carred out using EMD silica gel plates. Flash chromatography was performed using Sigma-Aldrich 60 Å 230-400 mesh silica gel. All solvents were concentrated using rotary evaporation. All ^{1}H and ^{13}C NMR spectra were acquired on Bruker AM-400 MHz spectrometer. All ^{1}H chemical shifts are reported in δ referenced to solvent. Coupling constants (*J*)

are reported in Hz. Electrospray (ESI±) spectra were obtained on a Waters Micromass ZQ LC/MS.

Compound 4.1 To Boc-Cys(NPys)-OH (500 mg, 1.33 mmol) dissolved in DMF (5 mL) was added pyridine (120 μL, 1.46 mmol) followed by pentafluorophenyl trifluoroacetate (260 μL, 1.53 mmol). After 45 min the reaction was diluted with ethyl acetate (50 mL) and washed with 1N HCl (50 mL, 2 times), conc. $NaHCO_3$ (50 mL, 2 times), H_20 (50 mL), and brine (50 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated. Silica gel chromatography (20% ethyl acetate in hexanes) yielded 610 mg (85%) of **4.**1 as a pink solid that was used in the next reaction with no further characterization.

Compound 4.2 To a solution of PEO-biotin (50 mg, 0.12 mmol) in DMF (2 mL) was added diisopropylethylamine (25 μL, 0.14 mmol). After stirring for 15 min a solution of **4.1** (50 mg, 0.12 mmol) in DMF (2 mL) was added and the reaction allowed to stir for 16 h. At this time the reaction was concentrated and purified by silica gel chromatography (14:2:1 ethyl acetate:MeOH:H_2O) to yield 77 mg (83%) of **4.2** as a tan foam. ESI-MS calculated for $C_{31}H_{49}N_7O_{10}S_3$ $[M + H]^+$ 776.27, found 776.90.

Compound 4.3 Compound **4.2** (60 mg, 0.08 mmol) was dissolved in 95% trifluoroacetic acid in H_2O (5 mL) and stirred for 1 h. At this time the reaction was

concentrated and azeotroped with toluene (10 mL, 3 times) to yield 50 mg (94%) of **4.3** as a waxy solid that was used directly in the next step.

Compound 4.4 To a solution of 4-azido-2,3,5,6-tetrafluorobenzoic acid (50 mg, 0.21 mmol) in DMF (5 mL) was added pyridine (19 μL, 0.24 mmol) and pentafluorophenyl trifluoroacetate (42 μL, 0.25 mmol). After 45 min the reaction was diluted with ethyl acetate (10 mL) and washed with 1N HCl (10 mL, 2 times), conc. $NaHCO_3$ (10 mL, 2 times), H_20 (10 mL), and brine (10 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated to yield 67 mg (80%) of **4.4** that was used directly in the next step.

Compound 4.5 Diisopropylethylamine (10 μL, 0.06 mmol) was added to a solution of **4.3** (35 mg, 0.05 mmol) in DMF (1 mL). After 15 min a solution of **4.4** (24 mg, 0.06 mmol) in DMF (1 mL) was added and the reaction stirred for 16 h. The mixture was then concentrated and purified by silica gel chromatography (14:2:1 ethyl acetate:MeOH:H_20) to yield 32 mg (74%) of **4.5** as an off-white foam. ^{1}H NMR (400 MHz, CD_3OD) δ 9.00 (d, 1 H, *J* = 3.3 Hz), 8.73 (d, 1 H, *J* = 7.0 Hz), 7.64 (dd, 1 H, *J* = 4.6, 8.1 Hz), 5.02 (t, 1 H, *J* = 7.6 Hz), 4.60 (dd, 1 H, *J* = 4.9, 7.6 Hz), 4.42 (dd, 1 H, *J* = 4.4, 7.7 Hz), 3.65 (app t, 6 H, *J* = 5.4 Hz), 3.48 (app t, 6 H, *J* = 5.9 Hz), 3.32-3.27 (m, 4 H), 3.03 (dd, 1 H, *J* = 4.9, 12.7 Hz), 2.80 (app d, 1 H, *J* = 12.7 Hz), 2.27-2.22 (m, 4 H), 1.85-1.66 (m, 5 H), 1.57-1.52 (m, 2 H); ESI-MS calculated for $C_{33}H_{40}F_4N_{10}O_9S_3$ $[M + H]^+$ 893.92, found 893.88.

Crosslinking samples

T. maritima σ-A was cloned into pET15B (Novagen) with an N-terminal hexahistidine tag and a thrombin recognition site for removal of the affinity tag (Final sequence : MGSSHHHHHHSSGLVPRGSH - σ-A). Cysteine mutants (**Table 4-3**) were generated using the Quikchange II XL kit (Stratagene). BL-21(DE3) cells were transformed with wild-type or mutant Sigma A and grown in LB broth (100 μg/mL ampicillin) at 37 °C to an OD_{600} of ~0.6. Protein expression was induced with 0.8 mM IPTG at 37 °C for 4-6 h. Bacteria were harvested by centrifugation, treated at 65 °C for 20 minutes lysed using a french press and purified using the Ni-NTA system (Novagen). Protein samples were then transferred to 1x PBS by dialysis and treated with thrombin protease (Amersham, 50 U) overnight at room temperature. Proteolyzed samples were further purified by ion exchange chromatography (HiTrap SP HP, Amersham) with a 0.1-1M NaCl gradient in 20 mM Tris·HCl (pH 8.0). Fractions containing protein were concentrated and stored at -80 °C in 10% glycerol.

Attachment of crosslinker and crosslinking

The crosslinker **4.5** was attached to the purified cysteine mutant of σ-A by disulfide exchange. Reaction mixtures contained 6 μM protein and 100 μM crosslinker in 50 mM sodium borate (pH 8.3), 8 mM Tris-HCl (pH 8.3), 80 mM KCl, 0.4 mM EDTA, 5 mM methionine, 0.6 mM tyrosine, 2% glycerol and 1%

DMSO. Reactions were carried out for 20 minutes at room temperature. Reaction mixtures were then buffer exchanged to 1x PBS using a centrifugal filter device (Vivascience, 10,000 MWCO) and concentrated to approximately 25 μM protein. Crosslinking was then induced by treatment with a 325 nM HeCd laser for 10-15 seconds.

Gel Shift Assay

5'-Fluorescein labeled T7A1 promoter DNA fragment (5'-AGGTATTGACAACATG-3') and the complementary, non-labeled strand were ordered from IDT DNA. Oligos were suspended in 1x TE buffer and annealed by mixing the complementary strands (0.5 mM each), heating to 100 °C for 5 minutes and allowing to slowly return to room temperature. The annealed mixture was then diluted with 10 mM tris·HCl (pH 8.5) to 10 μM. Crosslinker labeled σ-A constructs were used from stock solutions at 20 μM in 1x PBS. Labeled σ-A and promoter DNA solutions were combined in a 1:1 ratio (final concentrations 5 μM DNA, 10 μM DNA, 0.5x PBS, 5 mM tris·HCl (pH 8.5)) and incubated at 37 °C for 10 minutes. Glycerol was added from an 80% stock solution to a final concentration of 5%. Samples were separated on a 1% agarose gel in 0.5x TBE and visualized by UV fluorescence (Versa Doc Model 3000 Imaging System, Biorad).

CNBr digestion

CNBr digestion was performed by heating protein (1 mg/mL for crosslinking samples, 5 mg/mL for WT sample used to identify CNBr digestion products) with 1.3% SDS at 42 °C for 10 minutes. Reactions were carried out in 0.1 N HCl, 100 mM CNBr, 1% SDS overnight at room temperature. Reactions were quenched by neutralization with 1N NaOH.

BNPS-Skatole Digestion

BNPS-Skatole digestion was performed in 70% acetic acid with 3.33 mg/mL BNPS-skatole (Sigma) overnight at room temperature. Protein concentration was 0.5 μg/mL for the digests in **Figure 4-37** and 1 μg/mL for the digests in **Figure 4-38**. Following overnight digestion, samples were treated with 67mM DTT at room temperature for 20 minutes. Samples were extracted with ether and aqueous fractions were lyophilized and resuspended in SDS loading buffer.

Two Step Digestion

BNPS-Skatole digested samples were separated by SDS-PAGE and visualized by coomassie stain. Bands were excised with a clean razor blade and homogenized in 1N HCl, ~0.5-1% SDS with mortar and pestle then incubated at 100 °C for 5 minutes. Samples were digested with 100 mM CNBr overnight at room temperature as above.

Mass Spectrometry identification of CNBr digestion fragments

By extraction of digested fragments

Gel bands were excised and destained with a 1:1 solution of ferrocyanide (10 mg/mL) and thiosulfate (16 mg/mL) then rinsed with 50 mM NH_4HCO_3. Gel bands were further destained in 50% acetonitrile/50mM NH_4HCO_3 then shrunk in 100% acetonitrile and dried in a speed vac. The gel band was then reduced with 10 mM DTT in 50 mM NH_4HCO_3 at 56 °C for 45 minutes. After cooling to room temperature, the gel band was alkylated with 100 mM iodoacetamide in 50 mM NH_4HCO_3 for 30 minutes in the dark. The gel was then washed with 50% acetonitrile/50mM NH_4HCO_3 then shrunk in 100% acetonitrile and dried in a speed vac. Trypsin (0.2 μg in 50 mM NH_4HCO_3, 0.1% N-octyl β-D-glucopyranoside , 5 mM $CaCl_2$) was added to the dry gel slice and allowed to absorb then incubated overnight at 37 °C. Tryptic fragments were then eluted 3 times (elution 1 : 10% acetic acid, elution 2 : 30% acetonitrile/5% TFA, elution 3 : 50% acetonitrile/5% TFA) and the supernatant was dried in a speed vac. Extracted protein was then resuspended in water and analyzed by MALDI-TOF (Applied Biosystems DE-STR MALDI-TOF mass spectrometer) and ESMS/MS (Applied Biosystems QSTAR XL tandem mass spectrometer). CNBr fragments were identified as the smallest digestion product that contained all the tryptic fragments found in a given band.

RP-HPLC of CNBr digest

CNBr digested wild type σ-A was resolved on a Vydac C4 analytical column using a 30-55% gradient of buffer B (0.1% (v/v) TFA, 90% (v/v) acetonitrile) in buffer A (0.1% (v/v) aqueous TFA). HPLC was carried out using a Hewlett-Packard 1100 series instrument with 214 nm and 280 nm detection. Peak fractions were collected and analyzed by ESMS. CNBr fragments were identified by comparing experimentally obtained masses to the masses of predicted CNBr digestion products.

Avidin enrichment

Crosslinked, CNBr digested samples were neutralized with 1N NaOH and reduced by treatment with DTT (50 mM). Samples were boiled and then diluted 25-fold in 50 mM Tris, 4 mM DTT (pH 8.0) and applied to monomeric avidin resin (Promega). Samples were incubated with gentle rocking for 1 hour at 4 °C then washed 3 times with 50 mM Tris, 4 mM DTT (pH 8.0). Elution was attempted first with 50 mM Tris, 0.1% SDS, 4 mM DTT, 5 mM biotin and, after that failed, by boiling in SDS loading buffer.

References

1. Shogren-Knaak, M.A., P.J. Alaimo, and K.M. Shokat, *Recent Advances in Chemical Approaches to the Study of Biological Systems.* Annu. Rev. Cell Dev. Biol., 2001. **17**: p. 405-433.
2. Bishop, A.C., et al., *A chemical switch for inhibitor-sensitive alleles of any protein kinase.* Nature, 2000. **407**(6802): p. 395-401.
3. Ubersax, J.A., et al., *Targets of the cyclin-dependent kinase Cdk1.* Nature, 2003. **425**(6960): p. 859-64.
4. Wang, H., et al., *Inducible protein knockout reveals temporal requirement of CaMKII reactivation for memory consolidation in the brain.* Proc. Natl. Acad. Sci. U S A, 2003. **100**(7): p. 4287-92.
5. Hwang, Y.W. and D.L. Miller, *A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein.* J. Biol. Chem., 1987. **262**(27): p. 13081-5.
6. Powers, T. and P. Walter, *Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases.* Science, 1995. **269**(5229): p. 1422-4.
7. Kapoor, T.M. and T.J. Mitchison, *Allele-specific activators and inhibitors for kinesin.* Proc. Natl. Acad. Sci. U S A, 1999. **96**(16): p. 9106-11.
8. Shi, Y. and J.T. Koh, *Functionally orthogonal ligand-receptor pairs for the selective regulation of gene expression generated by manipulation of charged residues at the ligand-receptor interface of ER alpha and ER beta.* J. Am. Chem. Soc., 2002. **124**(24): p. 6921-8.
9. Spencer, D.M., et al., *Controlling Signal Transduction with Synthetic Ligands.* Science, 1993. **262**: p. 1019-1024.
10. Pruschy, M.N., et al., *Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012.* Chem. Biol., 1994. **1**(3): p. 163-72.
11. Freiberg, R.A., et al., *Specific triggering of the Fas signal transduction pathway in normal human keratinocytes.* J. Biol. Chem., 1996. **271**(49): p. 31666-9.
12. Pollock, R. and T. Clackson, *Dimerizer-regulated gene expression.* Curr. Op. Biotech., 2002. **13**: p. 459-467.
13. Spencer, D.M., et al., *A general strategy for producing conditional alleles of Src-like tyrosine kinases.* Proc. Natl. Acad. Sci. U S A, 1995. **92**(21): p. 9805-9.
14. Kohler, J.J. and C.R. Bertozzi, *Regulating Cell Surface Glycosylation by Small Molecule Control of Enzyme Localization.* Chem. Biol., 2003. **10**: p. 1303-1311.
15. de Graffenried, C.L., et al., *A small-molecule switch for Golgi sulfotransferases.* Proc. Natl. Acad. Sci. U S A, 2004. **101**(48): p. 16715-16720.
16. Johnsson, N. and A. Varshavsky, *Split ubiquitin as a sensor of protein interactions in vivo.* Proc. Natl. Acad. Sci. U S A, 1994. **91**: p. 10340-10344.

17. Hu, C.-D. and T.K. Kerppola, *Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis.* Nat. Biotechnol., 2003. **21**: p. 539-545.
18. Wilson, C.G., T.J. Magliery, and L. Regan, *Detecting protein-protein interactions with GFP-fragment reassembly.* Nat. Methods, 2004. **1**(3): p. 255-62.
19. Magliery, T.J., et al., *Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism.* J. Am. Chem. Soc., 2005. **127**(1): p. 146-57.
20. Magliery, T.J. and L. Regan, *Reassembled GFP: detecting protein-protein interactions and protein expression patterns.* Methods Biochem. Anal., 2006. **47**: p. 391-405.
21. Luker, K.E., et al., *Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals.* Proc. Natl. Acad. Sci. U S A, 2004. **101**: p. 12288-12293.
22. Paulmurugan, R. and S.S. Gambhir, *Firefly luciferase enzyme fragment complementation for imaging in cells and living animals.* Anal. Chem., 2005. **77**(5): p. 1295-302.
23. Mohler, W.A. and H.M. Blau, *Gene expression and cell fusion analyzed by lacZ complementation in mammalian cells.* Proc. Natl. Acad. Sci. U S A, 1996. **93**(22): p. 12423-7.
24. Rossi, F., C.A. Charlton, and H.M. Blau, *Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation.* Proc. Natl. Acad. Sci. U S A, 1997. **94**(16): p. 8405-10.
25. Wehrman, T., et al., *Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments.* Proc. Natl. Acad. Sci. U S A, 2002. **99**(6): p. 3469-74.
26. Remy, I. and S.W. Michnick, *Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays.* Proc. Natl. Acad. Sci. U S A, 1999. **96**(10): p. 5394-9.
27. Galarneau, A., et al., *Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions.* Nat. Biotechnol., 2002. **20**(6): p. 619-22.
28. Spencer, D.M., et al., *Functional analysis of Fas signaling in vivo using synthetic inducers of dimerization.* Curr. Biol., 1996. **6**(7): p. 839-47.
29. Sakamoto, K.M., et al., *Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation.* Proc. Natl. Acad. Sci. U S A, 2001. **98**(15): p. 8554-9.
30. Stankunas, K., et al., *Conditional Protein Alleles Using Knockin Mice and a Chemical Inducer of Dimerization.* Mol. Cell, 2003. **12**: p. 1615-1624.
31. Liu, K.J., et al., *Chemical rescue of cleft palate and midline defects in conditional GSK-3β mice.* Nature, 2007. **446**(7131): p. 79-82.

32. Banaszynski, L.A., et al., *A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules.* Cell, 2006. **126**(5): p. 995-1004.
33. Metzger, D., et al., *Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase.* Proc. Natl. Acad. Sci. U S A, 1995. **92**(15): p. 6991-5.
34. Feil, R., et al., *Ligand-activated site-specific recombination in mice.* Proc. Natl. Acad. Sci. U S A, 1996. **93**: p. 10887-10890.
35. Indra, A.K., et al., *Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases.* Nucleic Acids Res., 1999. **27**(22): p. 4324-7.
36. Buskirk, A.R., et al., *Directed evolution of ligand dependence: Small-molecule-activated protein splicing.* Proc. Natl. Acad. Sci. U S A, 2004. **101**(29): p. 10505-10510.
37. Yuen, C.M., et al., *Control of transcription factor activity and osteoblast differentiation in Mammalian cells using an evolved small-molecule-dependent intein.* J. Am. Chem. Soc., 2006. **128**(27): p. 8939-46.
38. Gruber, T.M. and C.A. Gross, *Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space.* Annu. Rev. Microbiol., 2003. **57**: p. 441-466.
39. Dombroski, A.J., et al., *Polypeptides Containing Highly Conserved Regions of Transcription Initiation Factor Sigma-70 Exhibit Specificity of Binding to Promoter DNA.* Cell, 1992. **70**: p. 501-512.
40. Dombroski, A.J., W.A. Walter, and C.A. Gross, *Amino-terminal amino acids modulate sigma-factor DNA-binding activity.* Genes & Dev., 1993. **7**: p. 2446-2455.
41. Murakami, K.S., S. Masuda, and S.A. Darst, *Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 A Resolution.* Science, 2002. **296**: p. 1280-1284.
42. Murakami, K.S., et al., *Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex.* Science, 2002. **296**: p. 1285-1290.
43. Vassylyev, D.G., et al., *Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution.* Nature, 2002. **417**: p. 712-719.
44. Laudano, A.P. and R.F. Doolittle, *Studies on synthetic peptides that bind to fibrinogen and prevent fibrin polymerization. Structural requirements, number of binding sites, and species differences.* Biochemistry, 1980. **19**(5): p. 1013-9.
45. Ueno, H., et al., *Inhibition of PDGF beta receptor signal transduction by coexpression of a truncated receptor.* Science, 1991. **252**(5007): p. 844-8.
46. Amaya, E., T.J. Musci, and M.W. Kirschner, *Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos.* Cell, 1991. **66**(2): p. 257-70.

47. Yin, H., et al., *Computational design of peptides that target transmembrane helices.* Science, 2007. **315**(5820): p. 1817-22.
48. Belshaw, P.J., et al., *Controlling programmed cell death with a cyclophilin-cyclosporin-based chemical inducer of dimerization.* Chem. Biol., 1996. **3**(9): p. 731-8.
49. Ptashne, M., *How eukaryotic transcriptional activators work.* Nature, 1988. **335**(6192): p. 683-9.
50. Ptashne, M. and A. Gann, *Transcriptional activation by recruitment.* Nature, 1997. **386**(6625): p. 569-77.
51. Chien, C.T., et al., *The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest.* Proc. Natl. Acad. Sci. U S A, 1991. **88**(21): p. 9578-82.
52. Fields, S. and R. Sternglanz, *The two-hybrid system: an assay for protein-protein interactions.* Trends Genet., 1994. **10**(8): p. 286-92.
53. Castellano, F., et al., *Inducible recruitment of Cdc42 or WASP to a cell-surface receptor triggers actin polymerization and filopodium formation.* Curr. Biol., 1999. **9**(7): p. 351-60.
54. Belshaw, P.J., et al., *Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins.* Proc. Natl. Acad. Sci. U S A, 1996. **93**(10): p. 4604-7.
55. Lin, H., et al., *Dexamethasone-Methotrexate: An Efficient Chemical Inducer of Protein Dimerization In Vivo.* J. Am. Chem. Soc., 2000. **122**: p. 4247-4248.
56. Massoud, T.F., et al., *Reporter gene imaging of protein-protein interactions in living subjects.* Curr. Opin. Biotechnol., 2007. **18**(1): p. 31-7.
57. Young, M.W. and S.A. Kay, *Time Zones: A Comparative Genetics of Circadian Clocks.* Nat. Rev. Genet., 2001. **2**: p. 702-715.
58. Kubori, T. and J.E. Galan, *Temporal regulation of salmonella virulence effector function by proteasome-dependent protein degradation.* Cell, 2003. **115**(3): p. 333-42.
59. Riedl, S.J. and Y. Shi, *Molecular mechanisms of caspase regulation during apoptosis.* Nat. Rev. Mol. Cell Biol., 2004. **5**(11): p. 897-907.
60. Perkins, N.D., *The Rel/NF-κβ family: friend and foe.* Trends Biochem. Sci., 2000. **25**(9): p. 434-40.
61. Zhou, P., et al., *Harnessing the ubiquitination machinery to target the degradation of specific cellular proteins.* Mol. Cell, 2000. **6**(3): p. 751-6.
62. Janse, D.M., et al., *Localization to the proteasome is sufficient for degradation.* J. Biol. Chem., 2004. **279**(20): p. 21415-20.
63. Picard, D., S.J. Salser, and K.R. Yamamoto, *A movable and regulable inactivation function within the steroid binding domain of the glucocorticoid receptor.* Cell, 1988. **54**(7): p. 1073-80.
64. Picard, D., *Posttranslational regulation of proteins by fusions to steroid-binding domains.* Methods Enzymol., 2000. **327**: p. 385-401.

65. Kellendonk, C., et al., *Regulation of Cre recombinase activity by the synthetic steroid RU 486.* Nucleic Acids Res., 1996. **24**(8): p. 1404-11.
66. Brocard, J., et al., *Spatio-temporally controlled site-specific somatic mutagenesis in the mouse.* Proc. Natl. Acad. Sci. U S A, 1997. **94**(26): p. 14559-63.
67. Brocard, J., et al., *A chimeric Cre recombinase inducible by synthetic,but not by natural ligands of the glucocorticoid receptor.* Nucleic Acids Res., 1998. **26**(17): p. 4086-90.
68. Logie, C. and A.F. Stewart, *Ligand-regulated site-specific recombination.* Proc. Natl. Acad. Sci. U S A, 1995. **92**(13): p. 5940-4.
69. Nichols, M., et al., *FLP recombinase/estrogen receptor fusion proteins require the receptor D domain for responsiveness to antagonists, but not agonists.* Mol. Endocrinol., 1997. **11**(7): p. 950-61.
70. Giriat, I., T.W. Muir, and F.B. Perler, *Protein splicing and its applications.* Genet. Eng. (N Y), 2001. **23**: p. 171-99.
71. Muir, T.W., *Semisynthesis of Proteins by Expressed Protein Ligation.* Annu. Rev. Biochem., 2003. **72**: p. 249-289.
72. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angew. Chem. Int. Ed. Engl., 2005. **44**(45): p. 7342-72.
73. Paulus, H., *Protein Splicing and Related Forms of Protein Autoprocessing.* Annu. Rev. Biochem., 2000. **69**: p. 447-496.
74. Mootz, H.D. and T.W. Muir, *Protein Splicing Triggered by a Small Molecule.* J. Am. Chem. Soc., 2002. **124**: p. 9044-9045.
75. Muralidharan, V. and T.W. Muir, *Protein ligation: an enabling technology for the biophysical analysis of proteins.* Nat. Methods, 2006. **3**(6): p. 429-38.
76. Perler, F.B., *InBase: the Intein Database.* Nucleic Acids Res., 2002. **30**(1): p. 383-4.
77. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation.* Science, 1994. **266**(5186): p. 776-9.
78. Muir, T.W., D. Sondhi, and P.A. Cole, *Expressed protein ligation: a general method for protein engineering.* Proc. Natl. Acad. Sci. U S A, 1998. **95**(12): p. 6705-10.
79. Shi, J. and T.W. Muir, *Development of a Tandem Protein Trans-Splicing System Based on Native and Engineered Split Inteins.* J. Am. Chem. Soc., 2005. **127**: p. 6198-6206.
80. Giriat, I. and T.W. Muir, *Protein semi-synthesis in living cells.* J. Am. Chem. Soc., 2003. **125**(24): p. 7180-7181.
81. Adam, E. and F.B. Perler, *Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in Escherichia coli DNA gyrase subunit A.* J. Mol. Microbiol. Biotechnol., 2002. **4**(5): p. 479-87.

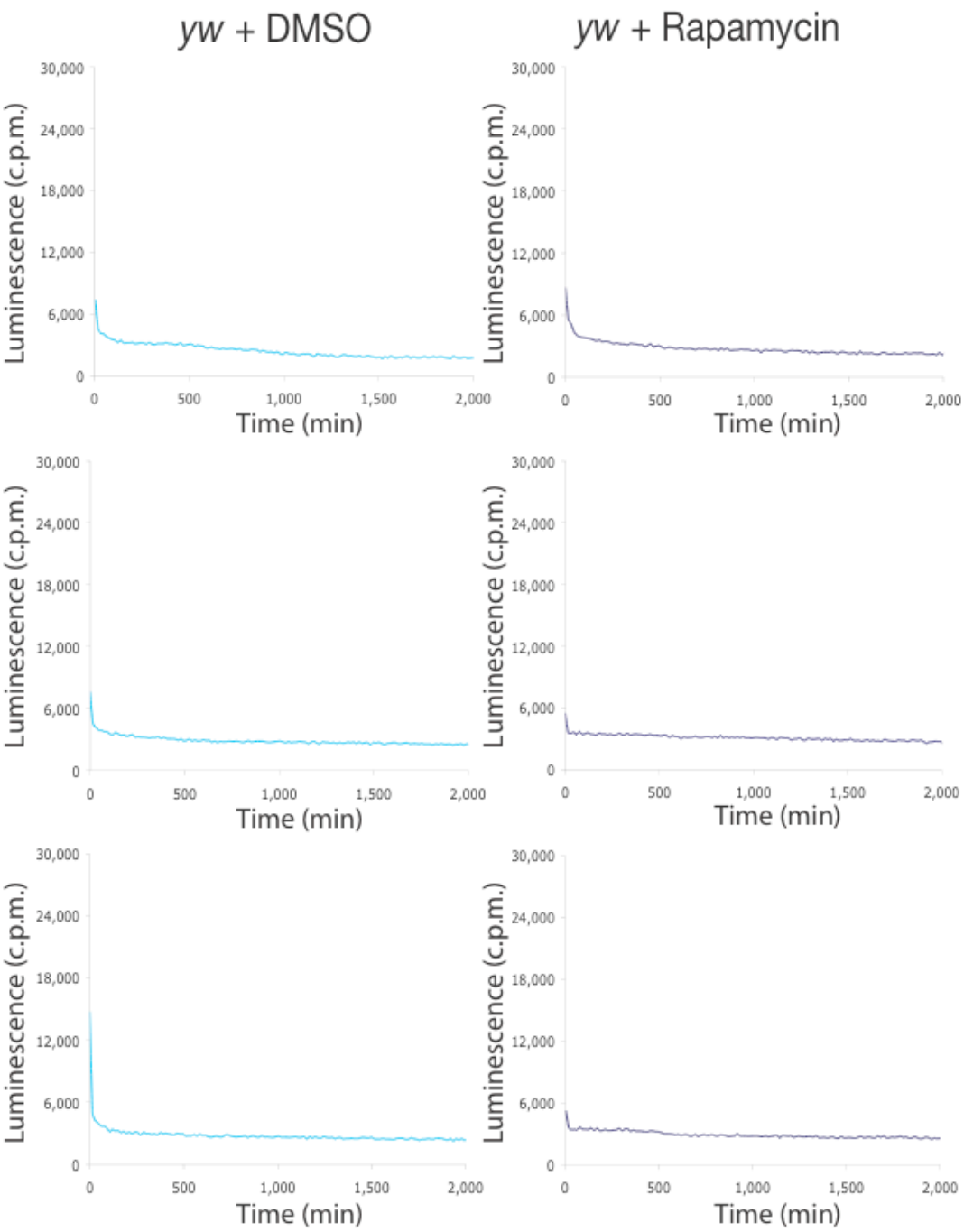
82. Zeidler, M.P., et al., *Temperature-sensitive control of protein activity by conditionally splicing inteins.* Nat. Biotechnol., 2004. **22**(7): p. 871-876.
83. Mootz, H.D., et al., *Conditional Protein Splicing: A New Tool to Control Protein Structure and Function in Vitro and in Vivo.* J. Am. Chem. Soc., 2003. **125**: p. 10561-10569.
84. Mootz, H.D., E.S. Blum, and T.W. Muir, *Activation of an Autoregulated Protein Kinase by Conditional Protein Splicing.* Angew. Chem. Int. Ed. Engl., 2004. **43**: p. 5189-5192.
85. Welsh, D.K. and S.A. Kay, *Bioluminescence imaging in living organisms.* Curr. Opin. Biotechnol., 2005. **16**(1): p. 73-8.
86. Chong, S., et al., *Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system.* J. Biol. Chem., 1996. **271**(36): p. 22159-68.
87. Chong, S. and M.Q. Xu, *Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs.* J. Biol. Chem., 1997. **272**(25): p. 15587-90.
88. Brandes, C., et al., *Novel features of drosophila period Transcription revealed by real-time luciferase reporting.* Neuron, 1996. **16**(4): p. 687-92.
89. Stempfl, T., et al., *Identification of circadian-clock-regulated enhancers and genes of Drosophila melanogaster by transposon mobilization and luciferase reporting of cyclical gene expression.* Genetics, 2002. **160**(2): p. 571-93.
90. Price, J., L., et al., *double-time Is a Novel Drosophila Clock Gene that Regulates PERIOD Protein Accumulation.* Cell, 1998. **94**: p. 83-95.
91. Kloss, B., et al., *The Drosophila Clock Gene double-time Encodes a Protein Closely Related to Human Casein Kinase Iε.* Cell, 1998. **94**: p. 97-107.
92. Ruel, L., et al., *Functional significance of a family of protein kinases encoded at the shaggy locus in Drosophila.* Embo J., 1993. **12**(4): p. 1657-69.
93. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu. Rev. Biochem., 1998. **67**: p. 425-79.
94. Baker, R.T., J.W. Tobias, and A. Varshavsky, *Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and UBP3, and functional analysis of the UBP gene family.* J. Biol. Chem., 1992. **267**(32): p. 23364-75.
95. Bogyo, M., et al., *Covalent modification of the active site threonine of proteasomal beta subunits and the Escherichia coli homolog HslV by a new class of inhibitors.* Proc. Natl. Acad. Sci. U S A, 1997. **94**(13): p. 6629-34.
96. Budihardjo, I., et al., *Biochemical pathways of caspase activation during apoptosis.* Annu. Rev. Cell Dev. Biol., 1999. **15**: p. 269-90.

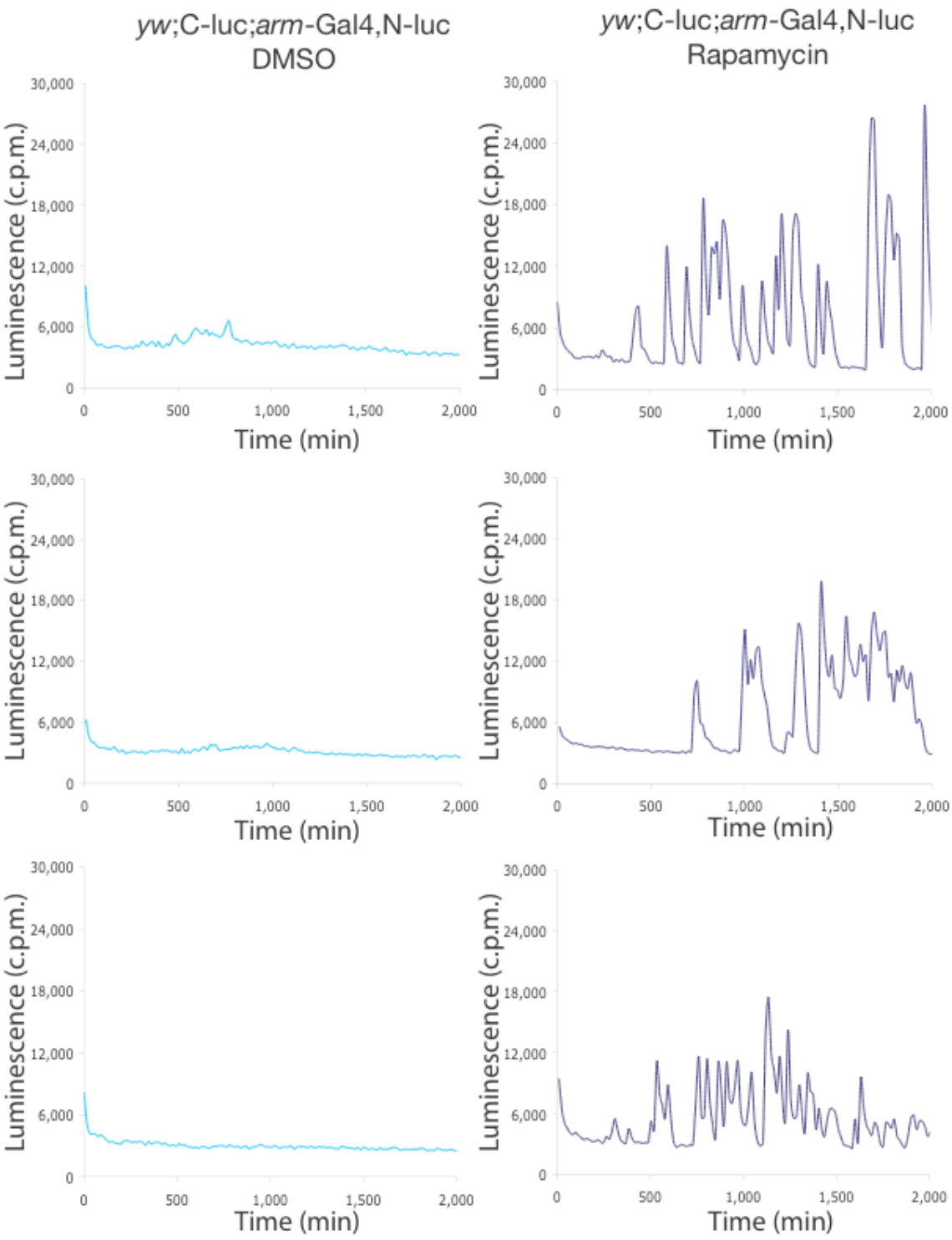
97. Frame, M.C., et al., *v-Src's hold over actin and cell adhesions.* Nat. Rev. Mol. Cell Biol., 2002. **3**(4): p. 233-45.
98. Yeatman, T.J., *A renaissance for SRC.* Nat Rev Cancer, 2004. **4**(6): p. 470-80.
99. Qiao, Y., et al., *Chemical rescue of a mutant enzyme in living cells.* Science, 2006. **311**(5765): p. 1293-7.
100. Ohno, M. and T. Abe, *Rapid colorimetric assay for the quantification of leukemia inhibitory factor (LIF) and interleukin-6 (IL-6).* J. Immunol. Methods, 1991. **145**(1-2): p. 199-203.
101. Siegel, P.M. and J. Massague, *Cytostatic and apoptotic actions of TGF-β in homeostasis and cancer.* Nat. Rev. Cancer, 2003. **3**(11): p. 807-21.
102. Shi, Y. and J. Massague, *Mechanisms of TGF-β signaling from cell membrane to the nucleus.* Cell, 2003. **113**(6): p. 685-700.
103. Chacko, B.M., et al., *The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization.* Nat. Struct. Biol., 2001. **8**(3): p. 248-53.
104. Inman, G.J., et al., *SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7.* Mol. Pharmacol., 2002. **62**(1): p. 65-74.
105. Van Obberghen-Schilling, E., et al., *Transforming growth factor β 1 positively regulates its own expression in normal and transformed cells.* J. Biol. Chem., 1988. **263**(16): p. 7741-6.
106. Chen, Y.G., F. Liu, and J. Massague, *Mechanism of TGFβ receptor inhibition by FKBP12.* Embo J., 1997. **16**(13): p. 3866-76.
107. Stockwell, B.R. and S.L. Schreiber, *TGF-β-signaling with small molecule FKBP12 antagonists that bind myristoylated FKBP12-TGF-β type I receptor fusion proteins.* Chem. Biol., 1998. **5**(7): p. 385-95.
108. Fukuchi, M., et al., *Ligand-dependent degradation of Smad3 by a ubiquitin ligase complex of ROC1 and associated proteins.* Mol. Biol. Cell, 2001. **12**(5): p. 1431-43.
109. Stankunas, K., et al., *Conditional protein alleles using knockin mice and a chemical inducer of dimerization.* Mol. Cell, 2003. **12**(6): p. 1615-24.
110. Ebright, R.H., *RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II.* J. Mol. Biol., 2000. **304**(5): p. 687-98.
111. Darst, S.A., *Bacterial RNA polymerase.* Curr. Opin. Struct. Biol., 2001. **11**(2): p. 155-62.
112. Borukhov, S. and E. Nudler, *RNA polymerase holoenzyme: structure, function and biological implications.* Curr. Opin. Microbiol., 2003. **6**(2): p. 93-100.
113. Browning, D.F. and S.J. Busby, *The regulation of bacterial transcription initiation.* Nat. Rev. Microbiol., 2004. **2**(1): p. 57-65.
114. Burgess, R.R., et al., *Factor stimulating transcription by RNA polymerase.* Nature, 1969. **221**(5175): p. 43-6.

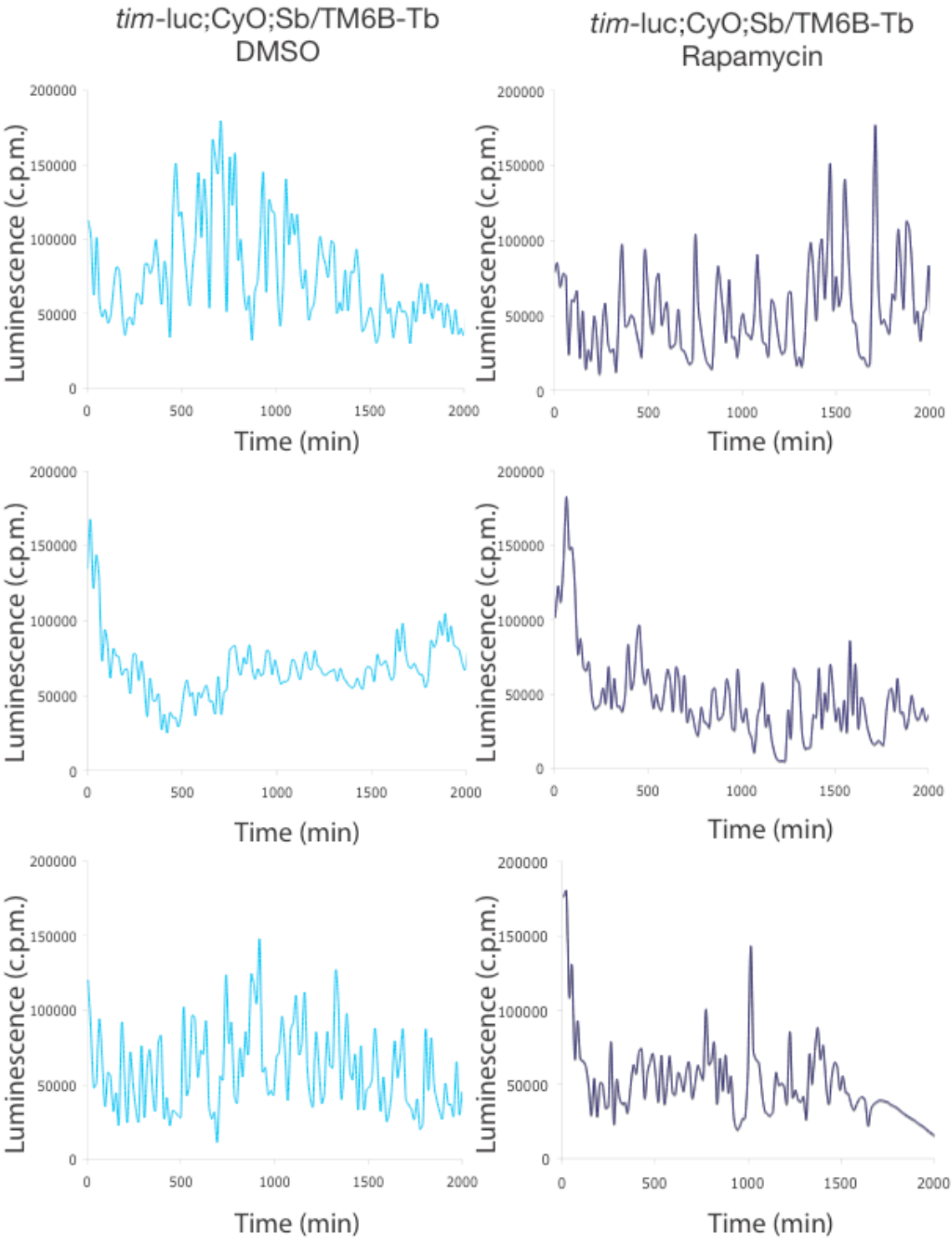
115. Travers, A.A. and R.R. Burgess, *Cyclic re-use of the RNA polymerase σ factor.* Nature, 1969. **222**(5193): p. 537-40.
116. Murakami, K.S. and S.A. Darst, *Bacterial RNA polymerases: the wholo story.* Curr. Op. Struct. Biol., 2003. **13**: p. 31-39.
117. Campbell, E.A., et al., *Crystal structure of the Bacillus stearothermophilus anti-sigma factor SpoIIAB with the sporulation sigma factor sigmaF.* Cell, 2002. **108**(6): p. 795-807.
118. Campbell, E.A., et al., *Crystal Structure of Escherichia coli σ^E with the Cytoplasmic Domain of Its Anti-σ RseA.* Mol. Cell, 2003. **11**: p. 1067-1078.
119. Camarero, J.A., et al., *Autoregulation of a bacterial σ factor explored by using segmental isotopic labeling and NMR.* Proc. Natl. Acad. Sci. USA, 2002. **99**(13): p. 8536-8541.
120. Malhotra, A., E. Severinova, and S.A. Darst, *Crystal Structure of a σ70 Subunit Fragment from E. coli RNA Polymerase.* Cell, 1996. **87**: p. 127-136.
121. Campbell, E.A., et al., *Structure of the Bacterial RNA Polymerase Promoter Specificity Sigma Subunit.* Molecular Cell, 2002. **9**: p. 527-539.
122. Sorenson, M.K., S.S. Ray, and S.A. Darst, *Crystal Structure of the Flagellar sigma/Anti-sigma Complex Sigma-28/FlgM Reveals an Intact Sigma Factor in an Inactive Conformation.* Mol. Cell, 2004. **14**: p. 127-138.
123. Lane, W.J. and S.A. Darst, *The structural basis for promoter -35 element recognition by the group IV sigma factors.* PLoS Biol., 2006. **4**(9): p. e269.
124. Gross, C.A., et al., *The functional and regulatory roles of sigma factors in transcription.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 141-55.
125. Barne, K.A., et al., *Region 2.5 of the Escherichia coli RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters.* Embo J, 1997. **16**(13): p. 4034-40.
126. Siegele, D.A., et al., *Altered promoter recognition by mutant forms of the sigma 70 subunit of Escherichia coli RNA polymerase.* J Mol Biol, 1989. **206**(4): p. 591-603.
127. Gardella, T., H. Moyle, and M.M. Susskind, *A mutant Escherichia coli sigma 70 subunit of RNA polymerase with altered promoter specificity.* J. Mol. Biol., 1989. **206**(4): p. 579-90.
128. Wilson, C. and A.J. Dombroski, *Region 1 of σ^{70} is Required for Efficient Isomerization and Initiation of Transcription by Escherichia coli RNA Polymerase.* J. Mol. Biol., 1997. **267**: p. 60-74.
129. Cavanagh, J., et al., *Protein NMR Spectroscopy; Principles and Practice, 1st ed.* 1996, San Diego, CA: Academic Press.
130. Wishart, D.S. and B.D. Sykes, *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.* J. Biomol. NMR, 1994. **4**(2): p. 171-80.
131. Holm, L. and C. Sander, *Protein structure comparison by alignment of distance matrices.* J. Mol. Biol., 1993. **233**(1): p. 123-38.

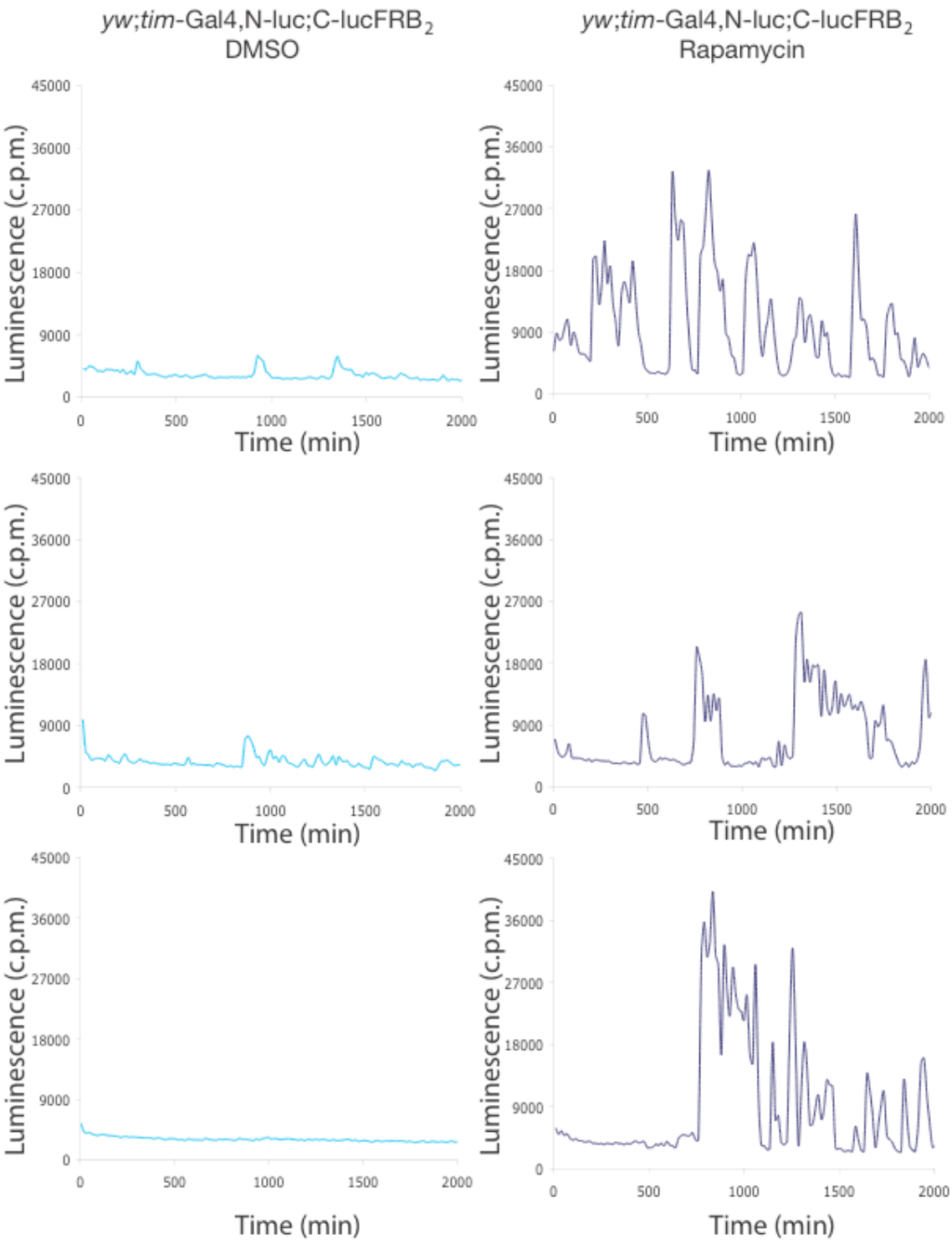
132. Bowers, C.W. and A.J. Dombroski, *A mutation in region 1.1 of σ^{70} affects promoter DNA binding by Escherichia coli RNA polymerase holoenzyme.* Embo J., 1999. **18**(3): p. 709-716.
133. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-96.
134. Callaci, S., E. Heyduk, and T. Heyduk, *Core RNA polymerase from E. coli induces a major change in the domain arrangement of the sigma 70 subunit.* Mol. Cell, 1999. **3**(2): p. 229-38.
135. Mekler, V., et al., *Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex.* Cell, 2002. **108**(5): p. 599-614.
136. Kobe, B. and B.E. Kemp, *Active site-directed protein regulation.* Nature, 1999. **402**(6760): p. 373-6.
137. Brand, A.H. and N. Perrimon, *Targeted gene expression as a means of altering cell fates and generating dominant phenotypes.* Development, 1993. **118**: p. 401-415.
138. Karess, R.E. and G.M. Rubin, *Analysis of P transposable element functions in Drosophila.* Cell, 1984. **38**(1): p. 135-46.
139. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol. Biol., 2004. **278**: p. 313-52.
140. Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulse field gradients.* Prog. NMR Spectrosc., 1999. **34**: p. 93-158.
141. Yamazaki, T., J.D. Forman-Kay, and L.E. Kay, *Two-dimensional NMR experiments for correlating C-13β and H-1δ/ε chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings.* J. Am. Chem. Soc., 1993. **115**: p. 11054-11055.
142. Cornilescu, G., F. Delaglio, and A. Bax, *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* J. Biomol. NMR, 1999. **13**(3): p. 289-302.
143. Linge, J.P., et al., *ARIA: automated NOE assignment and NMR structure calculation.* Bioinformatics, 2003. **19**(2): p. 315-6.
144. Habeck, M., et al., *NOE assignment with ARIA 2.0: the nuts and bolts.* Methods Mol. Biol., 2004. **278**: p. 379-402.
145. Linge, J.P., et al., *Refinement of protein structures in explicit solvent.* Proteins, 2003. **50**(3): p. 496-506.
146. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J. Biomol. NMR, 1996. **8**(4): p. 477-86.

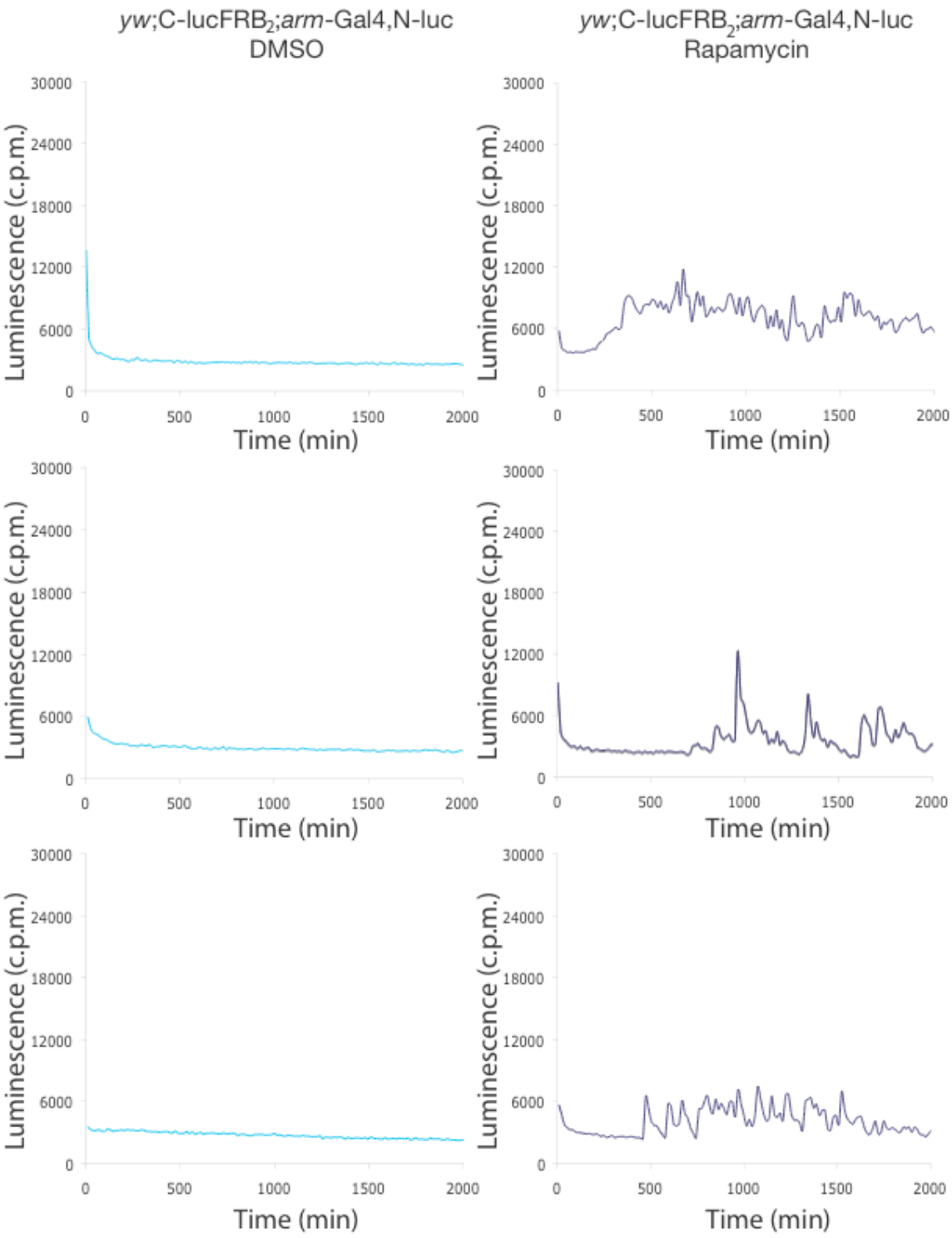
Appendix 1 : Additional Fly luciferase traces




yw;C-luc;*arm*-Gal4,N-luc
DMSO
yw;C-luc;*arm*-Gal4,N-luc
Rapamycin
Luminescence (c.p.m.)
Time (min)
30,000
24,000
18,000
12,000
6,000
0
500
1,000
1,500
2,000


tim-luc;CyO;Sb/TM6B-Tb
DMSO
tim-luc;CyO;Sb/TM6B-Tb
Rapamycin
Luminescence (c.p.m.)
Time (min)
0
50000
100000
150000
200000
500
1000
1500
2000


yw;tim-Gal4,N-luc;C-lucFRB2
DMSO
yw;tim-Gal4,N-luc;C-lucFRB2
Rapamycin
Luminescence (c.p.m.)
Time (min)
0
9000
18000
27000
36000
45000
500
1000
1500
2000


yw;C-lucFRB2;arm-Gal4,N-luc
DMSO
yw;C-lucFRB2;arm-Gal4,N-luc
Rapamycin
Luminescence (c.p.m.)
Time (min)
0
6000
12000
18000
24000
30000
500
1000
1500
2000