

**UCSF**

**UC San Francisco Electronic Theses and Dissertations**

**Title**

Mechanisms of pre-mRNA splicing in yeast

**Permalink**

<https://escholarship.org/uc/item/2wj5816h>

**Author**

Hadjivassiliou, Haralambos Antonis

**Publication Date**

2014

Peer reviewed|Thesis/dissertation

**Mechanisms of Pre-mRNA Splicing in Yeast**

by

**Haralambos A. Hadjivassiliou**

**DISSERTATION**

**Submitted in partial satisfaction of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

in

**Biochemistry and Molecular Biology**

in the

**GRADUATE DIVISION**

of the

**UNIVERSITY OF CALIFORNIA, SAN FRANCISCO**



## **Acknowledgements**

I would first like to thank my mentor Christine Guthrie and her husband John Abelson for teaching me how science should be done. Countless times I would be inspired by their probing wit, their ethical integrity, their childlike excitement for new discovery and their unending generosity and graciousness. These are not empty platitudes, but rather statements of fact and I will forever be grateful for my experiences in the lab.

Secondly I would like to acknowledge my wife Mary Matyskiela, aka Bubba, aka the little chicken. I do not know where I would be with out my better half. From editing my bioreg proposal the night before it was due to throwing me a life preserver when things got rough, my chicken was there for me. She is an amazing woman and every day I am grateful that we share this life together. We also have a lot of fun together.

Thirdly I would like to acknowledge David Pincus, who was my best friend during graduate school. David and I spent much time discussing science and life and I will always hold onto the fond memories we made at the Parkside. David is an amazing scientist and friend and I know that he will always do spectacular things.

Finally I want to recognize the newest little light in my life, Stella Matyskiela Hadjivassiliou, aka Stella Bo, aka le petit jambon. This little girl was born to Bubba and I on September 22, 2013 and now she is almost a year. She is a source of happiness and purpose. The glee and excitement that she expresses so often is a constant reminder of the joy of life. Big Bo, this one's for you!

Methodological accomplishments described in Chapter 2 were published in *Methods in Enzymology*: Abelson J, Hadjivassiliou H, Guthrie C. Preparation of fluorescent pre-mRNA substrates for an smFRET study of pre-mRNA splicing in yeast. *Methods Enzymology* 2010;472:31-40. doi: 10.1016/S0076-6879(10)72017-6. PubMed PMID: 20580958.

Chapter 3 is reproduced from work published in *RNA* in collaboration with Oren Rosenberg: Hadjivassiliou H, Rosenberg OS, Guthrie C. The crystal structure of *S.cerevisiae* Sad1, a catalytically inactive deubiquitinase that is broadly required for pre-mRNA splicing. *RNA*. 2014 May;20(5):656-69.

Chapter 4 contains ideas that were the basis of an NIH and NSF proposal

Chapter 5 is adapted from a grant written by Haralambos Hadjivassiliou and submitted to the QB3. The proposal was intended to combine 30 years of yeast genetics and biochemistry from Christine Guthrie's lab with the quantitative high-throughput screening approaches developed in Nevan Krogan's lab, the innovative electron microscopy pioneered in Yifan Cheng's lab, and the robust modeling programs engineered in Andrej Sali's lab.

## **Abstract**

The spliceosome is one of the cells largest and important molecular machines and yet it remains as one the least understood systems. This is a result of many experimentally challenging features inherent to the biology of splicing. In addition to having the largest number of unique proteins and RNAs associated with one process, the spliceosome has a complex assembly process that is both conformationally and compositionally dynamic. Classical genetics, much of it done in the Guthrie lab, revealed which proteins and RNAs were required for the process of splicing, but the molecular details of their interactions is still relatively unknown. In the past 10 years new biochemical tools have been developed, such as smFRET and metal substitution studies, which have increased our abilities to probe the structure and activity of the spliceosome. These techniques and others outlined in this proposal will be the foundation for the next ten years of research on the spliceosome. I believe that through these efforts and building on recent breakthroughs, the molecular mechanism of pre-mRNA splicing will be determined. These are truly exciting times to be studying the spliceosome.

## Table of Contents by Chapter

Abstract.....	v
List of Tables.....	vii
List of Figures.....	viii
Chapter 1: Introduction to the spliceosome.....	1
Chapter 2: SmFRET of the spliceosome.....	6
Chapter 3: Crystal structure of Sad1.....	27
Chapter 4: Molecular toggles of the spliceosome.....	84
Chapter 5: An integrated approach towards the structure of the spliceosome.....	98

## List of Tables

Table 1: Strains tested for complementation of <i>sad1</i> Δ::kan by 5-FOA selection.....	54
Table 2: Sad1 data collection and refinement statistics.....	55
Table 3: ZnF-UBP domain alignment with the ZnF-UBP domain of Sad1 (amino acids 29-124).....	57
Table 4: USP domain alignment with the iUSP domain of Sad1 (amino acids 125- 448).....	58



## List of Figures by Chapter

### Chapter 2:

Figure 1: Coupling of U1 displacement and U6 5' splice site recognition.....	19
Figure 2: Dynamics of U6 snRNA with predicted conformations and FRET values.....	20
Figure 3: Experimental design for Aim 1.....	21
Figure 4: Predicted FRET values for U6* in various molecular contexts.....	22
Figure 5: U6* reconstitutes wild-type splicing activity.....	23

### Chapter 3:

Figure 1: Sad1 is a core splicing factor necessary for spliceosome assembly.....	66
Figure 2: Sad1 does not bind ubiquitin <i>in vitro</i> .....	67
Figure 3: The crystal structure of full-length Sad1.....	68
Figure 4: The ZnF-UBP domain of Sad1 binds one Zn atom. ....	69
Figure 5: The L2A and $\beta$ 4-5 loops of Sad1's ZnF-UBP domain are contracted relative to other ZnF-UBPs.....	70
Figure 6: The iUSP domain of Sad1 aligns well with known USP structures.....	72
Supplemental Figure 1: Sequence alignment of Sad1 homologues from <i>S.</i> <i>cerevisiae</i> .....	73
Supplemental Figure 2: Measuring the binding affinity of Sad1 and USP5 ZnF-UBP for ubiquitin by biolayer interferometry.....	74
Supplemental Figure 3: Sad1 does not bind to yeast GST-UBLs <i>in vitro</i> .....	75
Supplemental Figure 4: Molprobit statistics for the Sad1 structure.....	76

Supplemental Figure 5: The ZnF-UBP from USP5 is missing the B site CCHH zinc binding motif.....77

**Chapter 4:**

Figure 1: SmFRET of the spliceosome as predicted by John Abelson's toggle model.....94

Figure 2: SmFRET of the spliceosome as predicted by Charles Query and Magda Konarska's toggle model.....95

**Chapter 5:**

Figure 1: Flow chart for an integrated approach to the structure of the spliceosome.....103

## 1. Introduction to the spliceosome

The path towards a mature mRNA that is competent for translation begins in the nucleus of eukaryotic organisms. Here genes are transcribed and the corresponding pre-mRNAs are processed to mature mRNAs in preparation for export from the nucleus. In addition to 5' end capping and 3' poly-A addition, many pre-mRNAs contain introns, which must be removed in order to be competent for translation. Introns are regions of non-coding sequence interspersed amongst the sequence that codes for the protein to be expressed, or the exons. Introns can be found in almost every gene in metazoan organisms. In *S. cerevisiae* only 5% of the genes contain introns, but these introns are found in some of the most highly expressed genes, such that intron containing pre-mRNAs account for a third of all transcription in the yeast cell (Pleiss et al. 2007). Intron boundaries are marked by conserved sequence elements that serve as recognition motifs for the spliceosome, a well-conserved enzyme that is capable of removing introns and splicing together the intervening sequences.

The removal of introns from eukaryotic pre-mRNAs is catalyzed by the spliceosome, a massive and dynamic enzyme composed of 5 RNAs that associate with over 150 proteins in yeast and over 200 proteins in humans (Rappsilber et al. 2002; Jurica and Moore 2003). Many of the core components of the spliceosome are remarkably well conserved. For instance, the 240 kDa protein Prp8 is 62% conserved from *S. cerevisiae* to *H. sapiens* (Shea et al. 1994). With the large number of spliceosomal factors required for activity there exists numerous ways splicing can be disrupted, this can lead to a wide range of physiological defects including cardiomyopathies (Ding et al. 2004), retinitis pigmentosa (Mozaffari-Jovin et al. 2014),

myotonic dystrophy (Echeverria and Cooper 2014), and myeloid and lymphoid malignancies (Chesnais et al. 2014). The catalytically active spliceosome is formed from the step-wise assembly of five small nuclear ribonucleic particles (snRNPs) onto an intron-containing pre-mRNA substrate. Each snRNP is composed of multiple proteins and a small nuclear RNA (snRNA). The first snRNP to bind the pre-mRNA substrate is the U1 snRNP, which recognizes the 5' splice site of the intron. This is followed by U2 snRNP binding to the branch point in the 3' region of the intron. This U1 and U2 snRNP-bound pre-mRNA complex is termed the "pre-spliceosome". The remaining three snRNPs then bind the pre-spliceosome as the "tri-snRNP", a super-complex composed of the U4, U5 and U6 snRNPs. Upon binding of the tri-snRNP to the pre-spliceosome, large-scale conformational rearrangements occur, resulting in the release of the U1 and U4 snRNPs and the formation of the activated state of the spliceosome.

In its activated state, the spliceosome is now competent to bind the final factors required to catalyze the first and second catalytic steps of splicing. For the first step of splicing the branchpoint adenosine performs a magnesium dependent transesterification reaction with the 5' splice site. This reaction is activated by the presence of Cwc25, an essential splicing factor (Chiu et al. 2009). Upon completion of this step, Prp16, an ATP dependent helicase, acts to remove the lariat-intermediate product from the active site of the spliceosome. In the second step of splicing, the now free 5' exon performs a transesterification reaction with the 3' splice site. This step is activated by two transiently associated factors, Slu7 and Prp18 (Ohrt et al. 2013). With the exons spliced, the Prp22 ATP dependent helicase releases the mRNA product and

the spliceosome disassembles allowing for the remaining snRNPs to be recycled and assembled upon the next intron containing pre-mRNA (Staley and Guthrie 1998; Wahl et al. 2009).

When I joined the lab in 2006 there was much that was unknown about the spliceosome. Fundamentally, a high-resolution structure of the spliceosome in any of its states of assembly or catalysis was and still is unknown. Though much work has been accomplished on individual spliceosomal proteins or spliceosomal sub-complexes. For instance in 2009 the crystal structure of the U1 snRNP was published and as of early 2014, we now know more than two thirds of the structure of the largest spliceosomal protein, Prp8. In addition to these structural limitations, quantitative studies of activity or structural dynamics were hampered by the inability to reconstitute in vitro splicing using purified components. As such in vitro splicing was and still is a poorly synchronized reaction that must be performed in fairly crude cellular lysates. As a person with a background in protein chemistry, protein x-ray crystallography and biophysics I saw this state of affairs as a challenge ripe with opportunity. I worked on three projects and wrote a proposal for a fourth all of which sought to expand our structural and biochemical understanding of the spliceosome.

## References

- Chesnais, V., Kosmider, O., Damm, F., Itzykson, R., Bernard, O.A., Solary, E., and Fontenay, M. 2014. Spliceosome mutations in myelodysplastic syndromes and chronic myelomonocytic leukemia. *Oncotarget* **3**(11): 1284-1293.
- Chiu, Y.F., Liu, Y.C., Chiang, T.W., Yeh, T.C., Tseng, C.K., Wu, N.Y., and Cheng, S.C. 2009. Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction. *Mol Cell Biol* **29**(21): 5671-5678.
- Ding, J.H., Xu, X., Yang, D., Chu, P.H., Dalton, N.D., Ye, Z., Yeakley, J.M., Cheng, H., Xiao, R.P., Ross, J. et al. 2004. Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* **23**(4): 885-896.
- Echeverria, G.V. and Cooper, T.A. 2014. Muscleblind-like 1 activates insulin receptor exon 11 inclusion by enhancing U2AF65 binding and splicing of the upstream intron. *Nucleic Acids Res* **42**(3): 1893-1903.
- Jurica, M.S. and Moore, M.J. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**(1): 5-14.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Luhrmann, R., and Wahl, M.C. 2014. Novel regulatory principles of the spliceosomal Brr2 RNA helicase and links to retinal disease in humans. *RNA Biol* **11**(4).
- Ohr, T., Odenwalder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzova, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P. et al. 2013. Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**(7): 902-915.

- Pleiss, J., Whitworth, G., Bergkessel, M., and Guthrie, C. 2007. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol* **5**: e90.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**(8): 1231-1245.
- Shea, J.E., Toyn, J.H., and Johnston, L.H. 1994. The budding yeast U5 snRNP Prp8 is a highly conserved protein which links RNA splicing with cell cycle progression. *Nucleic Acids Res* **22**(25): 5555-5564.
- Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**(3): 315-326.
- Wahl, M.C., Will, C.L., and Luhrmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**(4): 701-718.

## 2. SmFRET of the spliceosome

### Motivation

The spliceosome is conformationally dynamic and is typically assayed in cell-free extracts where, the population of intermediates is complex and heterogeneous. Thus with ensemble biochemical approaches, information on the order and timing of spliceosomal rearrangements is lost. This suggests that Fluorescence Resonance Energy Transfer (FRET) would be an ideal tool as it can be used to measure the rate of conformational changes at the single molecule level (smFRET). For my studies, I focused on the spliceosomal rearrangements leading to catalytic activation, a step that requires the ATP-dependent unwinding of U4 from U6 snRNAs and is catalyzed by Brr2, an RNA-dependent ATPase of the DEAD-box super-family. Through the careful placement of FRET probes in the U6 snRNA within an in vitro reconstituted snRNP, my goal was to monitor the transition from the duplex U4/U6 form to rearranged form of U6 alone, containing a well-characterized intra-molecular stem-loop (U6-ISL). In so doing, I would be able to obtain information on the timing of initiation of unwinding and the kinetics of the unwinding event. I would then exploit our large series of mutants (in both snRNAs and spliceosomal proteins) to gain mechanistic insights into these processes. In particular, I would test the model that coupled conformational rearrangements in spliceosome assembly allow the opportunity for proofreading.



One of the remaining challenges to understanding the mechanism of splicing is how the kinetics and timing of these ordered rearrangements are determined. Spliceosomal RNA/RNA and RNA/protein rearrangements are mediated by the RNA dependent DExD/H box ATPase family members. The helicase activity of spliceosomal DExD/H box ATPases is controversial: while in vivo, it has been shown that RNA-RNA rearrangements that occur during splicing are dependent on these helicases, in vitro, helicase activities are often weak and do not appear to be substrate specific (Silverman et al. 2003). This may be explained by the fact that, in the spliceosome, specificity and activity are regulated through unidentified protein-protein interactions that have not been present in in vitro helicase experiments. Thus context is important.

Genetic and biochemical studies have identified the DExD/H box ATPase Brr2 as the helicase that is required for the unwinding of the U4/U6 snRNA duplex in vivo and the activation of the spliceosome (Laggerbauer et al. 1998; Raghunathan and Guthrie 1998; Kim and Rossi 1999), but its helicase activity in vitro was poor. Our lab has provided a new insight into the activity of Brr2 through the discovery of a regulatory role for Prp8 (Maeder et al. 2009b). Through in vitro helicase studies, we have shown that Brr2 demonstrates higher helicase activity towards the U4/U6 snRNA duplex when it is in the presence of a C-terminal fragment of prp8, a core component of the spliceosome that is known to have physical and genetic interactions with Brr2 (van Nues and Beggs 2001; Kuhn et al. 2002).

It is not surprising that Brr2 is regulated; it joins the spliceosome associated with U4/U6 in the tri-snRNP complex and thus its helicase activity must be inhibited to prevent premature release of U6. Additionally, although it is present in the spliceosome

throughout the entire splicing cycle, it is only active at the beginning and end of the process. Factors that stimulate or inhibit its helicase activity would therefore insure the proper timing of the snRNA rearrangements it mediates. Mechanisms that prevent the premature unwinding of U4/U6 by Brr2 would also have important implications for proofreading at the 5' splice site. This can be achieved by coupling the initiation of U4/U6 unwinding to the displacement of the U1 snRNA and its replacement by U6 snRNA at the 5' splice site (Figure 1). The exchange of snRNAs at the 5' splice site is regulated by the helicase Prp28. Mutations in Prp28, Prp28-1, give rise to a fully assembled spliceosome but U1 is not displaced from the 5' splice site and U4/U6 unwinding is blocked. This coupling scheme would be important for proofreading of 5' splice site selection, as it would insure that splicing only proceeds once the proper binding of U6 at the 5' splice site has occurred. Indeed mutants in Prp8 have been identified that uncouple this process and in doing so suppress 5' splice site activation defects (Kuhn et al. 1999).

Additionally, Brr2 is regulated by Snu114, a GTPase that stimulates the unwinding of the U4/U6 duplex when bound to GTP and represses the unwinding when bound to GDP (Small et al. 2006). Additionally, ubiquitination of a factor in the spliceosome, most likely Prp8, is sufficient to inhibit the unwinding of the U4/U6 duplex, presumably to prevent premature U4/U6 snRNA unwinding (Bellare et al. 2008) (Figure 1). Thus it is likely that the initiation of U4/U6 unwinding is inhibited through multiple means until the proper establishment of a preceding step, at which point a switch is flipped and derepression and stimulation of helicase activity occurs.

In order to have an experimental system that allows us to monitor the events which take place during the activation of the spliceosome, our goal was to develop a single molecule FRET approach that would allow us to measure the time to the initiation of U4/U6 unwinding and the rate of its unwinding through the formation of the U6-ISL that occurs once U6 is separated from U4 (Figure 1). We would first examine U4/U6 unwinding in a reconstituted system with only the Brr2 helicase and the helicase stimulatory factor Prp8-CTF. This would provide kinetic information on the rate of U6-ISL formation and how it would be affected by the presence of a cofactor. Then, using splicing extracts we would examine the time to initiation of unwinding and the rate of U6-ISL formation during the activation phase of the spliceosome. The time to the initiation of unwinding would reflect the rearrangements that are necessary to proceed to the unwinding of the U4/U6 duplex and the rate of U6-ISL formation would reflect the impact of the full complement of regulatory cofactors on Brr2 helicase activity.

Single molecule analysis would be required due to the constraints of our in vitro splicing assay, for it has only been possible to study splicing in vitro with the use of yeast extracts. This poses a fundamental problem for ensemble FRET analysis, as the presence of flavins, NADP and other conjugated ring containing biomolecules results in the prohibitive autofluorescence of yeast extracts (Billinton and Knight 2001). Additionally, smFRET analysis allows asynchronous enzymatic events to be synchronized during the post-experimental data analysis phase. Both of these factors are pertinent to the study of the spliceosome, as in vitro splicing is notoriously inefficient and has been limited to end point enzymatic assays. Therefore, autofluorescence aside, the presumed molecular heterogeneity and asynchrony present in in vitro splicing

extracts would be averaged in ensemble FRET experiments to give a much less meaningful analysis of the conformational dynamics for any given step of the splicing cycle.

### Design

The Single molecule FRET system that we used is of the prism type Total Internal Reflection (TIR) type (Figure 3). We used a quartz slide that tethers RNAs via a Polyethylene Glycol (PEG) streptavidin-biotin linkage. The sample chamber consists of a coverslip sandwiched on top of the quartz slide through the use of double sided tape. Two holes are drilled in the cover slip, to allow for the exchange of reaction mixtures during the course of an experiment. Single molecule FRET trajectories are recorded with an electron-multiplying charge-coupled device (EM-CCD) camera at a resolution of ~ 10 ms. During the initial analysis of the data we would produce histograms for the range of smFRET states observed over the course of the experiment. This provides information about the different FRET states attainable by the sample and how well populated they are. For simple FRET traces we can extract dwell times in each state and determine the kinetics of interconversion. For more complex FRET traces we use hidden Markov modeling analysis to identify the kinetics of transitions between different smFRET states, and from this generate Transition Density Plots (TDPs) that graphically represent the distribution of different smFRET states and the kinetics for their interconversions (Joo et al. 2006; Roy et al. 2008).

My plan was to first characterize the transition of U4/U6\* to U6\* RNA, in both the presence and absence of Brr2. The T<sub>m</sub> for the natural U4/U6 hybrid is 53°C, and thus it would be expected to give a narrow distribution of low FRET values in the absence of

helicase activity (Brow and Guthrie 1988). I planned to measure the smFRET for cy3- and cy5-labeled U6 snRNA (U6\*) that had been immobilized on the quartz slide via a Polyethylene Glycol (PEG) streptavidin-biotin linkage. I also planned to measure the smFRET value of U6\* bound to U4 snRNA. To determine the baseline smFRET values for the U6\* and U4/U6\* molecules, the efficiency and dynamics of their smFRET was to be measured in the presence of Brr2 helicase reaction buffer and in yeast lysates in the absence of ATP.

To measure the U4/U6 unwinding in a system that contains the minimal number of components necessary to catalyze this reaction, I planned to use full length Brr2 protein that had been purified from yeast. SmFRET was to be used to monitor U6-ISL formation, as U4 is unwound from the surface of immobilized U4/U6\* substrates in reaction conditions that had already been established for the ensemble in vitro helicase reaction. I would also perform smFRET experiments with Brr2 in the presence of a C-terminal portion of prp8 that our lab has identified as having a stimulatory effect on Brr2's helicase activity (Figure 3). I would analyze the smFRET trajectories to produce histograms and TDPs in order to derive the heterogeneity of smFRET states in the sample and the kinetics for their transitions. This would provide us with kinetic reference points which would be representative of the minimal number of factors that can mediate the formation of the U6 U6-ISL and would indicate how the kinetics of formation are affected by the presence of a single stimulatory cofactor. This would also serve as reference point for understanding the smFRET of this transition in the more complex regulatory environment of the activation phase of the spliceosome.

Brr2 is required to unwind U4/U6 and we have identified one region of prp8 that can act as a cofactor, but during the activation of the spliceosome Brr2 activity is more heavily regulated. Brr2's helicase activity must be inhibited until it has joined the spliceosome and has formed the contacts that signify that it is appropriate for U6 to be released from U4 and integrated into the spliceosome. In this aim I would use splicing extracts and immobilized pre-mRNA to watch the recruitment to the pre-mRNA of the catalytically inert U4/U6\* duplex as it is found initially in the tri-snRNP. The initial FRET signal that we would see should be primarily low FRET, which would change once the spliceosome indicates to Brr2 that it can unwind U4/U6\*, at which point the FRET signal would increase and plateau when the U6-ISL is fully formed (Figure 4). With the development of our U6 smFRET approach in splicing extract we would be able to obtain the kinetic parameters that govern these different events and would thus be able to ask new questions about the factors that dictate these values.

I planned to make a yeast extract that has been depleted of the endogenous U6 snRNA through the use of RNase H-mediated oligonucleotide directed knockdown (Fabrizio and Abelson 1990). Then I would add U6\* to a concentration that is known to restore wild-type splicing levels. I would use this extract to prepare in vitro splicing reactions, and would flow this mixture into a TIR quartz slide chamber that has the yeast UBC4 pre-mRNA message attached to the surface of the slide through a biotin-streptavidin PEG linker. From work done in our lab and others we know that splicing can take place using similar experimental setups (Crawford et al. 2008). SmFRET trajectories would be recorded to produce histograms and TDPs. From the histogram, I would expect to see that the high FRET state, which corresponds to the formation of the

U6-ISL, would be well-populated in the presence of ATP, and that this would be poorly-populated in the absence of ATP.

In addition to comparing the rate at which maximum FRET is achieved in this system and the minimal system of Aim 1, it would be important to measure how long it takes the smFRET signal to begin to increase from the moment it first appears, as this would imply the time it takes for unwinding to initiate once the U4/U6 is recruited to the pre-mRNA. These two experimental parameters would be critical reporters for the different steps of the coupling process during spliceosome activation. The smFRET trajectories may contain FRET states not previously seen in Aim 1, and this could correspond to unexpected U6-ISL conformational changes at points beyond the activation point in the spliceosome assembly pathway. I would be able to control for this by performing an identical experiment with an extract that is made from a splicing mutant background, Prp2-1, which is required for the conversion of the activated spliceosome to the conformation that is capable of carrying out the first catalytic step, thus this experiment would be the reference point for maximal FRET. Lastly I would perform an experiment in a strain that has a mutated Prp28, which forms the spliceosome, but which is defective for the unwinding of the U4/U6 duplex, this would be the reference point for the low FRET value. From these experiments I would be able to define the time it takes for the initiation of U4/U6 unwinding to begin and the kinetics of U6\*-ISL formation in the context of the full complement of regulatory elements present in the spliceosome. If there is reversibility in the system then this would allow us to derive the equilibrium constant for at least one step in the activation of the spliceosome. As this experimental system is derived from yeast extract, through the

use of genetic perturbation I would then be able to analyze the determinants that dictate the observed kinetics of U6\* U6-ISL formation.

The spliceosome achieves fidelity through multiple mechanisms. One mechanism is through kinetic proofreading, wherein the rate of ATPase activity of the DExD/H box ATPases Prp16 and Prp22 has been described as a timer that discards incorrect substrates if they are unable to attain a conformation suitable for the next step in splicing before ATP is hydrolyzed (Burgess and Guthrie 1993; Mayas et al. 2006). Another mechanism of fidelity is achieved through the sequential rearrangements of snRNAs, wherein each state is coupled to the establishment of the previous one. One example of this is in the coupling of U4/U6 unwinding to the binding of U6 to the 5' splice site (Figure 1). Evidence for this coupling is supported by work that has shown that in spliceosomes where the U1/5' splice site is hyperstabilized, such that U6 cannot displace it, the U4/U6 duplex remains wound (Staley and Guthrie 1999). Likewise in mutants of Prp28, the DExD/H box ATPase that catalyzes the exchange of U1 for U6 at the 5' splice site, the U4/U6 duplex is maintained (Staley and Guthrie 1999).

Similarly a cold sensitive allele of U4, U4cs1, partially masks the region of U6 that interacts with the 5' splice site by extending its interactions with U6. This results in the formation of an aberrant spliceosome that contains U1 and the tri-snRNP and the U4/U6 duplex in a wound form (Li and Brow 1996). Interestingly, suppressors of the U4cs1 allele were found in Prp8, including Prp8-802 and others (Kuhn et al. 1999; Kuhn and Brow 2000). The authors hypothesized that because Prp8 is known to crosslink with the 5' splice site (Reyes et al. 1999) and have direct and genetic interactions with Brr2, that Prp8 serves as a proofreader of the U6/5' splice site interaction and that the



suppressor they identified bypasses this operation. A recent crystal structure of Prp8 lends further credence to this assertion. The Prp8 binding site for the 5' splice site was modeled based upon extensive homology to a RNase H domain, for which an RNA bound structure existed. The authors found that the suppressors of U4cs1 mapped to this region. They surmise that the basis for coupling of U6 5' splice site recognition to U4/U6 unwinding relies upon the proper positioning of the U4/U6 duplex for Brr2 unwinding after the U6 5' splice site duplex is formed. In the case of the U4cs1, the U4/U6 duplex is not positioned correctly so Brr2 cannot act on its substrate. Another hypothesis is that recognition of the U6 5' splice site duplex produces an allosteric switch that changes the regulation of Brr2 activity from repression to activation, that is Snu114's GDP is switched for GTP and ubiquitin is removed from Prp8. Additional suppressors of U4cs1 were identified in U4 and U6, each of which either destabilize the U4/U6 duplex (U4-G14C) or stabilize the U6-ISL (U6-A62U/C85A) (Kuhn et al. 1999), suggesting that the balance between the U4/U6 duplex and free U6 is delicately maintained by repressive factors and the inherent strengths of the competing snRNA structures.

What emerges is a two-state model for the activation of the spliceosome, wherein there is a transition from the catalytically inert U4/U6 duplex to the free U6, which is then able to bind to U2 and to form the U6-ISL, the necessary rearrangements for the first catalytic step. Prp8 may serve as a proofreader to inhibit this transition until U6 is properly bound at the 5' splice site. Previous work that addresses this issue was done with genetics and with native gels of isolated spliceosomes to report on the status of the U4/U6 duplex. To revisit this work, but now from the perspective afforded by smFRET,

would allow one to determine the timing of the two-state model and the factors which dictate this switch. The role of Prp8 as a proofreader would be determined by examining its contribution to regulating the time to the initiation of unwinding and/or the rate of unwinding. I would first measure smFRET from U4cs1 extracts that have been reconstituted with U6\*. Although this is a cs allele and the smFRET experiments would be done at room temperature, the time to the start of unwinding should be dramatically slowed relative to wild type splicing extracts but the effect of the U4cs1 allele upon the rate of unwinding is less clear. I would then attempt to rescue this slowed phenotype by pairing the U4cs1 allele with the suppressor Prp8-201. In this experiment the smFRET data should look more like that which is observed in the wild type splicing extracts. Yet, it would be interesting to examine which aspect of the unwinding is affected. Is it the time that it take for unwinding to initiate, or is it the rate of unwinding? If Prp8 is acting as a scaffold for the repressors of Brr2, could these be overcome with the addition of excess GTP or deubiquitinase enzymes in the context of the frozen U4cs1 spliceosomes. Another experiment would be to compare these smFRET data to the smFRET data extracted from the U6 and U4 mutants that suppress the U4cs1 mutation which presumably shift the equilibrium of the competing structures, ie. the bound and unbound states of U6. Again, would it be the time to unwinding or the rate of unwinding that is affected? Only a smFRET approach would allow us the resolution to answer these questions.

## Results

The foundation for using smFRET to study the spliceosome was established by John Abelson and others in our lab, in collaboration with Nils Walter's Laboratory at the

University of Michigan. This work focused on monitoring the splicing cycle with smFRET by placing probes in the 3' and 5' exons of the UBC4 pre-mRNA, with the expectation that as the pre-mRNA is spliced to produce the mature form, an increase in smFRET would be observed. John Abelson identified the UBC4 transcript as an ideal smFRET pre-mRNA for in vitro splicing based upon its high splicing efficiency. The initial analysis indicated that in the absence of ATP the UBC4 pre-mRNA populated high and low FRET states and that the forward and reverse rates were the same, suggesting reversibility and that the equilibrium constant was one. It was only upon the addition of ATP that a bias towards one direction was achieved. Important control experiments with mutations in the pre-mRNA that are known to block splicing also inhibited the appearance of the high FRET states observed in the presence of ATP for the wild type transcript, confirming that these values are dependent upon productive splicing.

In an alternative approach, I have chosen to use smFRET to obtain kinetic and conformational information on the formation of the well-characterized U6-ISL, which is one of the major rearrangements that take place during the activation of the spliceosome. I have produced a synthetic U6 snRNA that has cy3 and cy5 fluorescent probes in the base of the U6-ISL. From ensemble fluorimetric measurements I observe that there is low FRET efficiency in the extended conformation when bound to the U4 snRNA and there is high FRET efficiency when the U4 snRNA is removed allowing the U6-ISL to form. A similar trend is expected when U6 is bound to the U2 snRNA as the U6-ISL is present in the U2/U6 duplex (Figure 2). I have also determined that the cy3 and cy5 labeled U6 snRNA (U6\*) is functional, by showing that it can restore in vitro splicing to extracts that have been depleted of their wild type U6 (Figure 5).

With the characterization of this tool in ensemble experiments, I have demonstrated that these probes report faithfully on the predicted structures for U6 in the two conformations relevant to distinct stages during the activation of the spliceosome. I have also shown that the location of these probes is well tolerated by the spliceosome. From John Abelson's work and from others we know that the spliceosome can be recruited to the quartz surface via a tethered pre-mRNA and that splicing can take place. Therefore I believe that the experimental approach I have chosen is a meaningful step towards an smFRET analysis of the spliceosome during its activation (Figure 1).

## Figures

Figure 1: Coupling of U1 displacement and U6 5' splice site recognition to U4/U6 unwinding and the formation of the activated spliceosome

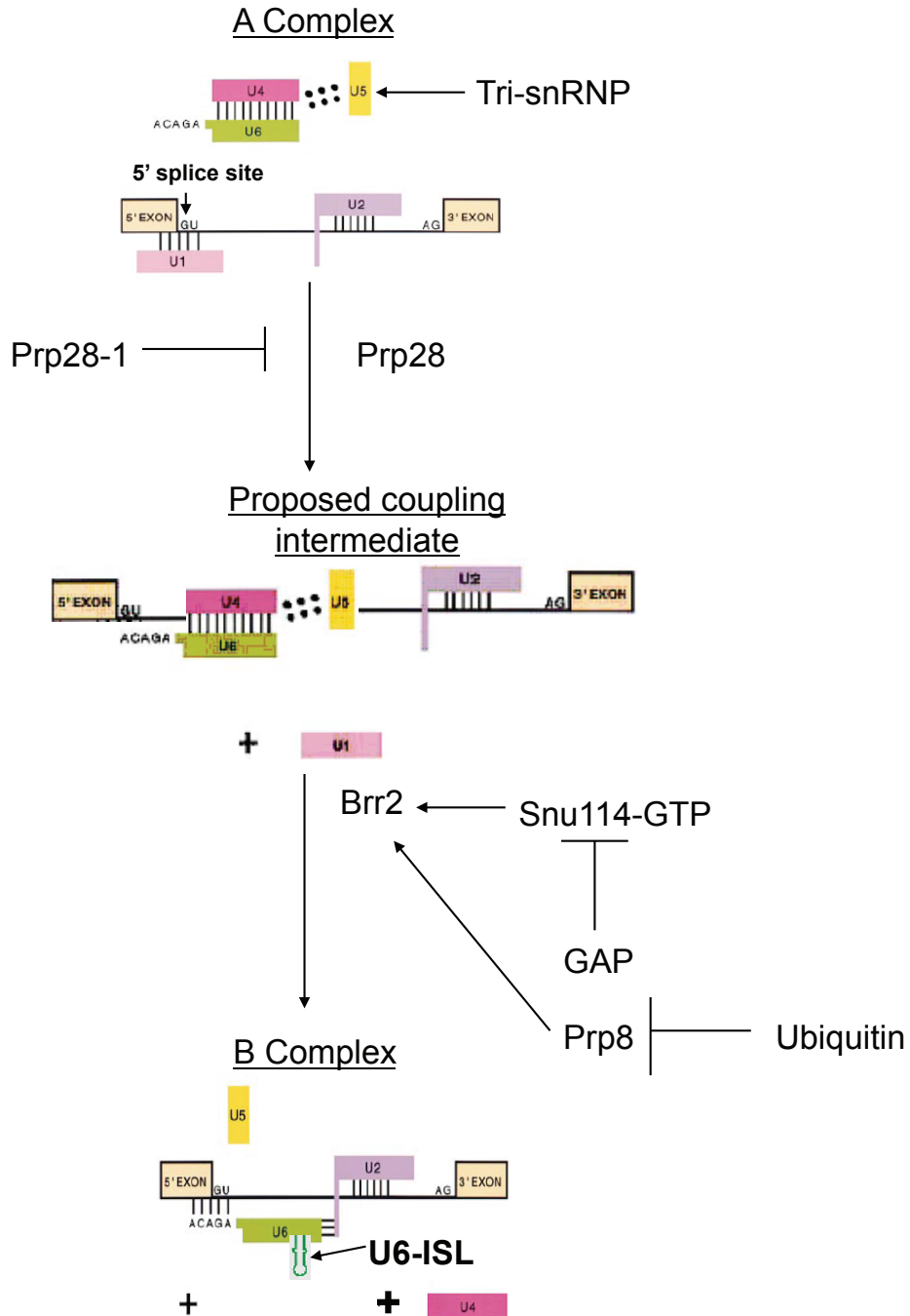
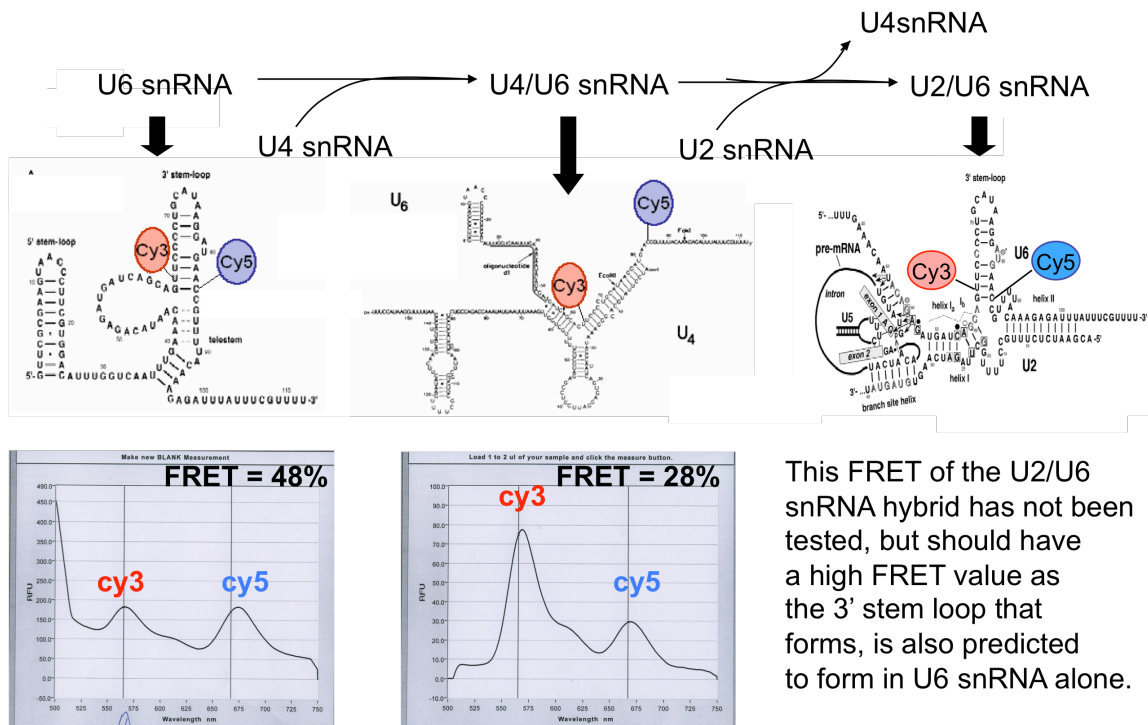


Figure 2: Dynamics of U6 snRNA With Predicted Conformations and the Corresponding FRET Values Obtained From Synthesized Cy3 and Cy5 Labeled U6 snRNA (U6\*), Alone and in Complex with U4 snRNA



This FRET of the U2/U6 snRNA hybrid has not been tested, but should have a high FRET value as the 3' stem loop that forms, is also predicted to form in U6 snRNA alone.

Figure 3: Experimental design for Aim 1. The first experiment will be to examine U6\* and U4/U6\* smFRET values alone. The second experiment will be to measure the FRET change that occurs when Brr2 unwinds the U4/U6\* duplex, alone or in the presence of the stimulatory cofactor Prp8-CTF.

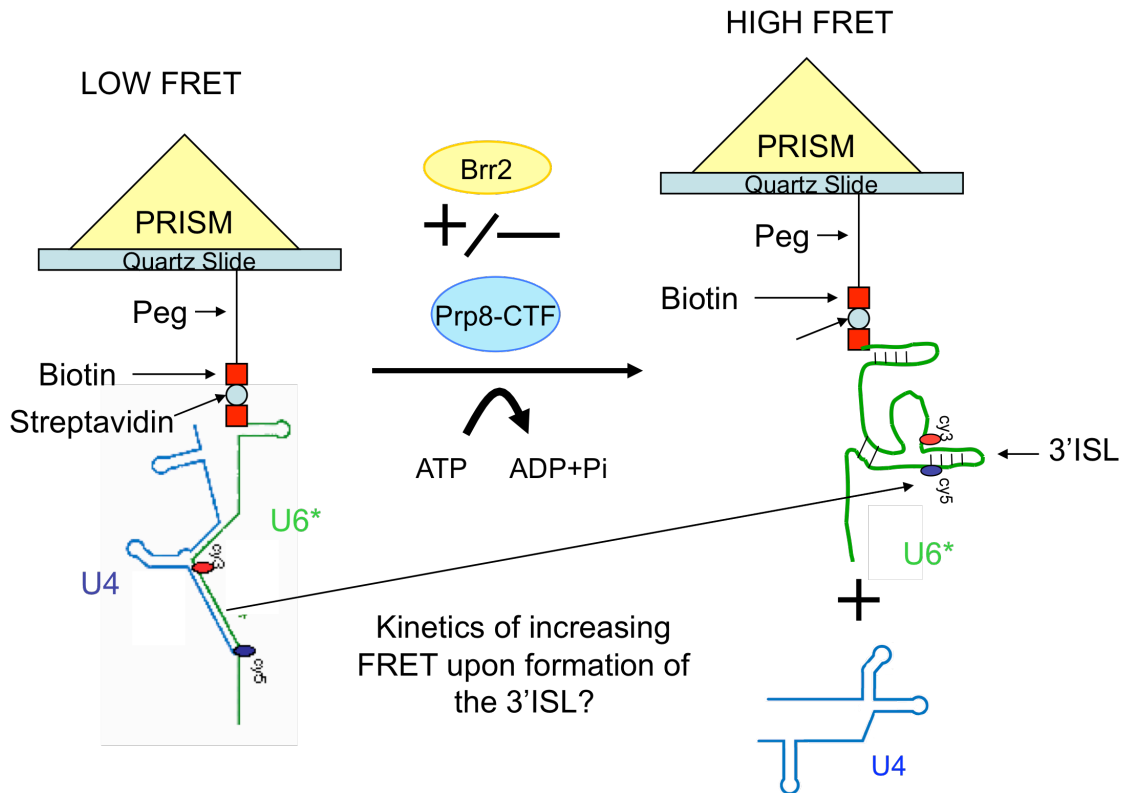


Figure 4: The expected FRET values from U6\* in cell lysate as predicted by its involvement in distinct steps of the assembly process

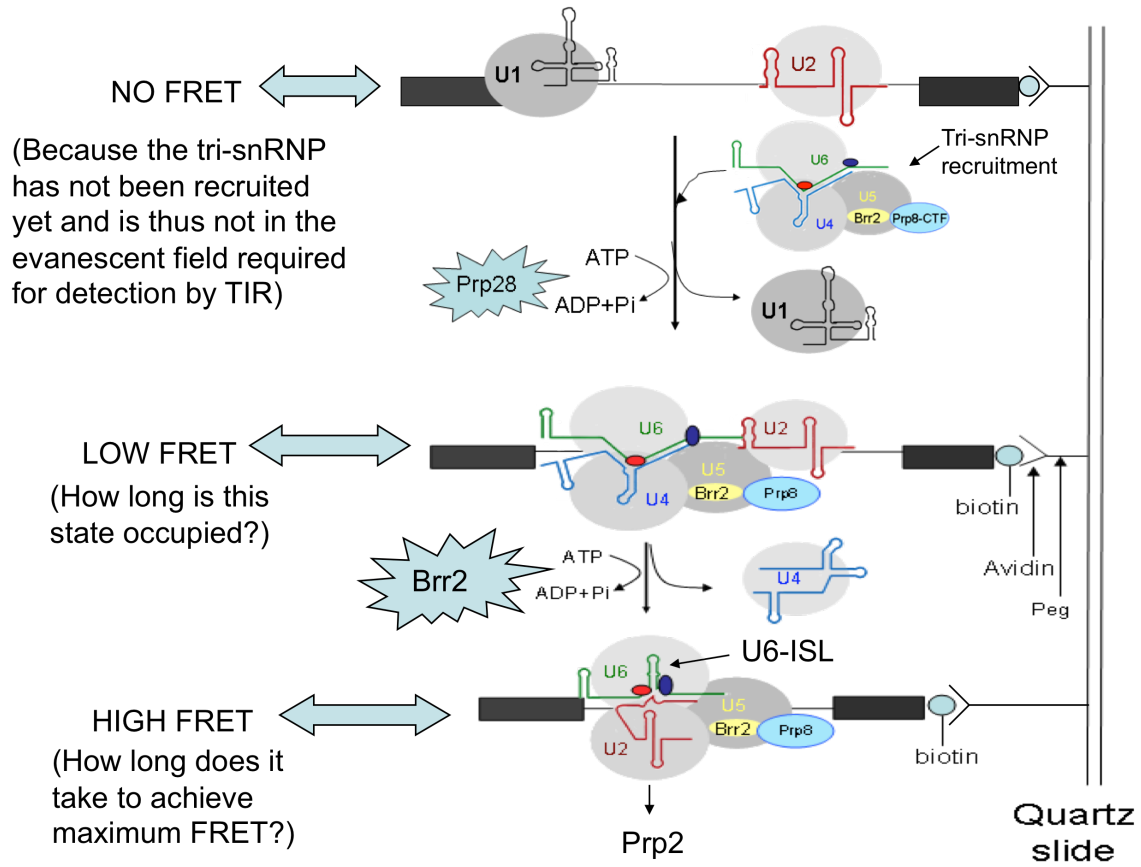
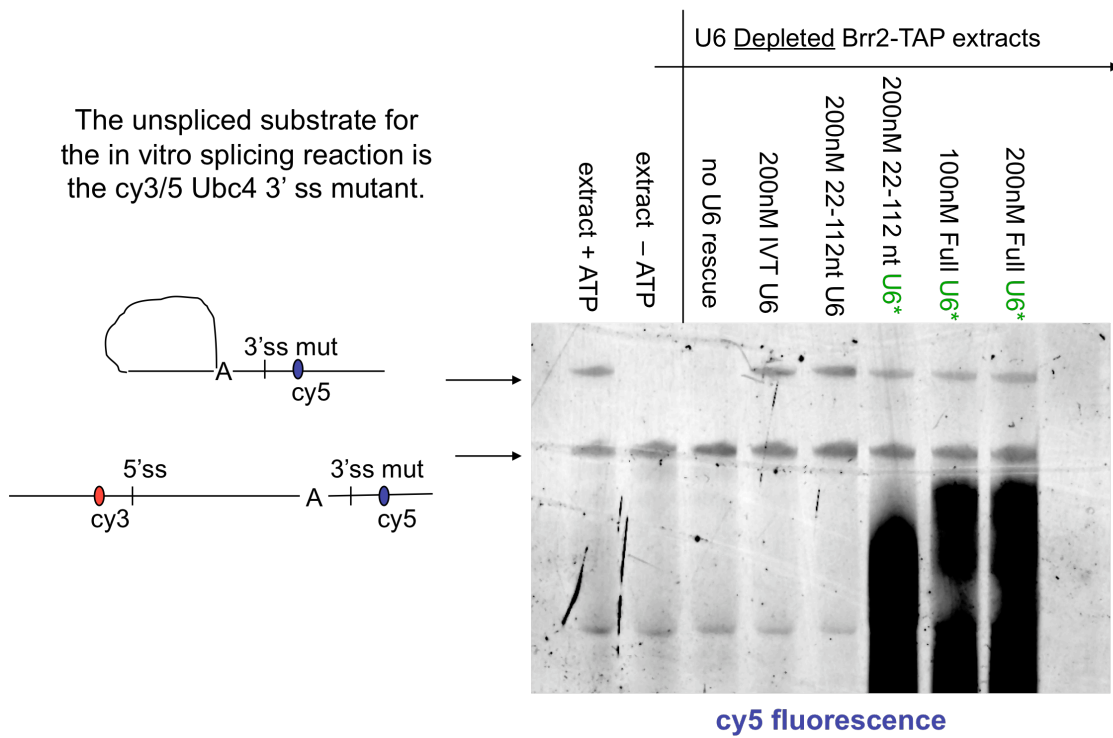




Figure 5: Cy3/5 U6 (U6\*) snRNA Reconstitutes Splicing in U6 Depleted Brr2-TAP extracts



## References

- Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. 2008. A role for ubiquitin in the spliceosome assembly pathway. *Nat Struct Mol Biol* 15(5): 444-451.
- Billinton, N. and Knight, A.W. 2001. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Anal Biochem* 291(2): 175-197.
- Brow, D.A. and Guthrie, C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334(6179): 213-218.
- Burgess, S.M. and Guthrie, C. 1993. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* 73(7): 1377-1391.
- Crawford, D.J., Hoskins, A.A., Friedman, L.J., Gelles, J., and Moore, M.J. 2008. Visualizing the splicing of single pre-mRNA molecules in whole cell extract. *RNA* 14(1): 170-179.
- Fabrizio, P. and Abelson, J. 1990. Point mutations in yeast U6 snRNA can specifically block the first or second step of pre-mRNA splicing in vitro. *Mol Biol Rep* 14(2-3): 135.
- Joo, C., McKinney, S.A., Nakamura, M., Rasnik, I., Myong, S., and Ha, T. 2006. Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* 126(3): 515-527.

- Kim, D.H. and Rossi, J.J. 1999. The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex. *RNA* 5(7): 959-971.
- Kuhn, A.N. and Brow, D.A. 2000. Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* 155(4): 1667-1682.
- Kuhn, A.N., Li, Z., and Brow, D.A. 1999. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol Cell* 3(1): 65-75.
- Kuhn, A.N., Reichl, E.M., and Brow, D.A. 2002. Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *Proc Natl Acad Sci U S A* 99(14): 9145-9149.
- Laggerbauer, B., Achsel, T., and Luhrmann, R. 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc Natl Acad Sci U S A* 95(8): 4188-4192.
- Li, Z. and Brow, D.A. 1996. A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element in vivo. *RNA* 2(9): 879-894.
- Maeder, C., Kutach, A.K., and Guthrie, C. 2009. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat Struct Mol Biol* 16(1): 42-48.
- Mayas, R.M., Maita, H., and Staley, J.P. 2006. Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nat Struct Mol Biol* 13(6): 482-490.

- Raghunathan, P.L. and Guthrie, C. 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol* 8(15): 847-855.
- Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. 1999. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* 5(2): 167-179.
- Roy, R., Hohng, S., and Ha, T. 2008. A practical guide to single-molecule FRET. *Nat Methods* 5(6): 507-516.
- Silverman, E., Edwalds-Gilbert, G., and Lin, R.J. 2003. DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene* 312: 1-16.
- Small, E.C., Leggett, S.R., Winans, A.A., and Staley, J.P. 2006. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol Cell* 23(3): 389-399.
- Staley, J.P. and Guthrie, C. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3(1): 55-64.
- van Nues, R.W. and Beggs, J.D. 2001. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* 157(4): 1451-1467.

### **3. The crystal structure of *S. cerevisiae* Sad1, a catalytically inactive deubiquitinase that is broadly required for pre-mRNA splicing**

#### Motivation

Sad1 is required for the assembly of the U4 and U6 snRNAs into the di-snRNP particle. Named for its snRNP assembly defect, the *sad1-1* temperature sensitive (ts) yeast strain is unable to assemble newly transcribed U4 snRNAs into the di-snRNP particle at the restrictive temperature, a phenotype also observed for the known snRNP assembly factor mutants *prp19-1* and *prp24-1* (Lygerou et al. 1999). Inactivation of Sad1 was shown to block the splicing of several pre-mRNA substrates, however it was unclear if Sad1 acted as a core splicing factor in yeast. Sad1 did not associate stably with any of the individual yeast snRNAs (Lygerou et al. 1999), but was detected in the penta-snRNP, a mature form of the spliceosome that contains all 5 snRNPs and can only be purified under low salt conditions (Stevens et al. 2002). The human homologue of Sad1, USP39, weakly associates with the tri-snRNP snRNAs under low salt conditions and was shown to be required for the stable association of the tri-snRNP to the pre-spliceosome (Makarova et al. 2001). Thus, Sad1 may be required for tri-snRNP recruitment to the pre-spliceosome in addition to its contributions to di-snRNP assembly. How Sad1 mediates these snRNP complex formations remains unknown.

Sad1 belongs to a group of spliceosomal proteins that contain ubiquitin-related domains (Korneta et al. 2012). Prp19 is a U-box ubiquitin ligase required for spliceosome activation and recycling (Chan et al. 2003). The tri-snRNP protein, Prp8, contains an inactive Jab1/MPN deubiquitinase domain in its C-terminus that has been linked to the regulation of U4/U6 snRNA duplex separation (Bellare et al. 2006; Maeder

et al. 2009a). Finally, the presence of ubiquitin in the tri-snRNP has been linked to both human and yeast tri-snRNP stability (Bellare et al. 2008; Song et al. 2010). The N-terminal domain of Sad1 is homologous to a zinc finger ubiquitin binding protein (ZnF-UBP) domain and the C-terminal domain is homologous to a ubiquitin-specific peptidase (USP) domain. Not all members of the ZnF-UBP domain family are competent to bind ubiquitin. For example, USP5 has two ZnF-UBP domains, one of which has been shown to bind the free C-terminal tail of ubiquitin with high affinity (Reyes-Turcu et al. 2006; Reyes-Turcu et al. 2008; Avvakumov et al. 2012). In contrast, the ZnF-UBP domain from Ubp8 does not bind ubiquitin, but rather it functions as a scaffold for co-factors, which then activate the USP domain of Ubp8 (Kohler et al. 2010; Samara et al. 2010; Samara et al. 2012). Sad1's USP domain (iUSP) is predicted to be catalytically inactive, as it is missing the essential nucleophilic cysteine in its USP catalytic triad. It is unclear whether this deviation would affect ubiquitin binding, or if Sad1's ZnF-UBP domain might be required to activate ubiquitin binding function.

To examine the mechanism of Sad1 function and its role in the regulation of the splicing reaction, we first analyzed *sad1-1* inactivation with splicing microarrays and found that, like other core splicing factors, Sad1 inactivation *in vivo* caused a strong block to most pre-mRNA splicing events. We then produced full-length recombinant Sad1 protein and showed that it could rescue *in vitro* splicing activity in *sad1-1* inactivated extracts. We observed that tri-snRNP joining to the pre-spliceosome was blocked in *sad1-1* temperature sensitive (ts) inactivated extracts, and that this defect could be rescued by the addition of recombinant Sad1. Finally, to gain insight into Sad1's function, we crystallized and solved the structure of full-length Sad1. We show

that Sad1's ZnF-UBP is in a conformation that is incompatible with ubiquitin binding. In addition, we observed that Sad1's iUSP domain preserves many of the features of active USP domains, but possesses a remodeled catalytic cleft. Together these results help us to interpret Sad1's role in splicing and raises questions for the further exploration of its molecular function.

## Results

### **Sad1 is composed of two ubiquitin related domains**

Sad1 is a member of the ZnF-UBP USP family of deubiquitinases. The ZnF-UBP domain is located at the N-terminus (amino acids 29 to 121) and the iUSP domain is located at the C-terminus (amino acids 148-448) (Fig. 1A). Although the iUSP domain has many of the residues conserved in active USPs, it and its homologues are predicted to be inactive. Whereas an active USP catalytic triad is CHN/D (Komander et al. 2009), Sad1 has DHD and its human homologue, USP39, has DSQ (Fig. 1A; Supplemental Fig. S1). Sad1's two domains are well-conserved amongst its homologues, with the exception of an additional N-terminal RS domain found in the human homologue (Fig. 1A; Supplemental Fig. S1). A 25 amino acid linker connects the N and C-terminal domains of Sad1.

### **Sad1 is a general splicing factor**

Sad1 appears to have a weak affinity for the spliceosome and it was unknown whether inactivating *Sad1 in vivo* would have a broad, genome-wide splicing defect, as is observed upon the inactivation of essential spliceosome factors (Pleiss et al. 2007).

We used pre-mRNA splicing microarrays to assess the breadth of Sad1's role in splicing after competitively hybridizing total mRNA from ts-inactivated *sad1-1* cells and *SAD1* cells treated under the same conditions. We cloned the *sad1-1* ts allele and *SAD1* into CEN/ARS vectors that can complement a *SAD1* deletion in yeast. The *sad1-1* ts mutation had not been mapped, and upon sequencing we discovered that it contains two mutations, L236P and S276F. These mutations likely destabilize the protein at 37 °C, as attempts to recombinantly express and purify the double mutant Sad1-1 protein yielded insoluble protein (data not shown), whereas the wild-type Sad1 protein was well expressed and soluble. Both CEN/ARS plasmids were able to complement the *SAD1* deletion at 30 °C, but only the wild-type plasmid was able to complement the *SAD1* deletion at 37 °C (Table 1). Splicing microarray analysis revealed a strong and broad block to pre-mRNA splicing in *sad1-1* after incubation at 37 °C for 30 minutes, as can be observed by the dramatic accumulation of intron signal for many pre-mRNAs in the *sad1-1* versus *SAD1* strain (Fig. 1B, middle lane, labeled "P" for pre-mRNA, represents the intron probes). The magnitude and breadth of this splicing defect is reminiscent of the inactivation of the core splicing factors, Prp5 and Prp8 (Pleiss et al. 2007). Thus *Sad1* plays a crucial role in pre-mRNA splicing.

To investigate the mechanism of *SAD1*'s function, we used an *in vitro* splicing assay (Lin et al. 1985). Upon the addition of a Cy5 fluorescently labeled *ACT1* pre-mRNA and ATP the spliceosome assembles and then catalyzes two chemical steps resulting in the removal of the intron and the ligation of the exons. The intermediates and the product of the splicing reaction are resolved by denaturing gel electrophoresis and visualized by a fluorimeter. Consistent with initial observations, splicing was



slightly inhibited by *sad1-1* extracts grown at 30 °C (Fig. 1C, compare lanes 2 and 4) (Lygerou et al. 1999), whereas splicing was entirely inhibited by heating the extracts to 37 °C for 10 minutes (Fig. 1C, lane 3). The *sad1-1* heat inactivated lysate was unable to complete the first chemical step of splicing of the *ACT1* pre-mRNA, as evidenced by the absence of splicing reaction intermediates (Fig. 1C, compare lanes 2 and 3). In order to further determine the role of Sad1 in the splicing reaction, we purified full-length recombinant Sad1 with an N-terminal 6-His tag, and found that it was able to fully rescue *in vitro* splicing activity (Fig. 1C).

### **Sad1 is required for the stable recruitment of the tri-snRNP**

Since *sad1-1* inactivated extracts were blocked at the first catalytic step, we sought to distinguish between a catalytic defect and a defect in the assembly of the spliceosome. Spliceosome assembly can be monitored by incubating extracts with a labeled pre-mRNA substrate and resolving the pre-mRNA bound spliceosome from its assembly intermediates by native gel electrophoresis (Cheng and Abelson 1987). Deleted two sentences As there was a low amount of ATP in both of our extracts, we saw a small amount of the pre-spliceosome assemble in the absence of exogenous ATP (Fig. 1D, compare lanes 1 and 5). After the addition of ATP to 2 mM to the wild-type extracts the pre-spliceosome band disappeared as the tri-snRNP joined the pre-spliceosome stably and a slower migrating band consisting of fully assembled spliceosomes appeared (Fig. 1D, lane 2). Depleting the U6 snRNA by anti-U6 oligo-directed RNase H digestion caused the tri-snRNP to fall apart and, as a result, even in the presence of 2 mM ATP, only the pre-spliceosome band is formed (Fig. 1C lane 3) (Fabrizio et al. 1989). Upon the addition of *in vitro* transcribed U6 snRNA, the tri-snRNP

reassembled and much of the pre-spliceosome matured to the fully assembled spliceosome (Fig. 1D, compare lanes 4 and 3).

To determine the role of Sad1 in spliceosome assembly we made extracts from *sad1-1* cells grown at 30 °C. As with the wild-type extract, in the absence of exogenous ATP, a small amount of the pre-spliceosome band formed (Fig. 1D, lane 5). Unlike the wild-type extract, upon the addition of 2 mM ATP, the *sad1-1* assembled spliceosome band was barely detectable and the pre-spliceosome band had only a slight decrease in intensity, reflecting the weakened capacity for this *sad1-1* to splice at 30 °C (Fig. 1D, lane 6 and Fig 1C, lane 2). Heat-shocking the *sad1-1* extracts for 10 minutes at 37°C led to a large accumulation of the pre-spliceosome species after the addition of 2 mM ATP, consistent with the total block to splicing that we observe by *in vitro* splicing (Fig. 1D, lane 8 and Fig. 2C, lane 3). This result looked similar to the U6 snRNA knockdown in wild-type extracts (Fig. 1D, lane 3) and addition of recombinant Sad1 to *sad1-1* lysates grown at 30 °C and, or heat-shocked *sad1-1* rescued these defects (Fig. 1D, lane 7 and lane 9). Therefore yeast Sad1 is required for the stable integration of the tri-snRNP into the pre-spliceosome to form the fully assembled spliceosome, similar to what was observed for USP39 (Makarova et al. 2001). Sad1 could achieve this by either interacting with the pre-spliceosome, the tri-snRNP, or both. Additionally Sad1 could be required to maintain the stability of the tri-snRNP, as it was originally identified as a di-snRNP assembly factor (Lygerou et al. 1999). As connections between ubiquitin and the tri-snRNP had been established previously (Bellare et al. 2008; Song et al. 2010) we were intrigued by the possibility that Sad1 might somehow employ its ubiquitin-related domains to affect tri-snRNP joining to the pre-spliceosome.

## **The two ubiquitin-related domains of Sad1 do not bind ubiquitin *in vitro***

Sad1 is composed of two domains that have the potential to bind ubiquitin (Fig. 1A). We found that both domains are required for Sad1's function *in vivo*, as each domain by itself is unable to complement a *sad1* deletion (Table 1). To determine whether Sad1 binds to ubiquitin, we developed an Isothermal Titration Calorimetry (ITC) assay to measure the interaction between Sad1 and ubiquitin. For comparison we used the ZnF-UBP from USP5, known to bind the C-terminal tail of ubiquitin through a WRY motif that is well conserved amongst the ZnF-UBP domains that have been shown to bind ubiquitin (Fig. 2A) (Reyes-Turcu et al. 2006; Bonnet et al. 2008; Ouyang et al. 2011). In contrast, the ZnF-UBP domains that do not bind ubiquitin have substitutions in the WRY motif (Allen and Bycroft 2007; Zhang et al. 2011), as does Sad1, which occur as YRL (Fig. 2A).

Using ITC, we were unable to detect an interaction between full-length Sad1 and ubiquitin. Titrating 1 mM ubiquitin into either 50  $\mu$ M Sad1 or buffer produces no observable response as measured by ITC (Fig. 2A). The ZnF-UBP from USP5 produced a dissociation constant of 2.4  $\mu$ M, consistent with previously published values (Fig. 2B) (Reyes-Turcu et al. 2006). In addition the iUSP domain alone of Sad1 does not bind to ubiquitin, as it too did not produce a measurable ITC response (Fig. 2B). The ZnF-UBP domain alone of Sad1 could not be recombinantly purified and was thus unsuitable for ITC analysis. As ITC only measures the enthalpy between two reactants, there was a possibility that Sad1 bound to ubiquitin with a purely entropic based interaction. To address this possibility we used an alternative method to investigate the potential binding between Sad1 and ubiquitin, by performing biolayer interferometry with

immobilized ubiquitin and USP5 ZnF-UBP, or Sad1 (Supplemental Fig. S2). Consistent with the ITC results, USP5 ZnF-UBP produced a robust response with a  $K_D$  similar to that previously observed and there was no specific response for Sad1.

We also used GST immunoprecipitations to investigate the possibility of an interaction between Sad1 and ubiquitin or the yeast ubiquitin-like proteins. We used the yeast pre-mRNA splicing factor Snu66, which has been shown to interact with Hub1 (Mishra et al. 2011) as a positive control. When purified N-terminal GST fusions of ubiquitin, SUMO, Nedd8, and Hub1 were incubated with either Sad1, ZnF-UBP USP5, or Snu66, the only observed interactions were between GST-Ubiquitin and the ZnF-UBP of USP5, and GST-Hub1 and Snu66 (Supplemental Fig. S3). As this absence of ubiquitin binding for Sad1 was contrary to our expectations, and given Sad1's critical role in splicing, we pursued a structural approach to gain further insights into Sad1 function.

### **The crystal structure of full-length SAD1**

Full-length Sad1 fused to a N-terminal 6-His tag was recombinantly expressed and purified to homogeneity from *E. coli* and was crystallized by hanging drop vapor diffusion methods. The crystal structure was solved by single anomalous diffraction using a tungsten derivative (Table 2). The structure of Sad1 was refined to 1.87 Å resolution with  $R_{work}$  of 17.1%,  $R_{free}$  of 20.2%, and good stereochemistry. The phased electron density maps allowed for the unambiguous placement of 414 of 448 residues of the full-length Sad1. The N-terminal 28 amino acids were disordered in the crystal and absent from the experimental map. The tips of the USP fingers sub-domain were also

disordered (Fig. 3A), which is a feature that has been observed in other USPs crystallized in the absence of ubiquitin (JCSG 2004; Kohler et al. 2010; Samara et al. 2010). Additionally, a loop in the pseudo-active site of Sad1's USP domain is disordered. As mentioned previously the *sad1-1* ts mutation, L236P and S276F, likely destabilize the protein at 37 °C. The L236P mutation introduces a proline in the middle of a central  $\alpha$ -helix of the iUSP domain and S276F introduces a hydrophobic residue in the solvent exposed tips of the iUSP fingers domain, both of which would be likely to challenge the stability of the Sad1 iUSP domain.

The orientation of Sad1's ZnF-UBP domain to its iUSP domain is specified by a 25 amino acid linker, composed of a short helix and two short anti-parallel  $\beta$ -strands (Fig. 3A). The linker sequence, amino acids 122 to 147, is not well conserved amongst Sad1 homologues. The length of Sad1's linker is only three amino acids shorter in the *S. pombe* and human homologues, suggesting some conservation in their inter-domain orientations (Supplemental Fig. S1). The linker is structurally distinct from the linker observed in the structure of Ubp8, another ZnF-UBP USP in the PDB, which is longer and composed of multiple short  $\alpha$ -helices (Kohler et al. 2010; Samara et al. 2010). Sad1 K199 from the iUSP domain makes hydrogen bonds to D43 from the ZnF-UBP domain. Sad1 K207 from the iUSP domain connects the two domains with a hydrogen bond to D41 from the ZnF-UBP domain (Fig. 3B). Lastly, inter-domain contacts are further specified through hydrophobic residues that are within van der Waals distances of each other. F206 from the iUSP domain and L140 from the linker domain interact with a hydrophobic patch from the ZnF-UBP domain, which is composed of the aliphatic

part of the conserved R37 side chain and the side chains of L40 and I117 residues (Fig. 3C)

### **The Zinc Finger-Ubiquitin Binding Domain of Sad1**

The ZnF-UBP domain from Sad1 shares the same (approximately 95 amino acid) mixed a-b core structure found in other ZnF-UBP proteins, possessing a 5-stranded anti-parallel b-sheet, flanked by two a-helices. Table 2 lists the RMSDs for these ZnF-UBPs aligned to Sad1's ZnF-UBP. The first a-helix,  $\alpha$ A, is connected to the end of the second b-strand through a loop, Loop L2A (Reyes-Turcu et al. 2006), that displays varied orientations in the different ZnF-UBP structures (Fig. 4A). The second a-helix,  $\alpha$ B, is held closely to the opposite side of the b-sheet wall through a network of hydrophobic residues.

ZnF-UBPs can bind up to 3 zinc atoms, as is the case with the ZnF-UBP from USP16 (Fig. 4B) (Pai et al. 2007). Unlike USP16, Sad1's ZnF-UBP only binds a single zinc atom at the B site through a CCHH zinc finger, which serves to connect the first and second b-strand to the end of the first a-helix (Fig. 4B and 4C). Sad1's B site CCHH zinc-binding finger is conserved amongst all of Sad1's splicing homologues and aligns well with the B site CCHH zinc fingers from the ZnF-UBP domains of USP16 and HDAC6 (Fig. 4C; Supplemental Fig. S1).

The B site CCHH finger may be generally important for the stability of ZnF-UBP domains. Alanine substitutions at any of the CCHH zinc-binding residues of Sad1's B site results in aggregated protein as determined by size exclusion chromatography (data not shown) and the Sad1 ACHH and the CCHA substitutions were lethal *in vivo* (Table 1). Every ZnF-UBP from human and yeast contains a B site CCHH zinc finger,

with the sole exception being USP5, which possesses residues that are engaged in compensatory interactions (Reyes-Turcu et al. 2006; Bonnet et al. 2008) (Supplemental Fig. S4).

**Sad1 does not possess a canonical ubiquitin binding site.**

Comparison of Sad1's ZnF-UBP primary sequence with those ZnF-UBPs that bind ubiquitin provided little insight into why ubiquitin binding was not observed for Sad1 (Fig. 2A). Inspection of the available ubiquitin-bound ZnF-UBP crystal structures in the PDB shows that the ubiquitin binding site lies between the L2A loop and the loop that connects the fourth and fifth  $\beta$ -strands, the  $\beta$ 4-5 loop (Reyes-Turcu et al. 2006; Dong 2008). Although overall RMSDs for Sad1's ZnF-UBP with other ZnF-UBPs in the PDB are low (Table 2), a structural alignment of Sad1's ZnF-UBP with the ubiquitin binding ZnF-UBPs from USP5 and HDAC6 reveals notable differences in the conformations of their L2A and  $\beta$ 4-5 loops. In Sad1, these loops are positioned closer to one another relative to the USP5 and HDAC6 ZnF-UBP structures (Fig. 5A). Additionally, the closed conformation of Sad1 does not align well with any of the ZnF-UBP structures available in the PDB (Fig. 5B).

As a result of the contraction between the L2A and  $\beta$ 4-5 loops in Sad1, homologous residues that would be used for ubiquitin recognition in USP5 are in a conformation that blocks ubiquitin binding by Sad1. Ubiquitin binding by USP5's ZnF-UBP is dependent upon the presence of a group of residues that are conserved in those ZnF-UBPs competent to bind ubiquitin (Fig. 2A) (Reyes-Turcu et al. 2006). R221 from USP5's loop L2A forms two hydrogen bonds to the main-chain carbonyl oxygen of ubiquitin G75 (Fig. 5D). Sad1's ZnF-UBP also has an arginine at this position, Sad1

R70 (Fig. 2A). However, instead of pointing towards solvent to make itself available for interaction with ubiquitin, the crystal structure of Sad1 reveals that R70 packs against the surface of Sad1's ZnF-UBP (Fig. 5C and D). This conformation of Sad1 R70 is stabilized by hydrogen bonds to Y58 from the first  $\beta$ -strand and N102 from the  $\beta$ 4-5 loop (Fig. 5C). The homologous residue of Sad1 Y58 corresponds to W209 of USP5, which interacts with ubiquitin and is conserved amongst ZnF-UBPs that bind ubiquitin (Fig. 2A). Although the positions of the planar rings of Sad1 Y58 and USP5 W209 align well, Sad1 Y58 introduces a hydroxyl group that functions as a hydrogen bond acceptor for the terminal amine of Sad1 R70 (Fig. 5C). Sad1 R70 is further stabilized through a second hydrogen bond to N102 (Fig. 5C). Sad1 N102 corresponds to D264 in USP5, which in USP5's ZnF-UBP points towards the solvent to interact with the terminal amines of ubiquitin R72 and R74 (Fig. 5C; Fig. 5D) (Reyes-Turcu et al. 2006). Aligning Sad1 to the crystal structure of USP5 bound to ubiquitin shows that, as a result of the unique stabilization of Sad1's R70, the R70 arginine side chain lies directly in the path of where ubiquitin would be expected to bind (Fig. 5D).

The more contracted conformation of Sad1's L2A and  $\beta$ 4-5 loops and R70's blockage of the ubiquitin binding path also results in an altered ZnF-UBP surface. Whereas the USP5 ZnF-UBP ubiquitin binding site appears as a prominent cavity on the surface of both the apo and the ubiquitin bound structures (Fig. 5E), the same area on Sad1, which corresponds to the space between the L2A and  $\beta$ 4-5 loops, is entirely filled by the contraction of these loops and the placement of Sad1's R70 side chain (Fig. 5E).



## **The absence of the A and C zinc binding sites in Sad1 allows for the observed contraction between the L2A loop and the $\beta$ 4-5 loop.**

Unlike Sad1, those ZnF-UBPs that are competent to bind ubiquitin have open structures stabilized by the presence of A and C zinc binding sites. In HDAC6, the A zinc binds the  $\beta$ 4-5 loop and the C zinc binds to the L2A loop, constraining the positions of these loops and leading to the observed open conformation (Fig. 5A). The ZnF-UBP from USP5 also has a more open conformation relative to Sad1. USP5 has the C site zinc finger, and although it is missing the A site zinc finger, it appears to compensate for this absence by constraining the  $\beta$ 4-5 loop with intra-loop hydrogen bonds between E263, S260 and S177 (Fig. 5F). The importance of these bonds is supported by the observation that the only deviation from these well-conserved residues is with the *S. cerevisiae* and *S. pombe* homologues of USP5, which contain a CHCC zinc finger (Fig. 5G). Shown below the alignment is the corresponding sequence in Sad1, which with the crystal structure shows that Sad1 possesses no such mechanism to constrain the position of its  $\beta$ 4-5 loop. With the absence of the A and C zinc fingers in Sad1 and its homologues, a more closed conformation between the L2A and  $\beta$ 4-5 loops is likely to be a general feature of the splicing-specific ZnF-UBPs, making ubiquitin binding unlikely.

## **The iUSP domain from Sad1**

The iUSP domain of Sad1 possesses the fingers, thumb and palm sub-domains that are characteristic of USPs and a Dali structural alignment search with the iUSP domain of Sad1 shows a high degree of structural similarity to other USP structures

(Fig. 6A; Table 3) (Hasegawa and Holm 2009; Komander et al. 2009). To analyze the structure of Sad1's iUSP domain, we used the USP domain from USP2 for comparison, as it has been well characterized biochemically and has been crystallized with ubiquitin allowing us to model ubiquitin onto Sad1 (Renatus et al. 2006).

A structural alignment of Sad1's iUSP domain to USP2 bound to ubiquitin shows similarities in the palm and thumb sub-domains, but also reveals the extent to which Sad1's iUSP fingers could not be built due to the high degree of disorder in this region (Fig. 6B). As can be observed from the structural alignment, USP2 binds one zinc atom in the tips of its fingers, the presence of which has been implicated in ubiquitin recognition (Renatus et al. 2006). However, zinc binding in the fingertips is not a prerequisite for ubiquitin binding, as USP14 and USP7 are active USPs that are missing this zinc binding site in their fingertips, instead these USPs compensate for this absence through the formation of a zinc ribbon-like fold (Hu et al. 2002; Hu et al. 2005). It is unclear how Sad1's fingertips are structured as it is missing the zinc binding residues identified in USP2 and the high degree of disorder in this region makes it impossible to determine if they adopt the compensatory zinc-ribbon like fold.

The overall fold and dimensions of Sad1's iUSP domain and catalytic core are similar to other USPs, however the consensus motif for the catalytic triad of active USP domains, CHD/N is changed (Komander et al. 2009). Instead of the essential nucleophilic cysteine, Sad1 and its homologues contain an aspartate, which in Sad1 is D159. Sad1 D159 participates in a hydrogen bond network with Sad1 H405 and D422, the conformations of which closely resemble that of the USP2 active catalytic triad, composed of C33, H314 and N331 (Fig. 6F). However, consistent with the hypothesis

that Sad1 is catalytically inactive, Sad1 H405 and D422 are not well conserved, and single alanine mutations of the residues in the pseudo-catalytic triad are tolerated *in vivo* and *in vitro*, suggesting that they are dispensable for Sad1 function (Table 1).

The USP domain of USP2 recognizes ubiquitin through two distinct sets of interactions. A large cup-like surface formed between the fingers, palm and thumb sub-domains recognizes the core of ubiquitin, residues 1-71, and a deep catalytic cleft formed between the palm and thumb sub-domains interacts extensively with the C-terminal tail of ubiquitin, residues 71-76 (Fig. 6C) (Renatus et al. 2006). Sad1's iUSP domain contains a similar cup-like feature that appears large enough to accommodate the core domain of ubiquitin in a hypothetical model of Sad1 bound to ubiquitin (Fig. 6D), however the cup is partially disordered in the fingers region, complicating the interpretation of Sad1's putative ubiquitin binding surface.

The structure of Sad1's iUSP domain reveals the presence of a cleft that is reminiscent of the catalytic cleft seen in USP2, but it is missing critical elements. The catalytic cleft of USP2 appears as a deep groove that binds to the last five amino acids of the ubiquitin tail (Fig. 6C) (Renatus et al. 2006). This groove is composed of the "Cys Box", the "His Box", the "QDE Box" and two loops, Blocking Loop 1 (BL1) and Blocking Loop 2 (BL2) (Hu et al. 2005; Renatus et al. 2006). Sad1 is missing the catalytic cysteine from the Cys Box, but has the histidine and aspartate residues from the His box (Fig. 6F). The QDE box contacts the C-terminus of ubiquitin in the active site of USP2, but in Sad1, the homologous sequence appears as PDL and is in an alternate orientation (Fig. 6E). Finally, BL1 and BL2, which abut the active site in USP2,

are larger and BL2 appears more flexible in Sad1 compared to USP2, as it was unable to be resolved in the crystal structure due to its high degree of disorder (Fig. 6E).

In USPs, BL1 and BL2 are connected to the palm sub-domain in the catalytic cleft and must move to accommodate the docking of the C-terminal ubiquitin tail (Hu et al. 2005) (Hu et al. 2005; Luna-Vargas et al. 2011). As is found with other USPs crystallized in the absence of ubiquitin, Sad1's BL1 is in the closed conformation, yet BL1 possesses an extra helical region that points towards the ubiquitin tail and is larger than the USP BL1s in the PDB (Fig. 6E). The BL2s from all USPs crystallized to date are short, well-ordered loops. In Sad1, BL2 is not only 5 amino acids longer than those of all other USPs, but is also disordered, suggesting an increased degree of flexibility (Fig. 6E). Thus, the increased bulk of BL1 and BL2 may prevent Sad1 from creating an open pocket for ubiquitin binding, which may account for the observed inability to recognize ubiquitin *in vitro*.

## Conclusions

### **Sad1 protein is a general splicing factor.**

We have examined the structure and function of the essential splicing factor Sad1, which is composed of two conserved and essential domains, a zinc finger ubiquitin binding domain and an inactive USP domain. Using pre-mRNA splicing microarrays (Fig. 1B), we show that Sad1 is a general splicing factor. Inactivation of *SAD1 in vivo* causes a rapid pre-mRNA splicing defect that is both broad and severe, and this phenotype is most likely due to a tri-snRNP joining defect, as inactivation of *Sad1 in vitro* causes an immediate block to spliceosome assembly at the pre-

spliceosome stage (Fig. 1D). Recombinant Sad1 was able to complement the spliceosome assembly defect and restore *in vitro* splicing (Fig. 1C) We determined a 1.8 Å crystal structure of full-length Sad1 (Fig. 3A) and found that although the two domains of Sad1 are broadly similar to the ubiquitin signaling families from which they are derived, they deviate sharply in the regions that would be predicted to interact with ubiquitin, which may explain the inability to observe ubiquitin binding to purified Sad1 *in vitro*.

### **Sad1 protein does not bind ubiquitin *in vitro***

Sad1 lacks two of the zinc-binding sites found in ZnF-UBPs. Taken together with the dramatic rearrangement of the ubiquitin binding pocket, it is clear that the ZnF-UBP domain of Sad1 does not have the high affinity ubiquitin binding site seen in other members of this family (Fig. 2A; Fig. 5D). There are three other structures of ZnF-UBP domains that do not bind ubiquitin and sequence alignments of all ZnF-UBPs in humans and yeast indicate that there may be more (Fig. 2A) (Bonnet et al. 2008). Thus, as has been suggested previously, there are two classes of ZnF-UBPs: those that can bind ubiquitin and those, like Sad1, that cannot (Bonnet et al. 2008). Based upon the structural studies presented here and the high degree of conservation in the residues that dictate Sad1's unique L2A and  $\beta$ 4-5 loop rearrangements, it is expected that the ZnF-UBP domain from all of Sad1's splicing homologues would similarly not bind ubiquitin (Supplemental Fig. S1).

While it is clear from the crystal structure of Sad1 why the ZnF-UBP domain is unable to bind ubiquitin, it is less obvious why Sad1's iUSP domain does not bind ubiquitin. Sad1's iUSP structure resembles other USP structures; it contains the

prominent, cup-like surface that contributes to the binding of the core domain of ubiquitin in other USPs. However, the catalytic cleft also contributes to ubiquitin binding in active USPs, and this region, although present in Sad1, is dramatically remodeled. The pseudo-active site formed by Sad1's catalytic cleft lacks the essential catalytic USP residues, and contains active site loops that are longer and more disordered than those that have been seen in other USP structures. Such changes may explain the inability to observe ubiquitin binding to the USP domain of Sad1 *in vitro* and may point to an unidentified role for these features *in vivo*.

### **The ZnF-UBP domain as an activator for the neighboring USP domain**

ZnF-UBPs domains have been shown to act as ubiquitin-dependent activators for the deubiquitinase activity of their neighboring USP domains. For example, the deubiquitinase activity of USP5 is stimulated when the free C-terminal tail of ubiquitin binds to its ZnF-UPB domain, most likely through contributions to substrate recruitment (Reyes-Turcu et al. 2006; Reyes-Turcu et al. 2008; Avvakumov et al. 2012).

The ZnF-UBP domain has also been shown to contribute to USP activity in a ubiquitin independent manner. Instead of binding ubiquitin, the ZnF-UBP from Ubp8 functions as an assembly lobe to recruit co-factors, which then stimulate the catalytic activity of its neighboring USP domain (Kohler et al. 2010; Samara et al. 2010; Samara et al. 2012). Analogous to our current examination of Sad1, initial studies of Ubp8 showed that purified full-length Ubp8 did not bind ubiquitin, nor did it possess *in vitro* deubiquitinase activity (Lee et al. 2005). It was only upon the purification of Ubp8 from cell lysates and the identification of its interacting partners that *in vitro* deubiquitinase activity could be observed (Ingvarsdottir et al. 2005; Lee et al. 2005). Finally Ubp8

crystallized in the presence of these factors revealed that they assembled upon the ZnF-UBP domain, which then allowed them to interact with the fingers and catalytic cleft of Ubp8's USP domain to activate USP activity (Kohler et al. 2010; Samara et al. 2010; Samara et al. 2012). Consistent with its ubiquitin independent function, the ZnF-UBP domain from Ubp8 is missing many of the residues important for ubiquitin recognition (Fig. 2A).

In this work, we show that Sad1's ZnF-UBP is of the class that does not bind ubiquitin and possesses an iUSP domain that closely resembles the USP fold (Table 3). Given UBP8's dependence upon its non-ubiquitin binding ZnF-UBP domain for activity, it is plausible that Sad1's function similarly depends upon activation through cofactor binding to its upstream ZnF-UBP domain. This could result in the activation of a ubiquitin receptor-like function in its neighboring iUSP domain given its high similarity to known USP structures (Table 3). Additionally, the ZnF-UBP of Sad1 may be essential for its recruitment to the spliceosome, as the co-factors that decorate the ZnF-UBP of Ubp8 for activation of the USP domain are also essential for its recruitment to the SAGA complex (Ingvarsdottir et al. 2005; Lee et al. 2005; Kohler et al. 2006).

If Sad1's ZnF-UBP does act as a scaffold, then these interactions occur in the pre-mRNA splicing lysate during *in vitro* splicing and assembly. When these co-factors are identified, we predict that Sad1's iUSP domain would become competent to interact with ubiquitin. An alternative to this model is that Sad1's iUSP domain may not need activation to bind ubiquitin, but rather may require the recognition of both ubiquitin and the spliceosomal protein to which it is conjugated. Finally, Sad1's iUSP domain may not bind ubiquitin and instead may serve an entirely different function in the splicing

reaction. With the resolution of which proteins directly interact with Sad1, which would be a challenge given Sad1's low affinity for splicing factors, we would distinguish between these potential mechanisms. The work presented here provides critical insight into the function of Sad1 by demonstrating Sad1's role in spliceosome assembly as a general splicing factor and by providing a structural perspective on the role of its essential and well-conserved ubiquitin-related domains.

## Materials and Methods

### **Strain Construction and 5-FOA selection**

The *sad1-1* allele was cloned from the original *sad1-1* isolate (Lygerou et al. 1999) along with 800 base pairs upstream of the start site and 900 base pairs downstream of the stop codon into a pRS313 *HIS3* CEN/ARS plasmid. The same primers were used to clone a wild-type copy of *SAD1* from *S. cerevisiae* S288C into both a pRS316 *URA3* CEN/ARS plasmid and a pRS313 *HIS3* CEN/ARS plasmid. The sequences were identical with the exception of T707C and C827T, resulting in the amino acid changes, L236P and S276F respectively. Subsequent mutants and truncations of *SAD1* were produced by standard PCR mutagenesis of the *SAD1* pRS313 plasmid. To produce a deletion strain of *SAD1*, the *SAD1* gene was deleted using PCR based homologous recombination in a diploid S288C strain using the KanMX marker (Longtine et al. 1998). After selection for kanamycin resistance, an individual colony was transformed with the wild-type *SAD1* pRS316 plasmid and after selection in the absence of uracil and the presence of kanamycin, the transformants were allowed to sporulate. Tetrads were dissected and grown on complete media, then replica plated to media lacking uracil and containing kanamycin for the selection of



haploid cells harboring both the KanMX marked *SAD1* deletion and the wild-type *SAD1* pRS316 plasmid, producing the *SAD1* pRS316/*sad1::kan* strain.

5-Fluoroorotic acid (5-FOA) screening was performed by transforming the *SAD1* pRS316/*sad1::kan* strain with either the wild-type *SAD1* pRS313 plasmid, an empty pRS313 plasmid, or a pRS313 plasmid containing a mutation, or truncation of *SAD1*. Selection was performed in the absence of histidine and in the presence of uracil to allow for the passive loss of the wild-type *SAD1* pRS316 plasmid. Individual colonies were patched, then replica plated to 5-FOA containing plates. Growth indicated the ability to lose the wild-type *SAD1* pRS316 plasmid and therefore the viability of the genetic manipulation. Transformants that survived on 5-FOA were further screened for temperature sensitivity by testing for growth at 30 °C and 37 °C.

### **Protein Purification**

Full-length Sad1, amino acids 1-448, was cloned from *S. cerevisiae* S288C into the pET-28b plasmid at the Bam HI and Xho I sites, which fuses a 6-His tag to the N-terminus. BL21 *E. coli* cells were transformed and induced with 0.1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside at an optical density of .8 at 600 nm and the cultures were allowed to grow overnight at 18 °C. Cells were flash frozen and resuspended in 50 mM HEPES pH 7.9, 500 mM NaCl, 20 mM Imidazole, 1 mM tris (2-chloroethyl) phosphate (TCEP). After sonication, the soluble lysate was incubated with Ni-NTA, then washed with 50 mM HEPES pH 7.9, 150 mM NaCl, 20 mM imidazole, 1 mM TCEP and eluted with wash buffer supplemented with imidazole to 300 mM. The eluate was concentrated and applied to a S200 Superdex size exclusion column equilibrated with

20 mM Hepes pH 7.9, 50 mM KCl and 1 mM TCEP. The peak fractions were pooled, concentrated to 50  $\mu$ M Sad1, then flash frozen. This protein was used for crystallization and for use in *in vitro* splicing and assembly assays.

### **Crystallization**

The initial screen to identify crystallization conditions for full-length Sad1 was conducted by the Hauptman-Woodward Institute with 50  $\mu$ M 6-His-Sad1. Under oil hits were optimized by vapor-diffusion hanging drop methods and the optimal crystal conditions were determined to be: 1:1 50  $\mu$ M 6-His Sad1 with 13% PEG 4000, 0.1 M sodium acetate and 0.1 M magnesium acetate pH 5.5 at 25 °C and were fully grown within 48 hours. Tungsten-derivative Sad1 crystals were produced by soaking native crystals for 2 hours in mother liquor supplemented with 10 mM sodium tungstate. Optimal cryogenic conditions were achieved by soaking crystals in a drop of mother liquor supplemented with PEG 400 to 25%.

### **Structure Solution**

Data were collected at 100 K at the Advanced Light Source and processed using the program suite Elves (Holton and Alber 2004). Phases were determined from a single wavelength anomalous dispersion experiment. A single tungsten site was found in the HYSS module of Autosol in Phenix (Adams et al. 2010), which was used to calculate phases with the program Phaser and then to generate an electron density map with Resolve for Autobuild. 375 of 414 residues were placed by Autobuild and then 39 residues were placed manually into the density. The final model was refined in Phenix to an  $R/R_{\text{free}}$  of 17.1/20.2 (Adams et al. 2010). Coordinates and Structure factors were deposited in the Protein Data Bank with accession codes 4MSX.

## **Pre-mRNA splicing arrays**

The following is adapted from previously designed *S. cerevisiae* pre-mRNA microarray studies (Pleiss et al. 2007). Briefly, 100 ml of *SAD1* pRS313/*sad1*Δ::kan and *sad1-1* pRS313/*sad1*Δ::kan cells were grown to mid-log phase at 25 °C, then split to two 50 ml cultures, which were either kept at 25 °C, or shifted to 37 °C for 30 minutes until cells were collected and frozen. RNA was extracted and split for dN9 cDNA synthesis in the presence of either Cy5, or Cy3 labeled dUTP. Cy3/Cy5 labeled *SAD1* and *sad1-1* cDNAs were competitively hybridized to pre-mRNA microarrays composed of probes that were designed to hybridize to sequence elements in the intron, the exon and the junction of every yeast intron-containing gene. Microarray fluorescence was measured at 532 and 635 nm with an Axon Instruments GenePix 4000B and images were processed with Axon Instruments GenePix Pro version 5.1. Six Technical replicate spots per array and dye flipped replicate arrays were combined and normalized as log<sub>2</sub> transformed ratios that were then averaged and subjected to centroid-linkage hierarchical clustering using Cluster 3.0 and displayed graphically to produce heat maps with Java Treeview.

## ***In vitro* splicing and assembly**

Splicing extracts were prepared as follows from *SAD1* pRS313/*sad1*Δ::kan and *sad1-1* pRS313/*sad1*Δ::kan grown at 30 °C to an optical density of 2 at 600 nm. Cells were resuspended in .4X volumes AGK buffer (20mM Hepes pH 7.9, 200 mM KCl, 2.5 mM MgCl<sub>2</sub>, protease inhibitors) and frozen. Cells were ball-milled five times at 10 Hz for 3 minutes and thawed lysates were centrifuged at 50,000 g for 30 minutes to remove membranes and centrifuged again at 125,000 g for 80 minutes to remove ribosomes.

Finally, lysates were buffer exchanged into buffer D (20 mM Hepes pH 7.9, 10% glycerol, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM DTT) using a desalting column.

*In vitro* splicing reactions were prepared with 8 µl of splicing extract, which was added to a reaction containing 3% PEG 8000, 60 mM potassium phosphate pH 7.0, 2.5 mM magnesium chloride and 10 units of RNasin. The *ACT1* pre-mRNA, transcribed with Cy5 labeled UTP (as 5% of the UTP) was incubated with *sad1-1* splicing extracts that were either inactivated completely with an incubation at 37 °C, or not. Full-length Sad1 was supplemented to the reactions in a 10-fold serial dilution starting from 105 nM. After the addition of ATP to 2 mM, the reactions were incubated at 25 °C for 30 minutes and stopped with the addition of SDS to 2.5 %, 1 mM EDTA. 10 units of Proteinase K were added and the reactions were incubated at 60 °C for 10 minutes, then loaded onto a 6% polyacrylamide, 6 M urea, TBE gel.

Using a P<sup>32</sup> labeled *ACT1* pre-mRNA, *In vitro* assembly reactions were prepared in an identical manner to that of the *in vitro* splicing reactions with the exception that the reactions were stopped with the addition of 2 X Native buffer (60 mM KPO<sub>4</sub>, 3 mM MgCl<sub>2</sub>, 3 % PEG 8000 and 8 % glycerol, 10 mg/ml heparin) and were then resolved on a 26 cm long, 0.5 mm thick, 4% 80:1 polyacrylamide gel that was run at 450 volts for 4.5 hours in TAE.

### **Isothermal titration calorimetry**

To express the iUSP domain of Sad1 for ITC, residues 124-448 were cloned and purified in a similar manner to full-length Sad1, but to aid in solubility, the Hepes pH 7.9 buffer was replaced with Tris-Cl pH 8.0 and the size exclusion buffer was 20 mM Tris-Cl pH 8.0, 150 mM NaCl and 1 mM TCEP. Full-length Sad1 was also purified in this

manner for the ITC experiments and showed no difference in potency in the in vitro splicing, or assembly add-back assays (data not shown). The plasmid to express the ZnF-UBP domain from USP5 was a gift from Dr. Hong-Yu Hu, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China and the protocol followed for purifying the protein did not vary from the published one (Zhang et al. 2011). Ubiquitin was purchased from Boston Biochemicals.

50  $\mu$ M of Sad1, 50  $\mu$ M of the iUSP domain from Sad1, 50  $\mu$ M of the ZnF-UBP domain from USP5 and 1 mM ubiquitin were dialyzed extensively against 25 mM Tris pH 7.4, 150 mM NaCl and 0.5 mM TCEP. The receptor cell was filled with 50  $\mu$ M Sad1, 50  $\mu$ M Sad1-iUSP, or 50  $\mu$ M USP5- ZnF-UBP. 1 mM ubiquitin was used as the titrant. Experiments were carried out on a Microcal VPITC at 25° C.

### **GST-UBL immunoprecipitations**

Hub1, SUMO, ubiquitin, and Rub1 were cloned from *S. cerevisiae* S288C into the pGEX 6-P-1 vector, producing N-terminal fusions with glutathione-S-transferase (GST). Proteins were purified in a similar manner to the 6-His tagged proteins above. 20 mM Tris-Cl pH 8.0, 150 mM NaCl and 1 mM TCEP was used as the resuspension and wash buffer, which was then supplemented with 20 mM glutathione for elution. Eluates were concentrated and dialyzed extensively against 20 mM Tris-Cl pH 8.0, 150 mM NaCl and 0.5 mM TCEP. Full-length Snu66 was cloned from *S. cerevisiae* S288C into pET28b, placing a 6-His tag on its N-terminus. 6-His Snu66 was purified in a manner identical to the Sad1 iUSP domain.

6 nmol of each of the GST-UBLs were incubated separately in 500  $\mu$ l of wash buffer with 12.5  $\mu$ l magnetic glutathione beads for 1 hour at 4 °C to allow for binding,

then beads were washed with 20 mM Tris-Cl pH 8.0, 150 mM NaCl and 1 mM TCEP. 500  $\mu$ l of 50  $\mu$ M Sad1, 50  $\mu$ M Snu66 or 50  $\mu$ M USP5-ZnF-UB were added to the beads and incubated for an additional hour at 4 °C before washing and eluting in 50  $\mu$ l wash buffer supplemented with 20 mM glutathione. Eluates were analyzed on 8% SDS-polyacrylamide gel electrophoresis.

### **Biolayer Interferometry**

Ubiquitin labeled with a single biotin on its N-terminus was purchased from Thermo Scientific. Sad1 and USP5 were prepared in a manner identical to that, which was used for the ITC experiments. Biolayer interferometry (BLI) was measured at 25 °C with an Octet384 (ForteBio). Biotin ubiquitin was immobilized on streptavidin biosensors at 10 mg/ml in assay buffer (25 mM Tris-Cl pH 8.0, 150 mM NaCl and 1 mM TCEP) for 5 minutes. Probes were then washed with assay buffer for 1 minute, regeneration buffer (.1% SDS, 10 mM glycine-HCl pH 2.5) for 30 seconds, and assay buffer again for 1 minute. All binding measurements were carried out at the indicated concentrations of either USP5 ZnF-UBP or Sad1, for the indicated amounts of time. A single biosensor was used to measure ubiquitin binding at each concentration. BLI responses to buffer alone and responses to streptavidin coated tips without immobilized ubiquitin were subtracted from each curve using the ForteBio analysis suite. Data for processed association binding curves were analyzed with GraphPad Prism 6.0 (GraphPad, San Diego, CA), where they were fit using the equation for a one-phase association. Saturation binding values from the plateaus of the one-phase associations ( $R_{eq}$ ) were plotted as a function of titrant concentration and fit with a single-site saturation binding model to obtain  $K_D$ s when possible.

## **Data Deposition**

The coordinates and the structure factors have been deposited in the Protein Data Bank with accession codes 4MSX.

Tables

**Table 1: Strains tested for complementation of *sad1*Δ::kan by 5-FOA selection**

<b>plasmid</b>	<b><i>sad1</i>Δ::kan at 30 °C</b>	<b><i>sad1</i>Δ::kan at 37 °C</b>
<i>sad1</i>	viable	viable
empty vector	lethal	not tested
<i>sad1-1</i>	viable	lethal
<i>sad1</i> ZnF-UBPΔ	lethal	not tested
<i>sad1</i> iUSPΔ	lethal	not tested
<i>sad1</i> C60A	lethal	not tested
<i>sad1</i> C63A	viable	viable
<i>sad1</i> H79A	lethal	not tested
<i>sad1</i> H85A	viable	viable
<i>sad1</i> D159A	viable	viable
<i>sad1</i> H405A	viable	viable
<i>sad1</i> D422A	viable	viable



**Table 2: Data collection and refinement statistics**

	Native	W derivative
<b>Data collection</b>		
Space group	P4 <sub>3</sub> 2 <sub>1</sub> 2	P4 <sub>3</sub> 2 <sub>1</sub> 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	139.315, 139.315, 86.737	138.685, 138.685, 87.755
Resolution (Å)	98.511-1.87	98.064-2.61
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.085(0.94)	0.128(1.13)
<i>I</i> /sig( <i>i</i> )	10.5(1.7)	10.7(1.6)
Completeness (%)	99.5(100.0)	100(100)
Redundancy/Multiplicity	7.7(7.6)	7.9(7.3)
Wavelength	1.1159	1.2146
<b>Refinement</b>		
Resolution (Å)	1.87	
No. reflections	70404 (6898)	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	17.1/20.2	
No. atoms	7109	
Protein	6780	
Ligand/ion	8	
Water	321	

<i>B</i> -factors		
Protein	43.20	
Ligand/ion	50.50	
Water	51.20	
R.m.s. deviations		
Bond lengths (Å)	.018	
Bond angles (°)	1.605	
Ramachandran favored (%)	98	
Ramachandran outliers (%)	0	

Statistics for the highest-resolution shell are shown in parentheses.

**Table 3: ZnF-UBP domains from the PDB with the highest structural similarity as determined by the DALI server for the ZnF-UBP domain of Sad1 (amino acids 29-124).**

Description	Chain	RMSD	# Res. Aligned	Total # Res.	% ID	Z score
HDAC6	3QV4-A	1.8	84	99	20	12.7
USP5	2PHD-C	2.2	85	673	13	10.1
USP16	2I50-A	2.2	83	122	22	9.6
USP33	2UZG-A	2.1	75	95	24	8.7
ETP1	2IDA-A	3.0	80	102	14	8.2
USP13	2L80-A	3.0	88	114	15	8.0
UBP8	3MHS-A	2.7	71	455	20	7.9

**Table 4: USP domains from the PDB with the highest structural similarity as determined by the DALI server for the iUSP domain of Sad1 (amino acids 125-448).**

Description	Chain	RMSD	# Res. Aligned	Total # Res.	% ID	Z score
USP4	2Y6E-B	2.3	244	339	23	25.5
USP8	2GFO-A	2.6	246	339	21	25.2
USP21	3MTN-A	2.5	240	315	18	25.1
USP2	2IBI-A	2.4	237	315	20	24.8
UBP8	3MHH-A	2.7	242	433	17	24.3
USP5	3IHP-A	2.6	245	673	16	7.9
USP7	1NBF-A	2.4	237	347	15	22.8
UBP6	1VJV-A	2.4	238	367	16	22.7
USP14	2AYN-A	2.3	234	337	13	22.5
CYLD	2VHF-A	3.2	219	341	18	17.7

## Figure Legend

Figure 1: Sad1 is a core splicing factor necessary for spliceosome assembly. (A) Domain topology for *S. cerevisiae* Sad1, the human homologue of Sad1, USP39, and a canonical active USP domain, with the catalytic residues denoted by asterisks. The ZnF-UBP domain is colored pale orange, a short linker domain is pale green and the iUSP domain is colored slate. USP39 possesses an additional N-terminal RS domain. (B) The yeast pre-mRNA splicing microarray is composed of probes that detect the pre-mRNA intron, “P”, the exon “T” for total, and the exon/exon junction for the mature mRNA, or “M”. *SAD1* and *sad1-1* cultures were grown at 25 °C to mid-log phase and were then shifted to 37 °C for 30 minutes. Total RNA was extracted and Cy3 and Cy5 labeled cDNA produced. The cDNAs were competitively hybridized to the array and the resulting log<sub>2</sub>-ratios for each gene were plotted after performing centroid-linkage hierarchical clustering. Shown side by side are the heat maps of plots produced by two biological replicate experiments, “a” and “b”, where increasing amounts of yellow signal for the intron probe, “P”, reflects intron accumulation and is indicative of a pre-mRNA splicing defect. (C) *sad1-1* splicing extracts were inactivated for 10 minutes at 37 °C and where indicated, a 10-fold decreasing dilution series of 105 nM full-length Sad1 was added. Pre-mRNA splicing reactions were incubated at 25 °C in the presence of 2mM ATP with Cy5-labeled pre-*ACT1* for 30 minutes then resolved on a 6% denaturing polyacrylamide gel. (D) Splicing extracts from *sad1-1* were prepared and inactivated in a manner identical to that of the in vitro splicing assay in part C, but the products of the reaction were resolved by native gel electrophoresis. An RNase H-catalyzed knockdown of the U6 snRNA in a *SAD1* wild-type strain is shown for comparison.

Figure 2: Sad1 does not bind ubiquitin *in vitro*. (A) A primary sequence alignment of ubiquitin binding and non-ubiquitin binding *H. sapiens* and *S. cerevisiae* ZnF-UBP domains. The WRY motif residues that are essential for ubiquitin binding in USP5 are annotated with asterisks; in Sad1 they appear as YRL. The residues that comprise the three zinc fingers found in ZnF-UBPs are denoted A, B and C; Sad1 only possesses the B site zinc finger. Conservation is indicated by shades of gray and gaps in the alignment are represented by dashes. (B) 1 mM of ubiquitin was titrated into an ITC receptor cell containing either 50 uM full-length Sad1, Sad1-iUSP, USP5 ZnF-UBP, or buffer at 25 °C. The data from the USP5 ZnF-UBP titration with ubiquitin was fit to a single exponential in good agreement with one binding site and a dissociation constant of 2.3 uM. ITC curve-fitting and data analysis were performed with the MicroCal-enabled Origin software.

Figure 3: The crystal structure of full-length Sad1. (A) The ZnF-UBP (pale orange) and the iUSP domain (slate) are connected through a 25 amino acid linker (pale green). (B) Intra-domain hydrogen bonding of K199 to D43 and K207 to D41 provides inter-domain stabilization. (C) Intra-domain hydrophobic packing between, F206, L140, R37 and a hydrophobic patch on the ZnF-UBP domain further specifies the orientation of the two domains.

Figure 4: The ZnF-UBP domain of Sad1 binds one Zn atom. (A) Sad1's ZnF-UBP domain (pale orange) with secondary structure elements labeled, including loop L2A

and the  $\beta$ 4-5 loop. This view is from a 50° rotation about the y-axis of the full-length molecule in figure 3B with the iUSP domain removed for clarity. (B) A structural alignment produced with Pymol of Sad1's ZnF-UBP to the ZnF-UBP domain from *H. sapiens* USP16 (yellow, PDB ID: 2I50). The three zinc atoms of USP16, A-C, appear as grey spheres and the B site zinc atom from Sad1 appears as a pale orange sphere. (C) Structural alignment of the B site zincs from SAD1 ZnF-UBP, *H. sapiens* HDAC6 ZnF-UBP (violet, PDB ID: 3C5K) and USP16 ZnF-UBP. The residues that comprise the CCHH zinc finger of Sad1 are labeled.

Figure 5: The L2A and  $\beta$ 4-5 loops of Sad1's ZnF-UBP domain are contracted relative to other ZnF-UBPs. (A) Structural alignment of Sad1's ZnF-UBP domain (pale orange) to the ZnF-UBP domains from *H. sapiens* HDAC6 (violet, PDB ID: 3C5K) and *H. sapiens* USP5 (cyan, PDB ID: 2G43). The zinc atoms of HDAC6 and USP5 are depicted as grey spheres and the zinc atom from Sad1 is depicted as an orange sphere. Structural alignments were performed in Pymol. (B) Structural alignment of Sad1's ZnF-UBP main-chain (pale orange) to all other ZnF-UBP domain main-chains available in the PDB (grey) (2G45, 2G43, 3IHP, 3C5K, 3PHD, 3GV4, 2UZG, 2I50, 2L80, 2IDA, 3M99, 3MHH and 3MHS). Sad1 is an outlier with respect to the distance between its L2A and  $\beta$ 4-5 loops. (C) Structural alignment of Sad1's ZnF-UBP domain to the apo ZnF-UBP domain from USP5 (PDB ID: 2G43). The ubiquitin interacting residue R221 of USP5 is pointed towards solvent. In contrast, the corresponding residue in Sad1, R70, is held close to the surface through two hydrogen bonds to Y58 and N102. Hydrogen bonds were determined by Pymol with a 3.1 Å cut-off. (D) Structural alignment of Sad1's ZnF-

UBP to the ZnF-UBP domain from USP5 bound to ubiquitin (raspberry) (PDB ID: 2G45), shows that the ubiquitin tail as it binds to USP5 would clash with the unique placement of Sad1 R70. (E) Surface representations of the ZnF-UBP domains from USP5 and Sad1, either with or without ubiquitin. The structural alignment from 4D was used to model Sad1 with ubiquitin and USP5 was removed for clarity. The deep depression present on the surface of USP5's ZnF-UBP domain, is entirely filled by the unique placement of Sad1 R70. (F) Structural alignment of the A site zinc finger of HDAC6 to the corresponding regions of USP5 and Sad1. The residues of USP5 that are engaged in compensatory hydrogen bond interactions are labeled. (G) Primary sequence alignment of USP5 homologues aligned with the T-coffee server. The *H. sapiens* USP5 residues that engage in compensatory A site zinc finger hydrogen bond interactions are highlighted. Shown below the alignment for reference are the equivalent residues in Sad1.

Figure 6: The iUSP domain of Sad1 aligns well with known USP structures. (A) The fingers, palm and thumb sub-domains of Sad1's iUSP domain (slate blue) are circled and the blocking loops indicated. (B) Structural alignment of the Sad1 iUSP domain to the USP domain from *H. sapiens* USP2 (lime) bound to ubiquitin (PDB ID: 2HD5) where ubiquitin was removed afterwards for clarity. The structural alignment was produced in Pymol. (C) Transparent surface representation of USP2 bound to ubiquitin (raspberry). The cup-like surface, which binds to the globular fold of ubiquitin and the catalytic cleft, which binds to the C-terminal tail of ubiquitin are highlighted. (D) A hypothetical model



of the transparent surface representation of Sad1's iUSP domain after a structural alignment of Sad's iUSP domain to USP2 bound to ubiquitin. USP2 was removed to allow modeling of Sad1 iUSP's capacity to accommodate ubiquitin in its partial cup like surface. (E) A zoomed image of the catalytic cleft from USP2 bound to ubiquitin aligned to Sad1's iUSP domain. The two blocking loops, the QDE loop of USP2 and the homologous PDL loop of Sad1 are labeled to highlight their different conformations. (F) A zoomed image of the catalytic triad from USP2 aligned to the pseudo-catalytic triad of Sad1's iUSP domain shows a high degree of structural similarity.

Supplemental Figure 1: Sequence alignment of Sad1 homologues from *S. cerevisiae* (*S.c.*) Sad1, *S. pombe* (*S.p.*) UBP10, and *H. sapiens* (*H.s.*) USP39 produced using the T-coffee server. Pseudo-catalytic triad residues of Sad1 are indicated by asterisks. Conserved inter-domain residues that interact via hydrogen bonds are indicated by blue spheres. Conserved inter-domain residues that participate in hydrophobic interactions are indicated by orange spheres. The residues that comprise the Sad1 ZnF-UBP B-site zinc finger are denoted with green Bs. The WRY motif residues that are essential for ubiquitin binding in USP5 are labeled with red arrowheads, these residues occur as YRL in Sad1 and its homologues. Conservation is indicated by shades of gray and gaps in the alignment are represented by dashes. Below the alignment is the USP consensus sequence derived from an alignment of all USPs from *H. sapiens* and *S. cerevisiae* (Ye et al. 2009). The Cys, QDE and His boxes that define the USP catalytic cleft are noted.

Supplemental Figure 2: Measuring the binding affinity of Sad1 and USP5 ZnF-UBP for ubiquitin by biolayer interferometry (BLI). (A) BLI association response curves for the indicated concentrations of USP5 ZnF-UBP binding to ubiquitin that was immobilized on streptavidin coated biosensors. The raw data represent the average of two experiments (red). Curves were fit using a one-phase association equation (black) to determine the amount of titrant bound at saturation (response at equilibrium,  $R_{eq}$ ). (B) BLI association response curves for the indicated concentrations of Sad1 binding to ubiquitin immobilized on streptavidin coated response tips. The raw data represent the average of two experiments (red). Curves were fit using a one-phase association equation (black) to determine the amount of titrant bound at saturation ( $R_{eq}$ ). (C) Saturation binding data from BLI experiments (A, B) for USP5 ZnF-UBP (triangles) and Sad1 (squares) plotted against concentration of titrant. The USP5 ZnF-UBP data was fit using a one-site saturation binding model and give a  $K_D$  value of 3.4 mM ( $R^2 = .9291$ ) for USP5 ZnF-UBP, while a  $K_D$  for Sad1 was indeterminable.

Supplemental Figure 3: Sad1 does not bind to yeast GST-UBLs *in vitro*. (A) GST-UBL immunoprecipitations with Sad1, or USP5. The only specific interaction that was observed was between GST-ubiquitin and the ZnF-UBP domain from USP5. (B) GST-UBL immunoprecipitations with Sad1, or Snu66. The only specific interaction observed was between GST-Hub1 and Snu66.

Supplemental Figure 4: The ZnF-UBP from USP5 is missing the B site CCHH zinc binding motif and instead possesses conserved residues that participate in compensatory hydrogen bonds (A) Structural alignment of the B site zinc finger of SAD1 aligned to the same region of the ZnF-UBP from *H. sapiens* USP5 (green cyan, PDB ID: 2G43). The residues of USP5 that are engaged in compensatory hydrogen bonds are labeled as calculated by Pymol. Water molecules appear as small marine spheres. Sad1's B site zinc atom is colored grey. (B) Primary sequence alignment of USP5 homologues aligned with the T-coffee server. The *H. sapiens* USP5 residues that engage in the compensatory B site zinc finger interactions in supplemental figure 3A are indicated and outlined in green cyan boxes, except for the USP5 homologues from *S. pombe* and *S. cerevisiae*, which are boxed in red to indicate their difference. Shown below the alignment for reference are the equivalent residues in Sad1.

# Figures

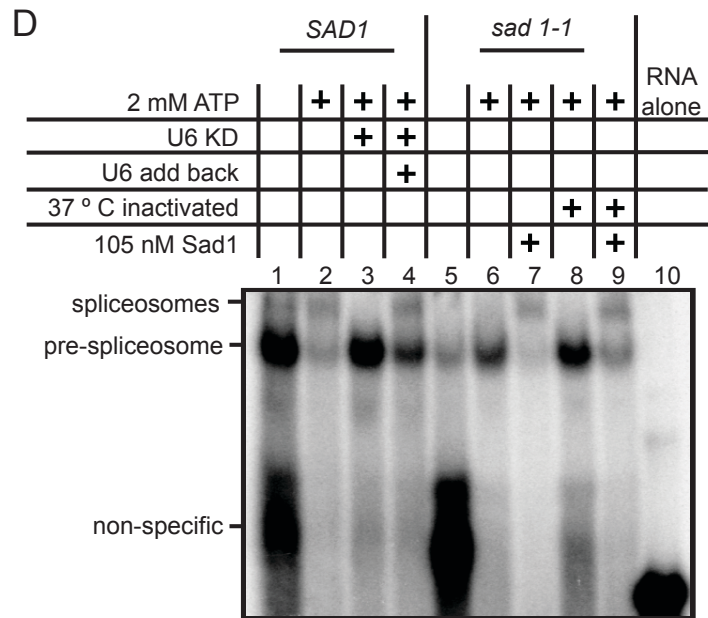
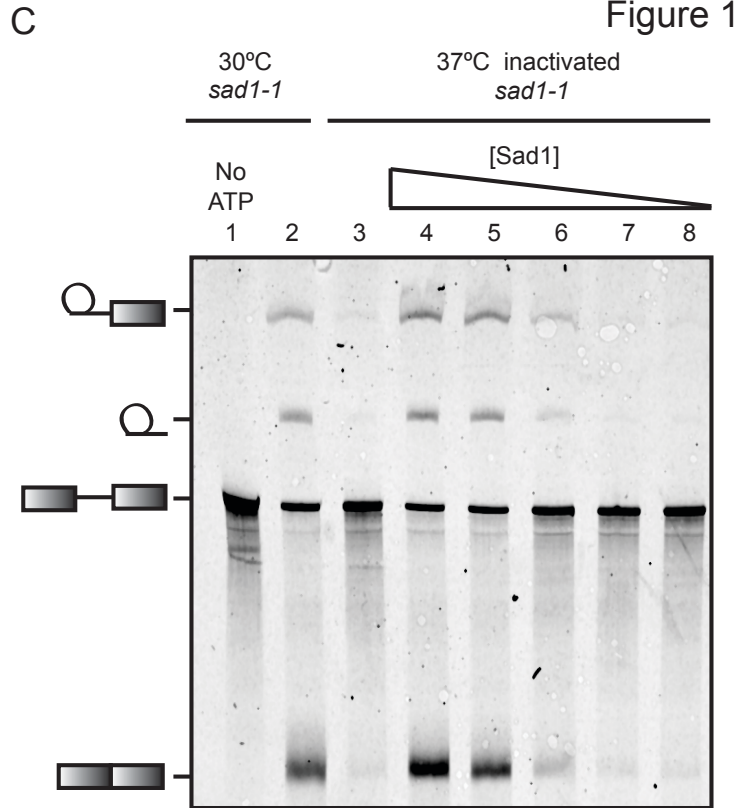
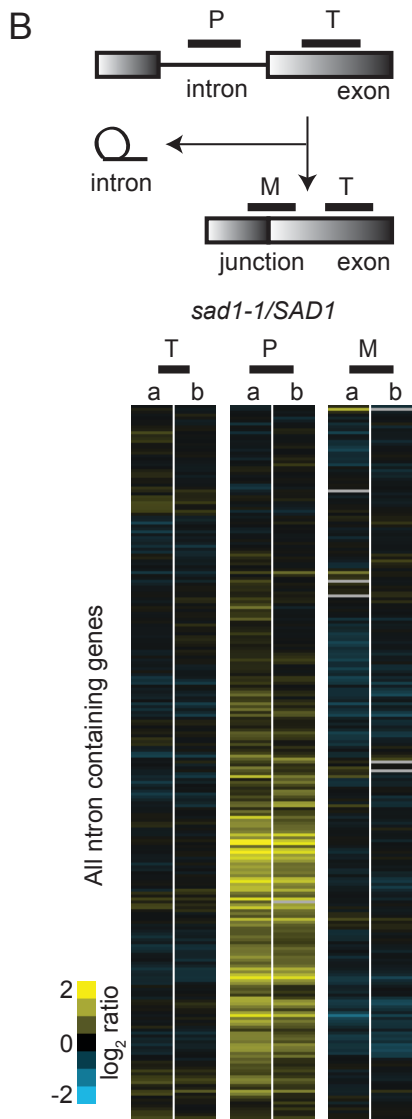
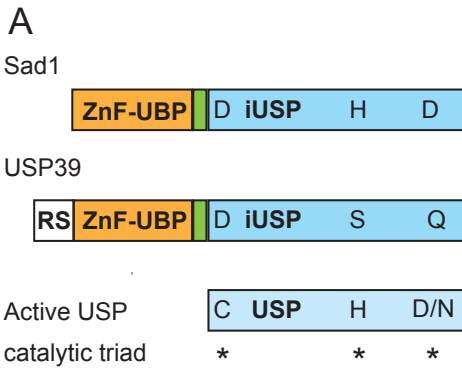
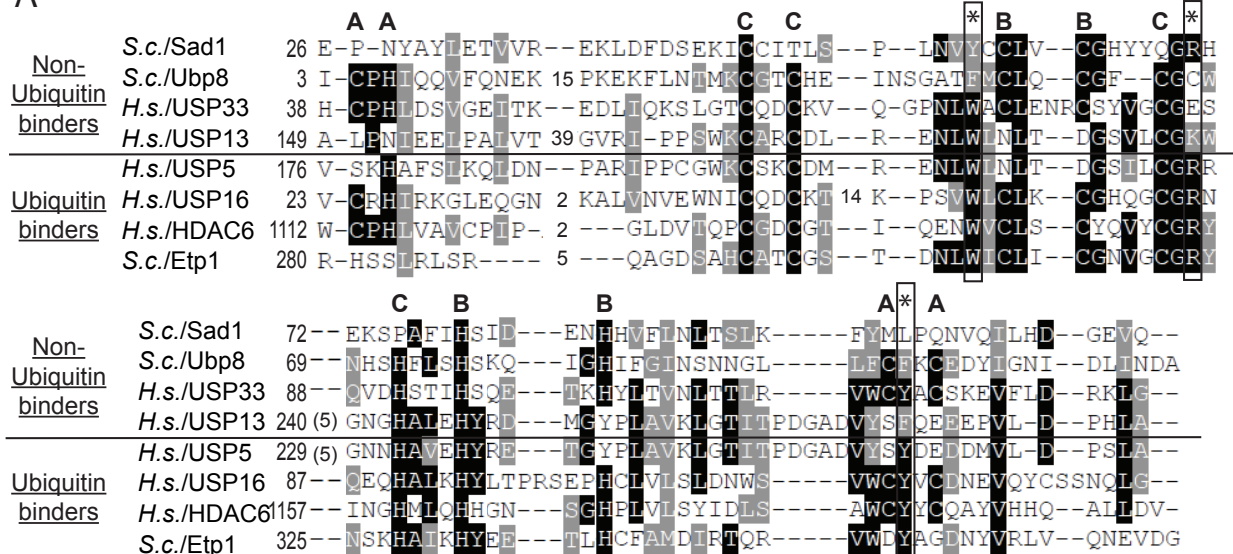


Figure 2

A



B

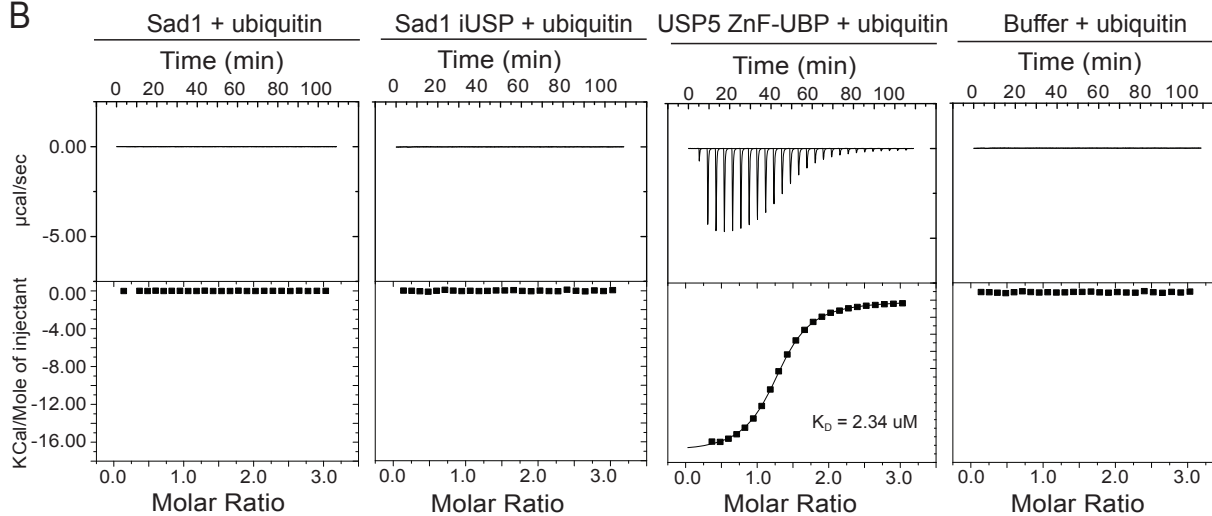


Figure 3

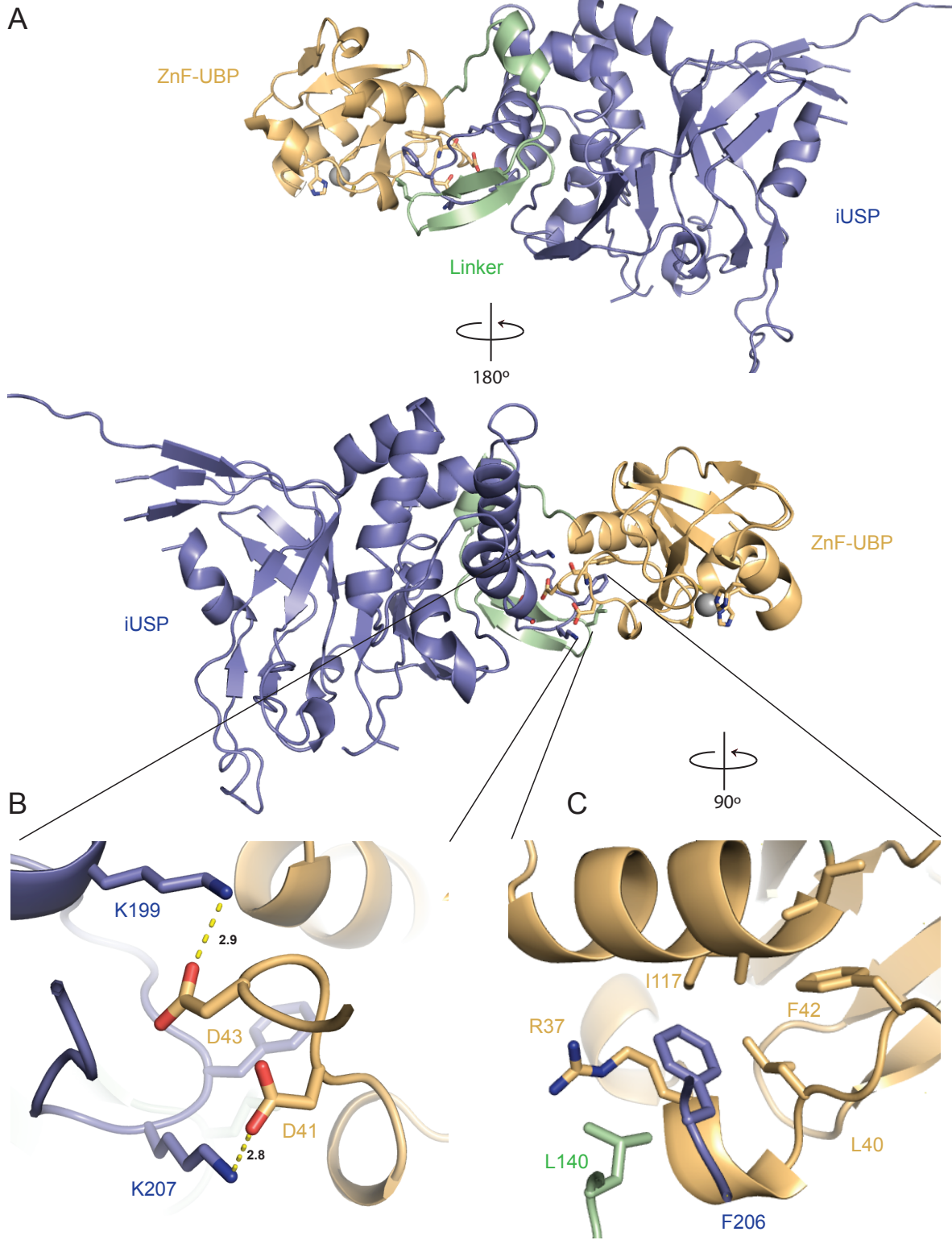
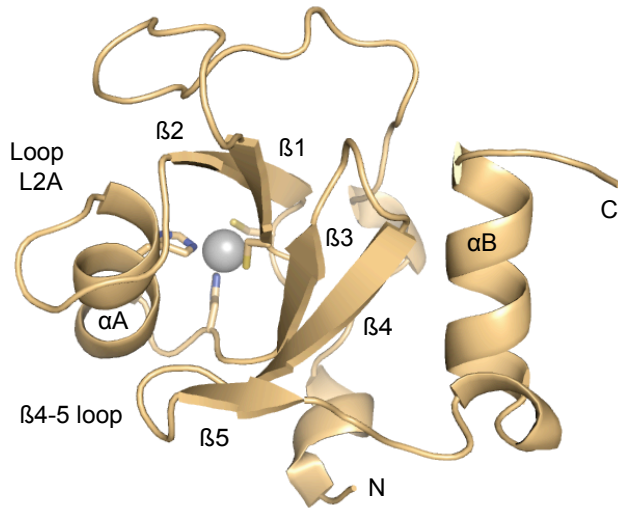
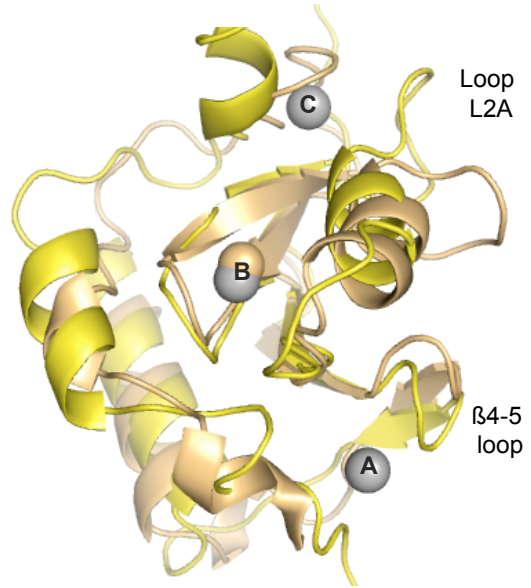


Figure 4

A Sad1 ZnF-UBP



B Sad1 ZnF-UBP  
USP16 ZnF-UBP



C Sad1 ZnF-UBP  
USP16 ZnF-UBP  
HDAC6 ZnF-UBP

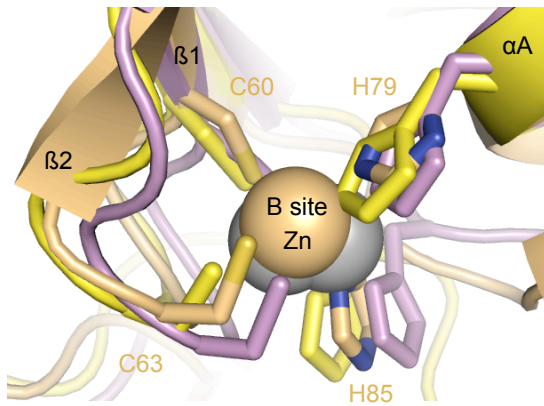


Figure 5

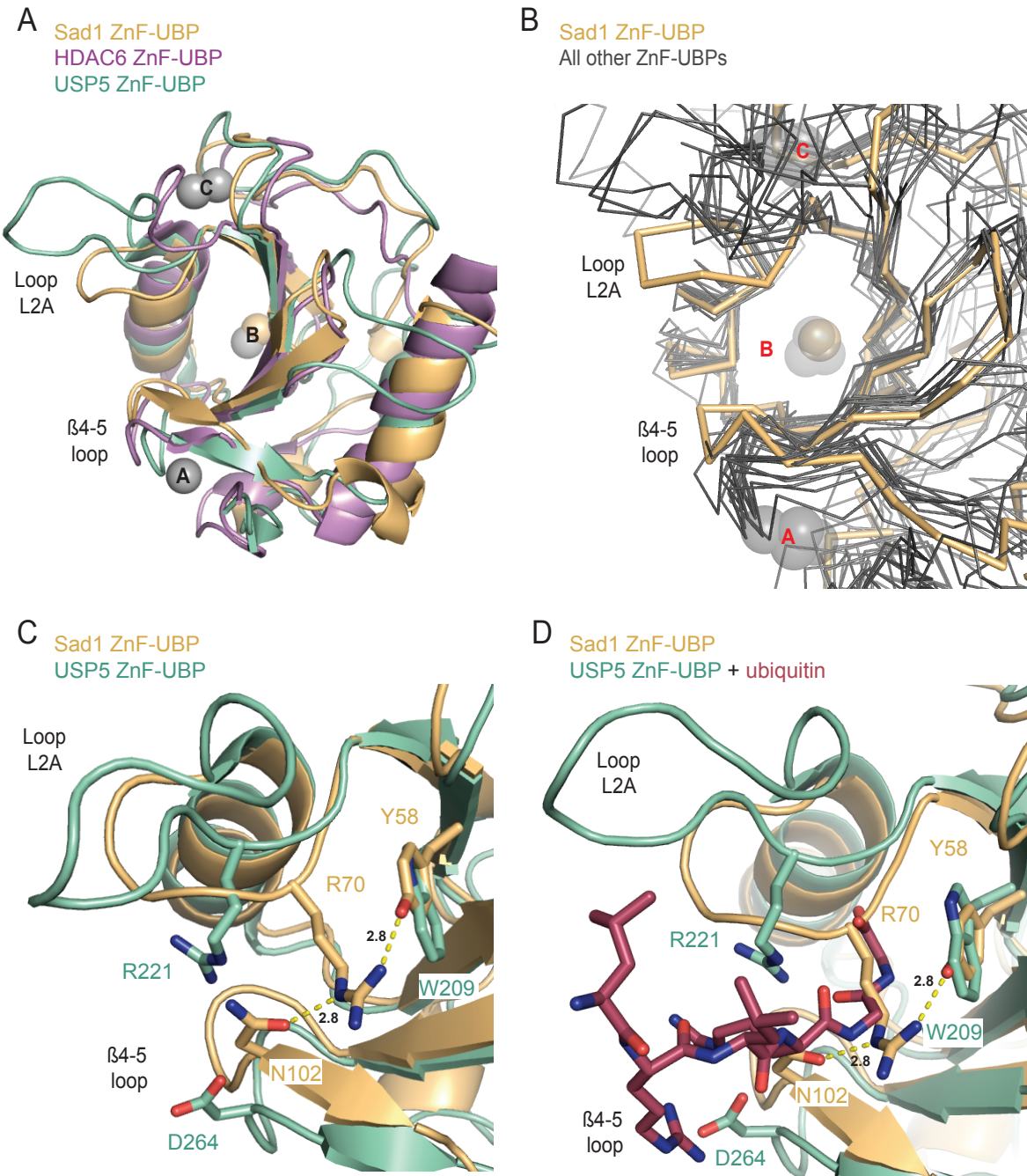




Figure 5

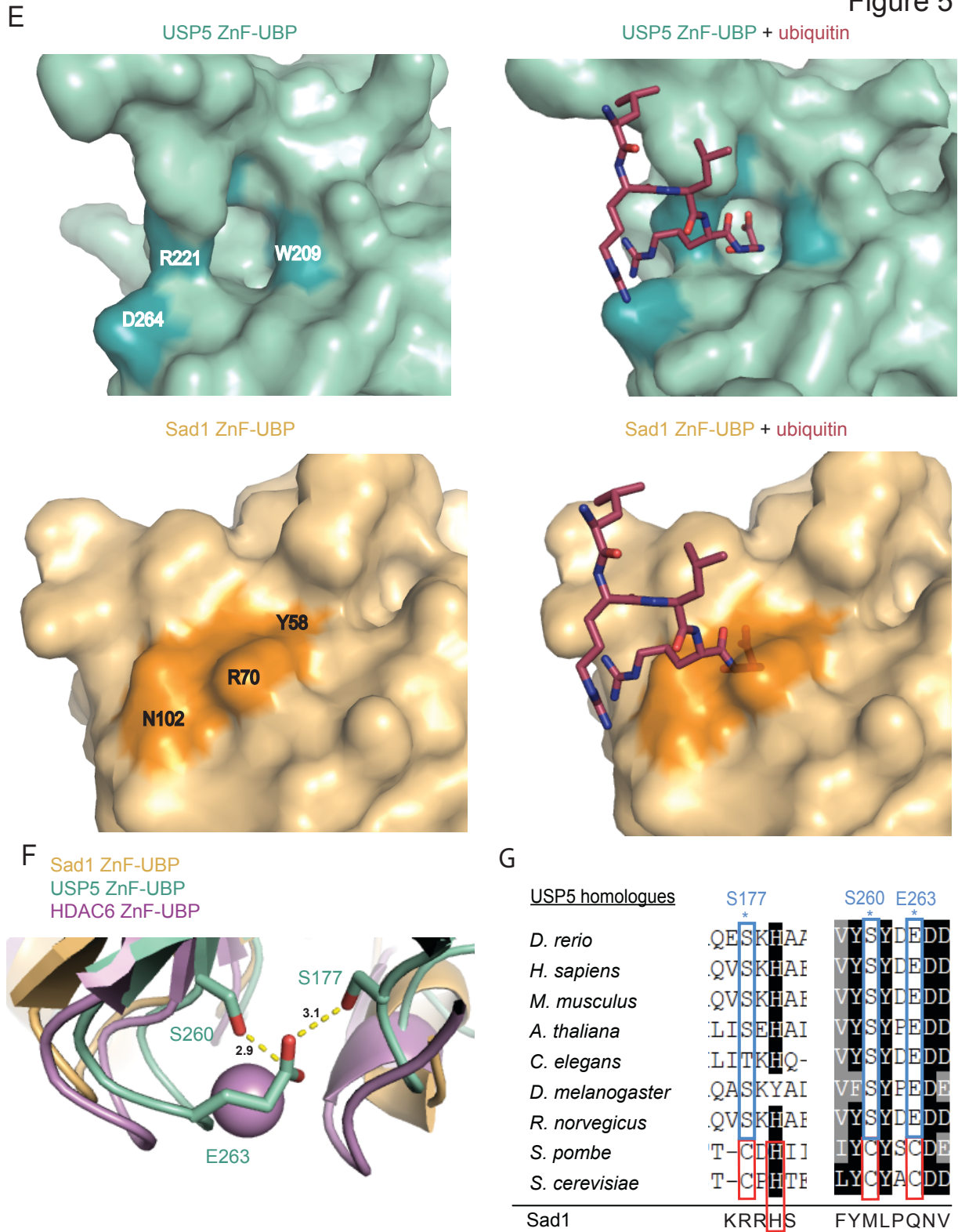


Figure 6

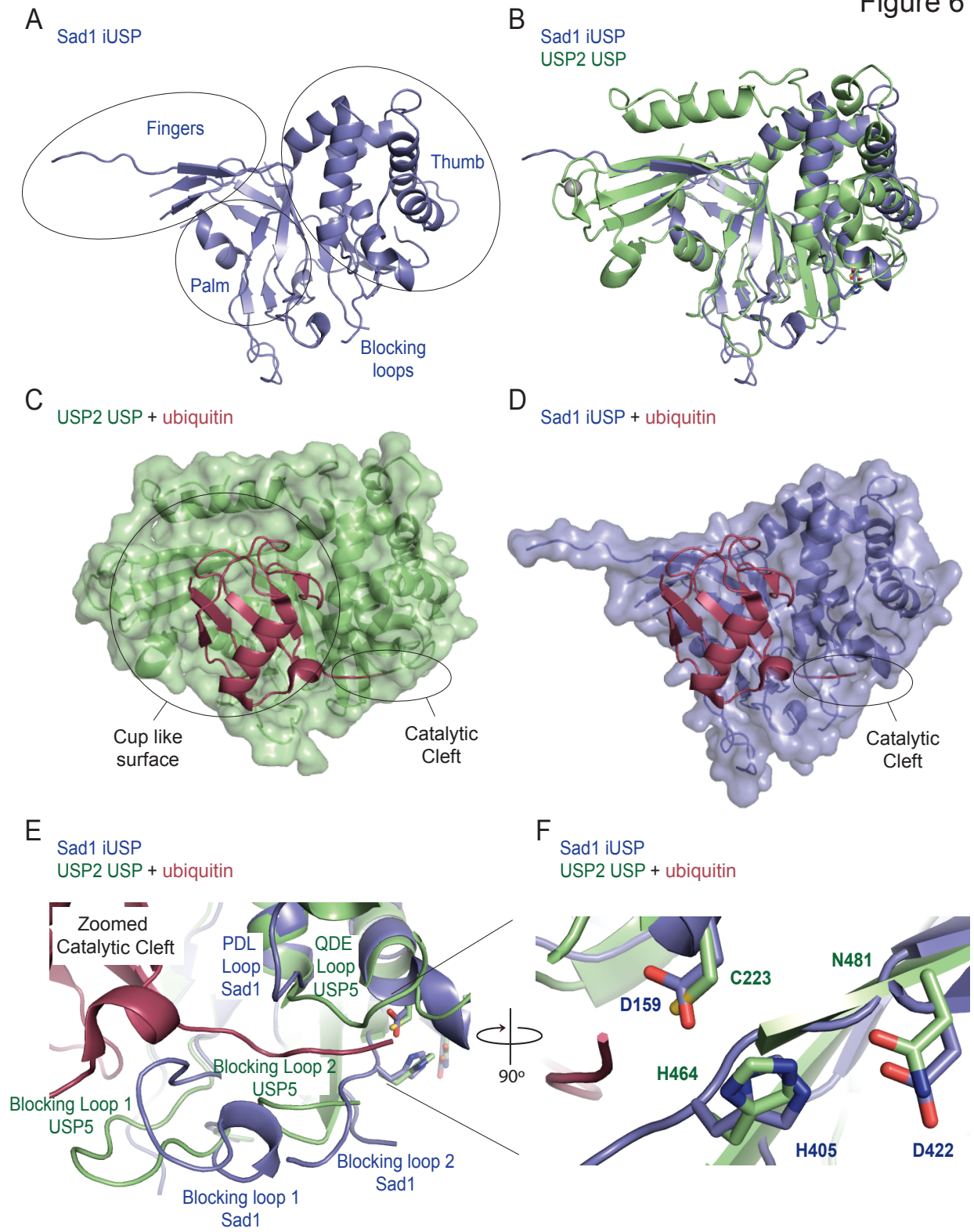


Figure S1

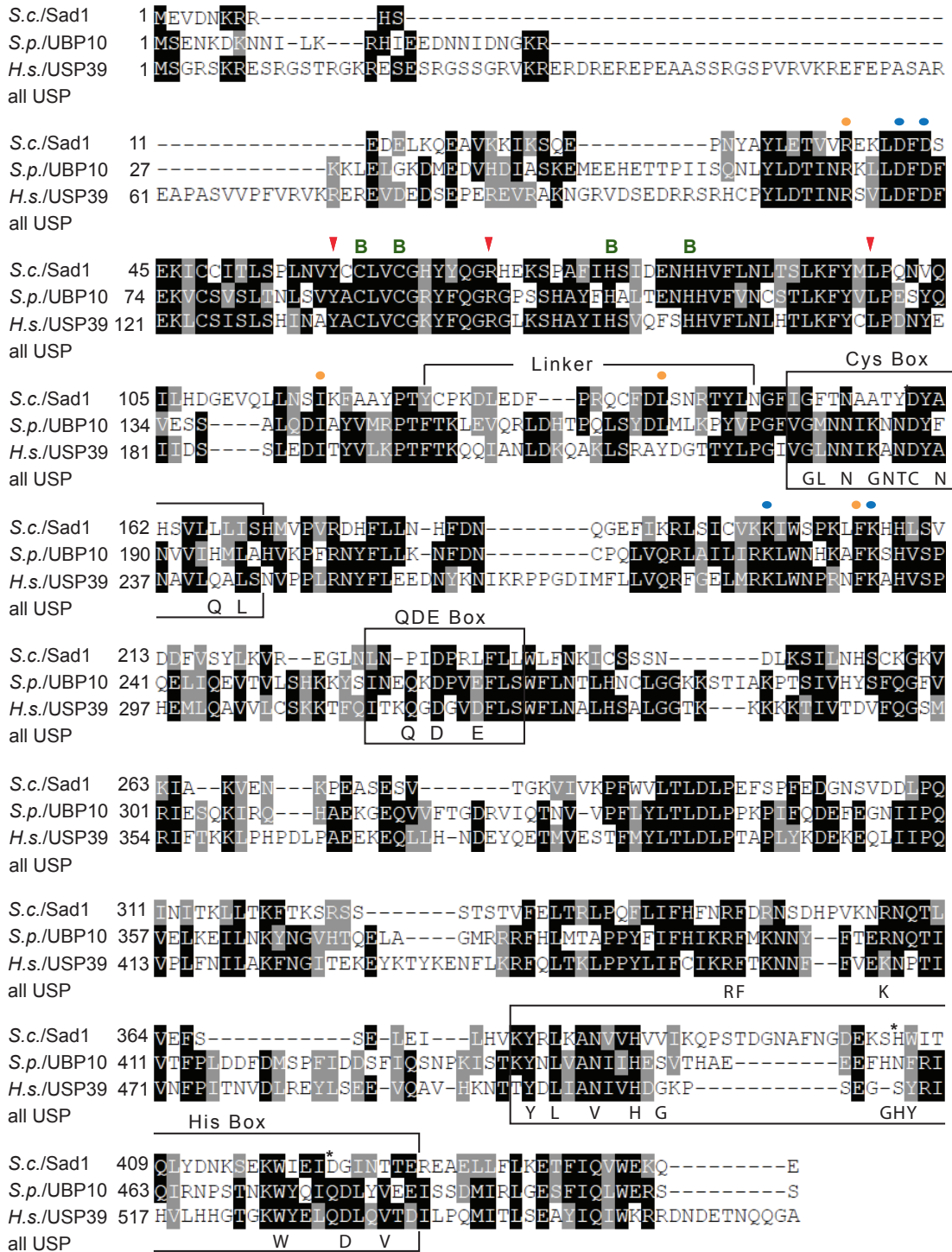


Figure S2

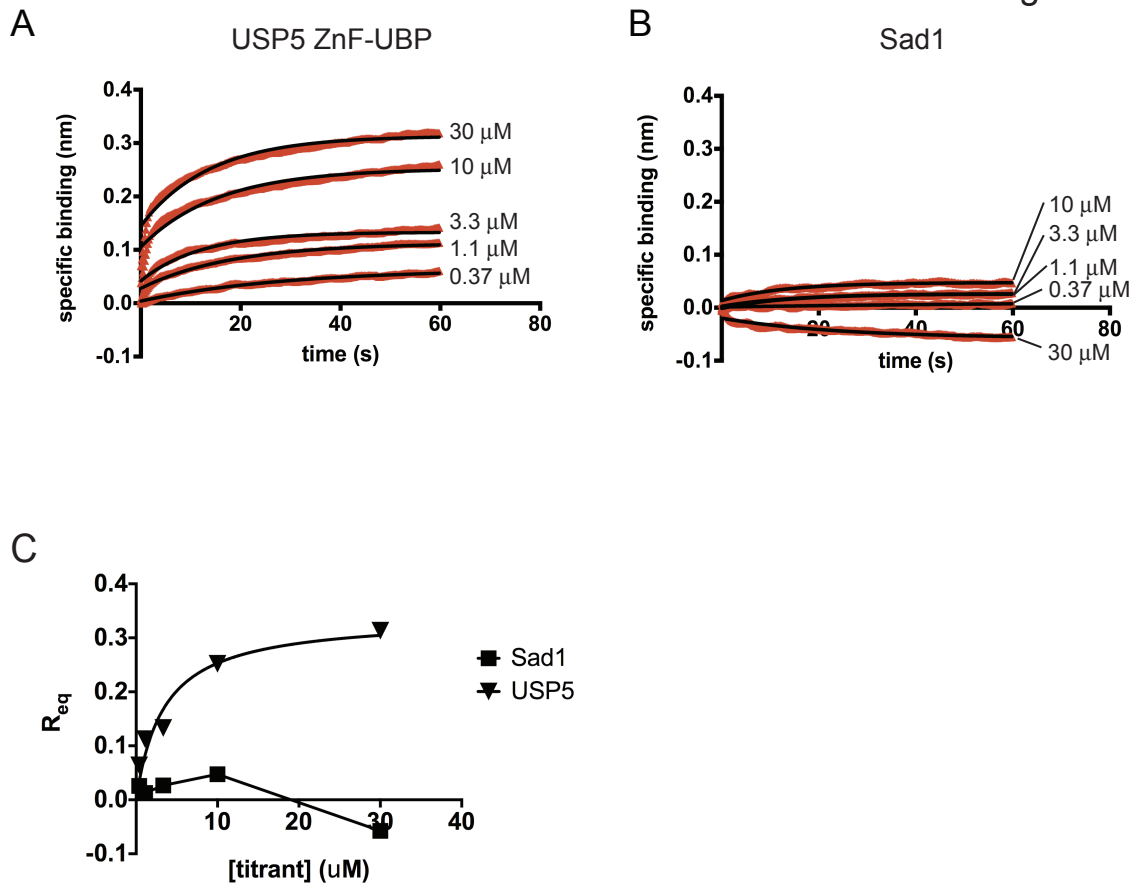
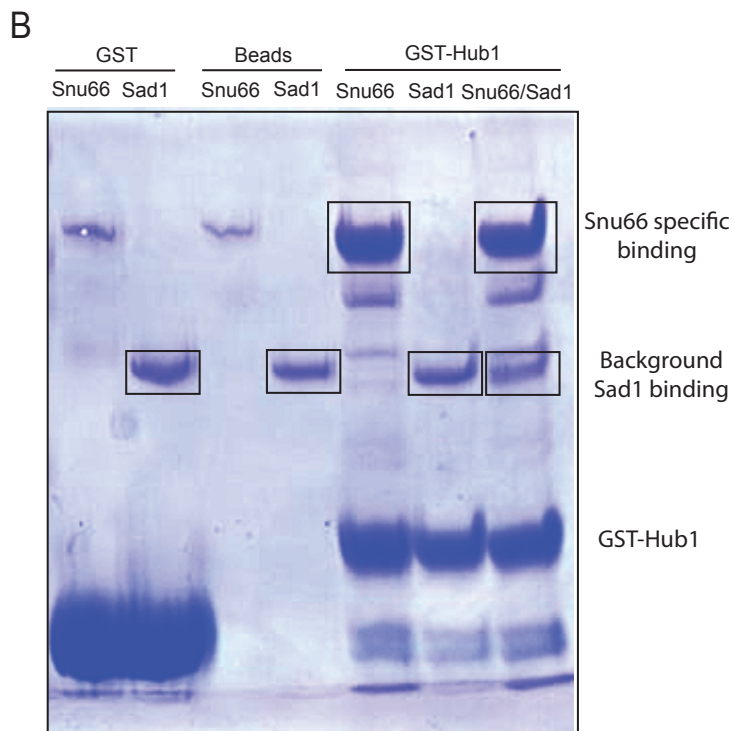
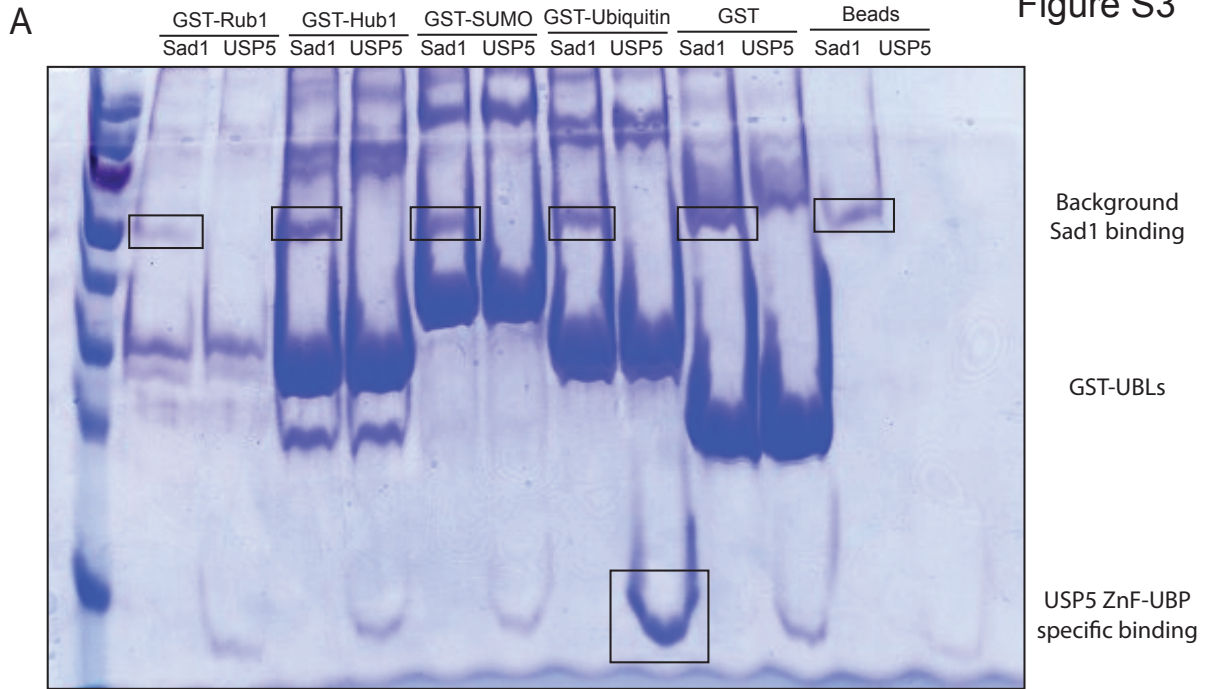


Figure S3



A

Sad1 Molprobity

**Summary statistics**

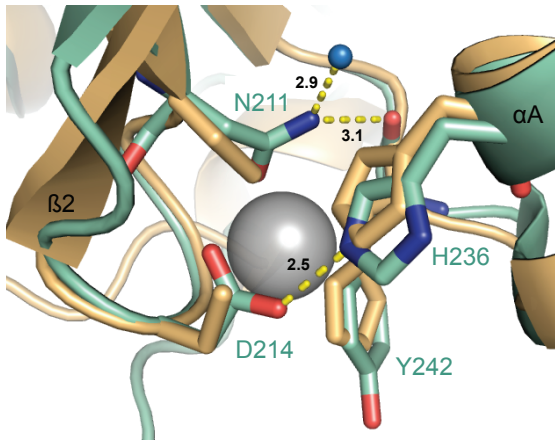
All-Atom Contacts	Clashscore, all atoms:	1.91		100 <sup>th</sup> percentile* (N=778, 1.870Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	0.81%	Goal: <1%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	378	97.67%	Goal: >98%
	MolProbity score <sup>†</sup>	1.03		100 <sup>th</sup> percentile* (N=11964, 1.870Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	0 / 1576	0.00%	Goal: 0%
Bad backbone angles:	0 / 1963	0.00%	Goal: <0.1%	

In the two column results, the left column gives the raw count, right column gives the percentage.

\* 100<sup>th</sup> percentile is the best among structures of comparable resolution; 0<sup>th</sup> percentile is the worst. For clashscore the comparative set of structures was selected in 2004, for MolProbity score in 2006.

<sup>†</sup> MolProbity score combines the clashscore, rotamer, and Ramachandran evaluations into a single score, normalized to be on the same scale as X-ray resolution.

A Sad1 ZnF-UBP  
 USP5 ZnF-UBP



B

USP5 homologues	N211	D214	H236	Y242
<i>H. sapiens</i>	NLWLNLT	DGSIL	CGRRYFDG	SGGNNHAVEHYRE-TGYP
<i>M. musculus</i>	NLWLNLT	DGSIL	CGRRYFDG	SGGNNHAVEHYRE-TGYP
<i>A. thaliana</i>	NLWLNLT	DGMIL	CGRKNWDG	TGGNNHAVEHYKE-TAYP
<i>C. elegans</i>	NLWLNLT	DGAVR	CGRSQFLSDG	KKTNGNGHMQDYFDS-TREP
<i>D. melanogaster</i>	NLWLNLT	DGSIM	CGRKFFDG	SGGNDHAVEHYRV-TGEP
<i>R. norvegicus</i>	NLWLNLT	DGSIL	CGRRYFDG	SGGNNHAVEHYRE-TGYP
<i>S. pombe</i>	NLWMCL	TCGALS	CGRKQYGG	GGGNGHALSHYDD-TGHP
<i>S. cerevisiae</i>	NLWLC	LHCGN	IGCGREQI-G	IDGHSHALDHYRSNNHP
Sad1	NVYCCL	LVCGH	YHQGR-HEKS	PAFIHSID ENHH

## References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 213-221.
- Allen, M. and Bycroft, M. 2007. The solution structure of the ZnF UBP domain of USP33/VDU1. *Protein Sci* **16**: 2072-2075.
- Avvakumov, G.V., Walker, J.R., Xue, S., Allali-Hassani, A., Asinas, A., Nair, U.B., Fang, X., Zuo, X., Wang, Y.X., Wilkinson, K.D. et al. 2012. Two ZnF-UBP domains in isopeptidase T (USP5). *Biochemistry* **51**(6): 1188-1198.
- Bellare, P., Kutach, A., Rines, A., Guthrie, C., and Sontheimer, E. 2006. Ubiquitin binding by a variant Jab1/MPN domain in the essential pre-mRNA splicing factor Prp8p. *RNA* **12**: 292-302.
- Bellare, P., Small, E.C., Huang, X., Wohlschlegel, J.A., Staley, J.P., and Sontheimer, E.J. 2008. A role for ubiquitin in the spliceosome assembly pathway. *Nat Struct Mol Biol* **15**(5): 444-451.
- Bonnet, J., Romier, C., Tora, L., and Devys, D. 2008. Zinc-finger UBPs: regulators of deubiquitylation. *Trends Biochem Sci* **33**: 369-375.
- Chan, S., Kao, D., Tsai, W., and Cheng, S. 2003. The Prp19p-associated complex in spliceosome activation. *Science* **302**: 279-282.
- Cheng, S. and Abelson, J. 1987. Spliceosome assembly in yeast. *Genes Dev* **1**: 1014-1027.



- Dong, A., Ravichandran, M., Schuetz, A., Loppnau, P., Li, Y., MacKenzie, F., Koziaradzki, I., Edwards, A.M., Arrowsmith, C.H., Weigelt, J., Bountra, C., Bochkarev, A., Dhe-Paganon, S., Min, J., Ouyang, H. 2008. Crystal structure of human HDAC6 zinc finger domain.
- Fabrizio, P., McPheeters, D., and Abelson, J. 1989. In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev* **3**: 2137-2150.
- Hasegawa, H. and Holm, L. 2009. Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* **19**(3): 341-348.
- Holton, J. and Alber, T. 2004. Automated protein crystal structure determination using ELVES. *Proc Natl Acad Sci U S A* **101**(6): 1537-1542.
- Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J., Gu, W., Cohen, R., and Shi, Y. 2002. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* **111**: 1041-1054.
- Hu, M., Li, P., Song, L., Jeffrey, P., Chenova, T., Wilkinson, K., Cohen, R., and Shi, Y. 2005. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J* **24**: 3747-3756.
- Ingvarsdottir, K., Krogan, N., Emre, N., Wyce, A., Thompson, N., Emili, A., Hughes, T., Greenblatt, J., and Berger, S. 2005. H2B ubiquitin protease Ubp8 and Sgf11 constitute a discrete functional module within the *Saccharomyces cerevisiae* SAGA complex. *Mol Cell Biol* **25**: 1162-1172.
- JCSG, J.C.f.S.G. 2004. Crystal structure of Ubiquitin carboxyl-terminal hydrolase 6 (yfr010w) from *Saccharomyces cerevisiae* at 1.74 Å resolution.

- Jurica, M.S. and Moore, M.J. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**(1): 5-14.
- Kohler, A., Pascual-Garcia, P., Llopis, A., Zapater, M., Posas, F., Hurt, E., and Rodriguez-Navarro, S. 2006. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitylation through its interaction with Ubp8 and Sgf11. *Mol Biol Cell* **17**(10): 4228-4236.
- Kohler, A., Zimmerman, E., Schneider, M., Hurt, E., and Zheng, N. 2010. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* **141**(4): 606-617.
- Komander, D., Clague, M., and Urb√©, S. 2009. Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* **10**: 550-563.
- Korneta, I., Magnus, M., and Bujnicki, J.M. 2012. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res* **40**(15): 7046-7065.
- Lee, K., Florens, L., Swanson, S., Washburn, M., and Workman, J. 2005. The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex. *Mol Cell Biol* **25**: 1173-1182.
- Lin, R.J., Newman, A.J., Cheng, S.C., and Abelson, J. 1985. Yeast mRNA splicing in vitro. *J Biol Chem* **260**(27): 14780-14792.
- Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**(10): 953-961.

- Luna-Vargas, M.P., Faesen, A.C., van Dijk, W.J., Rape, M., Fish, A., and Sixma, T.K. 2011. Ubiquitin-specific protease 4 is inhibited by its ubiquitin-like domain. *EMBO Rep* **12**(4): 365-372.
- Lygerou, Z., Christophides, G., and Skaraphin, B. 1999. A novel genetic screen for snRNP assembly factors in yeast identifies a conserved protein, Sad1p, also required for pre-mRNA splicing. *Mol Cell Biol* **19**: 2008-2020.
- Maeder, C., Kutach, A., and Guthrie, C. 2009. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat Struct Mol Biol* **16**: 42-48.
- Makarova, O., Makarov, E., and Lührmann, R. 2001. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J* **20**: 2553-2563.
- Mishra, S., Ammon, T., Popowicz, G., Krajewski, M., Nagel, R., Ares, H., and Tack, J. 2011. Role of the ubiquitin-like protein Hub1 in splice-site usage and alternative splicing. *Nature* **474**: 173-178.
- Ouyang, H., Ali, Y., Ravichandran, M., Dong, A., Qiu, W., MacKenzie, F., Dhe-Paganon, S., Arrowsmith, C., and Zhai, R. 2011. Protein aggregates are recruited to aggresome by histone deacetylase 6 via unanchored ubiquitin C termini. *J Biol Chem* **287**: 2317-2327.
- Pai, M., Tzeng, S., Kovacs, J., Keaton, M., Li, S., Yao, T., and Zhou, P. 2007. Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *J Mol Biol* **370**: 290-302.

- Pleiss, J., Whitworth, G., Bergkessel, M., and Guthrie, C. 2007. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol* **5**: e90.
- Renatus, M., Parrado, S., D'Arcy, A., Eidhoff, U., Gerhartz, B., Hassiepen, U., Pierrat, B., Riedl, R., Vinzenz, D., Worpenberg, S. et al. 2006. Structural basis of ubiquitin recognition by the deubiquitinating protease USP2. *Structure* **14**: 1293-1302.
- Reyes-Turcu, F., Horton, J., Mullally, J., Heroux, A., Cheng, X., and Wilkinson, K. 2006. The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* **124**: 1197-1208.
- Reyes-Turcu, F., Shanks, J., Komander, D., and Wilkinson, K. 2008. Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *J Biol Chem* **283**: 19581-19592.
- Samara, N., Datta, A., Berndsen, C., Zhang, X., Yao, T., Cohen, R., and Wolberger, C. 2010. Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* **328**: 1025-1029.
- Samara, N., Ringel, A., and Wolberger, C. 2012. A role for intersubunit interactions in maintaining SAGA deubiquitinating module structure and activity. *Structure* **20**: 1414-1424.
- Song, E., Werner, S., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J., Elledge, S., Kirschner, M., and Rape, M. 2010. The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev* **24**: 1434-1447.

- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. 2002. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* **9**(1): 31-44.
- Wahl, M.C., Will, C.L., and Luhrmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**(4): 701-718.
- Ye, Y., Scheel, H., Hofmann, K., and Komander, D. 2009. Dissection of USP catalytic domains reveals five common insertion points. *Mol Biosyst* **5**: 1797-1808.
- Zhang, Y., Zhou, C., Zhou, Z., Song, A., and Hu, H. 2011. Domain analysis reveals that a deubiquitinating enzyme USP13 performs non-activating catalysis for Lys63-linked polyubiquitin. *PLoS One* **6**: e29362.

## 4. Toggling and the Spliceosome

### Motivation

This chapter will summarize the last project that I worked on with other members of the lab, a collaboration formerly known as Team Toggle. This project stemmed from a hypothesis generated by John Abelson and developed further by the group. The hypothesis challenged an existing conformational model for the spliceosome, which stated that the spliceosome had two distinct conformational states, one for the first catalytic step, 5' splice site cleavage, and the other for the second catalytic step, 3' splice site cleavage. This model was supported by a series of alleles predominately in Prp8, but also in other splicing factors, which when present biased the spliceosome to primarily first or second step activity (Query and Konarska 2004; Konarska and Query 2005; Liu et al. 2007). John Abelson's hypothesis challenged the idea that there were two distinct conformational states for catalysis. Instead, he was that what we called first step mutants actually biased the spliceosome to a silent state of the spliceosome and that second step mutants biased the spliceosome to a catalytic conformation, which he proposed, would be capable of carrying out both catalytic steps of splicing. At the time, the idea was inspired by the crystallization of the RNaseH domain of Prp8, a region which contained alleles for both first and second step mutants. Interestingly it was revealed that these residues mapped to a unique flexible 15 amino acid beta-hairpin insertion (Yang et al. 2008). Structural determination of representative first and second step alleles revealed that first step alleles stabilized the hairpin form and second step alleles stabilized a rearrangement of the hairpin into a loop form (Schellenberg et al. 2013). By Charles Query and Magda Konarska's model, this suggested that in the

transition of the spliceosome from the first to the second catalytic step, Prp8 would toggle from the hairpin to the loop form. But John Abelson was intrigued by a previous observation that certain Prp8 toggle mutants could suppress the U4-cs1 mutation, an allele that imposes a block to splicing at a step earlier than the first step, suggesting that the Prp8 toggle took place at earlier points as well. It was also around this time that the crystal structures of a self-splicing Group II intron were captured during the first catalytic step and in an obligate silent state (Marcia and Pyle 2012).

The self-splicing Group II intron has many parallels to pre-mRNA splicing by the spliceosome. As has been shown biochemically for the spliceosome, the crystal structure of the catalytic state of the group II intron clearly showed the coordination of catalytic metals and the silent state was found to be metal free (Toor et al. 2008). Although presumably coincidental, this was intriguing as the loop form of the Prp8 hairpin was associated with a non-catalytic metal binding event, whereas the hairpin form, which by John Abelson's model would be present in the silent state, was found to be metal free. Further, at the 2014 RNA society meeting it was announced that the crystal structure of the group II intron captured in the second catalytic step had been determined and that the structure closely resembled that of the first step catalytic structure. This suggested that group II introns performed both catalytic steps with a similar structural conformation and that the catalytic steps were separated by an obligate, metal free silent state, which was presumably needed to facilitate the removal of the lariat intermediate from the active site and the entry of the 3' splice site for attack by the now free 5' exon.

These structural breakthroughs for the group II intron bore a striking resemblance to recent biochemical progress with the spliceosome. Mapping of the magnesium binding sites of the spliceosome's catalytic snRNA, U6, showed that the binding sites were virtually identical for the first and second catalytic steps, suggesting that like the group II intron, the spliceosome is in a similar conformation during each step (Fica et al. 2013). Additionally, studies on the interactions of the Bact spliceosome with Cwc25 provided evidence for a silent and a catalytic state of the spliceosome. SmFRET of a spliceosomal bound pre-mRNA showed that prior to binding Cwc25 for the first catalytic step, the branchpoint adenosine and the 5' splice site were apart from one another and experienced only transient interactions, suggestive of a non-catalytic or silent state. Then upon the addition of Cwc25 these interactions became well stabilized and the first step of splicing efficiently occurred, indicative of having attained a catalytic conformation (Krishnan et al. 2013). Further support for a silent and catalytic state of the spliceosome comes from unpublished smFRET work from John Staley. Also using smFRET, John Staley's lab showed that Prp16 was able to reverse the Cwc25 imposed stabilization. This was presumably to allow for the ejection of the product of the first step, the branchpoint lariat. As well, it was observed that after this dissociation, the 3' splice site was able to stably associate with the 5' splice site for the second catalytic step, but did so only in the presence of the second step factors Slu7 and Prp18.

What emerges from these varied lines of inquiry is a model for the spliceosome, where like the Group II intron, there is toggling between two states: a silent state where the catalytic groups are separated from one another, and a catalytic state, where the catalytic groups are stably associated with one another, ie. the 5' splice site and the



branchpoint adenosine for the first step and the free 5' exon and the 3' splice site for the second step. Unlike the group II intron, the spliceosome relies upon a whole host of proteinaceous factors, which help either stabilize or destabilize one conformation. Thus features such as the Prp8 hairpin toggle may play an important role in determining the state of the spliceosome and thus its activity. Figure 1 provides a summary of differences between the John Abelson's and the Charles Query and Magda Konarska model. John Abelson's hypothesis also accounts for other local toggle events, which could impact the probability of making a global transition between the catalytic and silent states. These local toggle events include the U2 stem IIC to stem IIA, the U4/U6 duplex to the U6 intra-stem loop and principally for the focus of team toggle, the RNaseH hairpin to loop toggle, wherein it is believed that the hairpin structure stabilizes the silent state and the loop structure stabilizes the catalytic state.

We may think of these structures as a closed one for the catalytic and a more open one for the silent state, but until further structural information emerges these are just guesses. It is also unclear in which state proofreading takes place although it has been shown that in the same unpublished paper discussed previously from Staley, that Prp16 and Prp22 act on the catalytic form of the spliceosome. Separately, it has also been shown that both ATPases associate with the spliceosome coincident with the accessory factors, Cwc25, Slu7 and Prp18, which stabilize the catalytic state of the spliceosome for the first and second chemical steps (James et al. 2002; Tseng et al. 2011).

## Design

Team toggle wrote a grant for submission to the NSF and the NIH, which sought to prove or disprove John Abelson's toggle hypothesis. The aims of the grant, were to firstly comprehensively test all Prp8 toggle alleles against splice site mutants using the CUP reporter assay. The second aim was to test the hypothesis that the toggling of the hairpin acted in concert with other conformational toggles in the spliceosome, for instance the U2 snRNA helix rearrangements and U6 intra-stem loop formation. The third aim sought to demonstrate that the activity of the first and second step accessory factors, Cwc25 and Slu7/Prp18 respectively, were determined by the equilibrium between the silent and catalytic states of the spliceosome and which would be influenced by the presence of a Prp8 toggle mutant. My contribution to the team was to make various toggle strains of which I hope the double hairpin alleles I designed are the most useful. Additionally I designed new experiments to complement those outlined in the grant.

These experiments seek to obtain direct evidence, which would distinguish between the two proposed models to describe the conformation dynamics of the spliceosome. Charles Query and Magda Konarska's model states that the spliceosome has two distinct conformations for each catalytic step of the spliceosome and that the hairpin toggle alleles bias the spliceosome towards the first step catalytic conformation and the loop toggle alleles bias the spliceosome towards the second step. Conversely, the John Abelson's model states that there is one catalytic state for both steps of splicing, which is biased by loop toggle alleles, and that there is another state, a silent state biased by hairpin toggle alleles, which exists before and after the catalytic steps. I

propose three broad experimental strategies towards distinguishing between these two models. The first is to repeat Ramya Krishnan's smFRET work on Cwc25 in the context of the toggle mutants. The second is to demonstrate that Prp8 toggle alleles can impact the dependence of Cwc25 and Slu7/Prp18. The third is to carefully measure the rate of splicing using the split exon assay, which allows for the monitoring of each catalytic step separately

### **SmFRET of the toggle mutants at the first and second steps**

To directly observe the contributions of the toggle alleles to the conformational state of the spliceosome, I propose using toggle allele strains and repeating the smFRET experiments used to monitor the Cwc25 mediated activation of the first step. Spliceosomes would be assembled upon UBC4 pre-mRNAs fluorescently labeled at the branchpoint and the 5' splice site. They would be incubated with splicing extracts prepared from a Prp2-1 ts strain combined with either wild type, or Prp8 toggle alleles. Inactivation of Prp2-1 blocks the splicing cycle at the B<sup>act</sup> stage just after unwinding of the U4/U6 duplex (Lardelli et al. 2010). In the original studies, the B<sup>act</sup> spliceosome displayed a stable low FRET state. When Prp2 was added, which advances the B<sup>act</sup> spliceosome to the B<sup>act\*</sup> spliceosome, mid to high FRET states were briefly observed but primarily a low FRET state was observed, indicative of dynamic silent state. Upon the addition of Cwc25 the high FRET states became stabilized. This suggested that the branchpoint adenosine and the 5' splice site could interact in the silent state, but did not stably associate until Cwc25 was present when the catalytic state was stabilized. This interpretation is borne out in biochemical studies from Reinhard Luhrmann lab, which

showed that in the absence of Cwc25 there is a small amount of first step activity in vitro and that this is then increased dramatically upon the addition of Cwc25 (Warkocki et al. 2009).

Using these observations as a starting point we would expect mutually exclusive results for the different classes of toggle alleles using the branchpoint adenosine to 5' splice site smFRET reporter. In Charles Query and Magda Konarska's model, where there are two catalytic conformations to the spliceosome, the hairpin or first step alleles of Prp8 would be expected to stabilize the spliceosome in a state competent for the first catalytic step that is with the catalytic groups proximal to one another. Therefore, even in the absence of Cwc25, with fluorescent reporters near the branchpoint adenosine and the 5' splice site, we would expect to see stable high FRET states. This would be in contrast to what was initially observed with the wild-type spliceosome, which in the absence of Cwc25 was stuck in a non-reactive silent state characterized by a predominantly low FRET state (Krishnan et al. 2013). Similarly, by the John Abelson's model we would expect a low FRET state for the hairpin toggle allele, as his model states that *instead* of stabilizing a conformation competent for the first catalytic step, the hairpin toggle allele stabilizes the silent state, wherein the catalytic groups are separated from one another. Charles Query and Magda Konarska's model, does not suggest a clear FRET prediction for the loop, or second step alleles of Prp8 as by their model, they would be expected to bias the spliceosome to a conformation competent for the second step and there is no obvious way to rationalize how this would impact the first step. Although it should be mentioned that the loop toggle alleles are capable of carrying out the first step in vivo and in vitro (Liu et al. 2007). On the contrary, John

Abelson's model does provide a clear prediction, as by his model, the loop toggle allele should stabilize the catalytic conformation and thus we would expect a stabilization of the high FRET state even in the absence of Cwc25, a result if observed might then allow bypass of Cwc25, or at least decrease the dependence upon it. Finally these experiments could be performed in parallel with a pre-mRNA containing fluorescent reporters near the 5' splice site and the 3' splice site to monitor the impact of the toggle alleles on the second step of splicing in the absence of second step factors. Although the ability to distinguish between the models is diminished, for in the case of the second step, both models would predict a high FRET state for the loop toggle alleles. And again, whereas John Abelson's model provides a clear prediction that the hairpin toggles alleles would have a low FRET state, Charles Query and Magda Konarska's model does not.

### **Interactions with Accessory factors**

Building off of the same logical introduction presented above, this second idea is really a complement to the previous one, but relies on in vitro splicing assays instead of smFRET. The smFRET study of the Bact\* spliceosome's interaction with Cwc25 by Ramya shows that Cwc25 must be present in order to stabilize the association of the branchpoint adenosine with the 5' splice site and consequently to get efficient first step activity. Thus Cwc25 shifts the equilibrium from the silent state of the spliceosome to the catalytic state of the spliceosome. But what if that equilibrium was shifted by the presence of a toggle mutant? If it were harder to exit the silent state, more Cwc25 would be necessary to achieve the same level of splicing observed in a wild-type

spliceosome. Likewise, if it were easier to enter the catalytic state, less would Cwc25 would be necessary. Thus again we are presented with an experimental design that would distinguish between the two competing models. If Charles Query and Magda Konarska's model is correct and hairpin toggle alleles bias the spliceosome towards a catalytic conformation then we would expect less Cwc25 would be necessary for the first step. Conversely, if John Abelson's model is correct and hairpin toggle alleles bias the spliceosome to the silent state, then we would expect more Cwc25 to be present and if the loop allele is present, then the catalytic conformation would be biased and less Cwc25 would be present. In fact although *CWC25* is an essential gene, the presence of a loop toggle allele may allow for bypass its deletion, this is a possibility, which would be checked. It is unclear what Charles Query and Magda Konarska's model would predict for loop toggle alleles within this framework. Finally, a parallel framework could again be used to ask the same question about the relationship between the toggle alleles and the second step factors Slu7 and Prp18.

To perform these experiments immunodepletion or temperature sensitivity would be used to inactivate the accessory factor in question to study either the first or second step. Then a titration of the purified factors would be added back to the extracts and in vitro splicing would be assayed. The first titration would be with wild-type extract to determine the concentration at which 50% of maximum activity is achieved ( $EC_{50}$ ). The toggle allele strains would be treated with a similar titration of the relevant accessory factor and the  $EC_{50}$  would also be determined. If a toggle allele stabilizes the silent state of the spliceosome, then we would expect that the measured  $EC_{50}$  would be greater than that which was measured for a wild-type spliceosome. Oppositely, if a

toggle allele stabilized the catalytic state of the spliceosome, then we would expect the  $EC_{50}$  to be lower than the wild-type. As this approach seeks to study the contributions of the toggle to both the first and second step of splicing, it may be necessary to utilize an established protocol that allows for monitoring of each step of splicing independently. This protocol would also be utilized in the final idea.

### **Kinetics of first and second step**

The final idea is to carefully measure the rate of each step of splicing using the split-exon assay (Schellenberg et al. 2013). The expectation from this approach is that if a toggle allele stabilizes the silent state of the spliceosome, then we would expect that the rate of splicing would be slower than that which was measured for a wild-type spliceosome. Oppositely, if a toggle allele stabilized the catalytic state of the spliceosome, then we would expect that the rate of splicing would be faster than that which was measured for a wild-type spliceosome. Thus the expectation would be that Charles Query and Magda Konarska's model would predict faster activity for the first step for hairpin alleles and slower activity with the loop alleles. Conversely John Abelson's model would predict that the hairpin alleles would slow down both steps relative to wild-type and the loop alleles would speed up both steps. To perform these experiments, spliceosomes would be presented with a pre-mRNA substrate that is lacking a 3' exon. The spliceosome will carry out the first step of splicing on this substrate and will then pause until the 3' exon is added in trans at which point the second step can take place, thereby allowing the measurement of both steps.

# Figures

FRET states 5'ss to BP for the first step and 5'ss to 3'ss for the second step

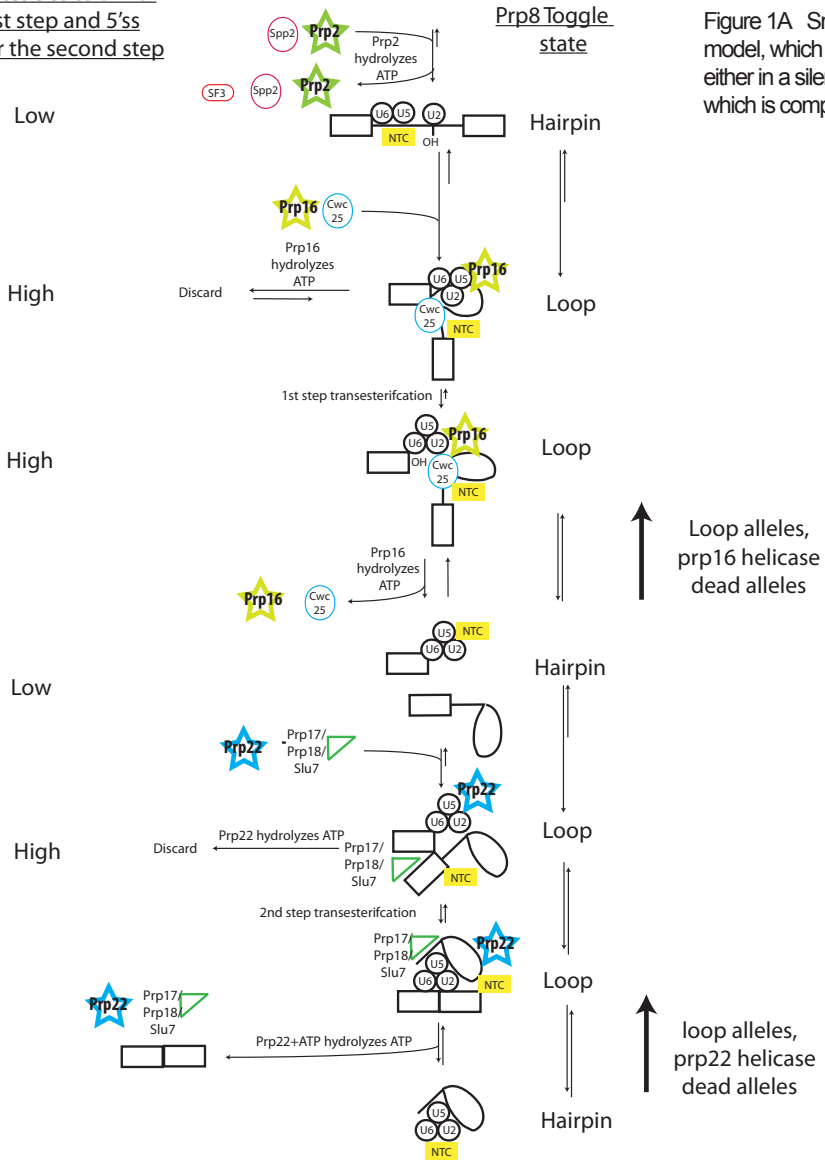


Figure 1A Sm FRET predictions for the Abelson model, which states that the spliceosome is either in a silent or a catalytic state, the latter of which is competent for both steps of splicing.



FRET states 5'ss to BP for the first step and 5'ss to 3'ss for the second step

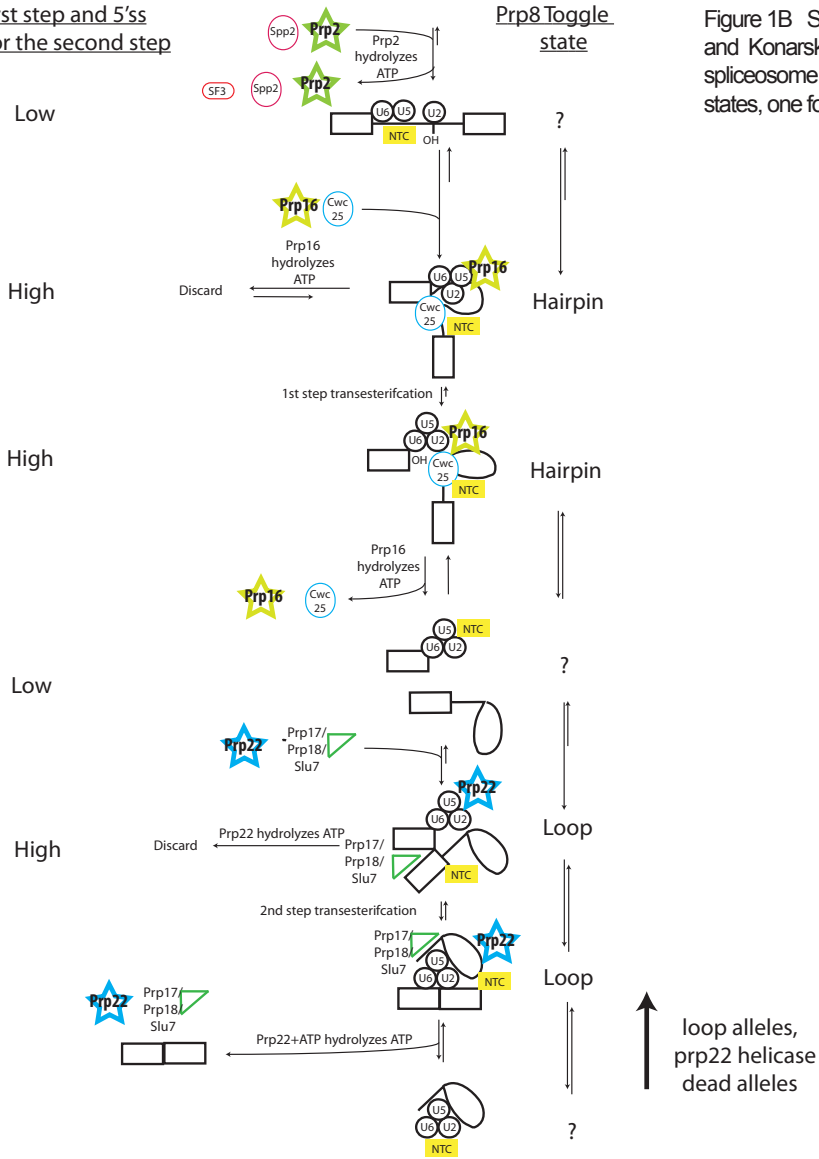


Figure 1B Sm FRET predictions for the Query and Konarska model, which states that the spliceosome exists in two distinct catalytic states, one for each catalytic step.

## References

- Fica, S.M., Tuttle, N., Novak, T., Li, N.S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**(7475): 229-234.
- James, S.A., Turner, W., and Schwer, B. 2002. How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* **8**(8): 1068-1077.
- Konarska, M.M. and Query, C.C. 2005. Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev* **19**(19): 2255-2260.
- Krishnan, R., Blanco, M.R., Kahlscheuer, M.L., Abelson, J., Guthrie, C., and Walter, N.G. 2013. Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing. *Nat Struct Mol Biol* **20**(12): 1450-1457.
- Lardelli, R.M., Thompson, J.X., Yates, J.R., 3rd, and Stevens, S.W. 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**(3): 516-528.
- Liu, L., Query, C.C., and Konarska, M.M. 2007. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat Struct Mol Biol* **14**(6): 519-526.
- Marcia, M. and Pyle, A.M. 2012. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**(3): 497-507.
- Query, C.C. and Konarska, M.M. 2004. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell* **14**(3): 343-354.

- Schellenberg, M.J., Wu, T., Ritchie, D.B., Fica, S., Staley, J.P., Atta, K.A., LaPointe, P., and MacMillan, A.M. 2013. A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat Struct Mol Biol* **20**(6): 728-734.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. 2008. Crystal structure of a self-spliced group II intron. *Science* **320**(5872): 77-82.
- Tseng, C.K., Liu, H.L., and Cheng, S.C. 2011. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**(1): 145-154.
- Warkocki, Z., Odenwalder, P., Schmitzova, J., Platzmann, F., Stark, H., Urlaub, H., Ficner, R., Fabrizio, P., and Luhrmann, R. 2009. Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat Struct Mol Biol* **16**(12): 1237-1243.
- Yang, K., Zhang, L., Xu, T., Heroux, A., and Zhao, R. 2008. Crystal structure of the beta-finger domain of Prp8 reveals analogy to ribosomal proteins. *Proc Natl Acad Sci U S A* **105**(37): 13817-13822.

## 5. Integrated Approach to the Structure of the Spliceosome

### Motivation

Although the study of yeast and human pre-mRNA splicing has generated lists of spliceosomal components, we currently lack structural information and direct protein-protein interaction information for many parts of even the core spliceosome machinery, let alone the fully assembled splicing reaction intermediates. High-throughput yeast two-hybrid analyses of human spliceosomal factors have yielded some information on the connectivity of spliceosomal proteins, but these studies do not provide information on the location or timing of the interactions within the spliceosomal cycle (Hegele et al. 2012). In addition, while limited structural studies have led to crystal structures of isolated proteins and low-resolution EM structures have been determined for some spliceosomal intermediates, these approaches have not yet provided a holistic understanding of how the ensemble of spliceosomal RNAs and proteins assemble and work together to splice a target pre-mRNA. **Thus the spliceosome remains one of the last great macromolecular complexes that has yet to be fully characterized by high-resolution structural approaches.**

A critical path towards obtaining a structure of the spliceosome should employ an integrated systems and structural approach. At its core, the approach would utilize the Integrative Modeling Platform (IMP) software recently developed by the Sali lab, which relies upon two types of data as inputs (Webb et al. 2014). The first type of data provides information on the overall shape and volume of the target complex, for this we would use 2D and 3D electron microscopy in collaboration with Yifan Cheng's lab. The second provides restraints on the organization of the components in the target complex,

both on their shapes and on who their neighbors are. For this we would collaborate with Nevan Krogan's lab to perform immunoprecipitations followed by mass spectrometry (MS), cross-linking followed by MS, and in vivo genetic interaction mapping. Together these data would then be integrated with the IMP software to generate a probabilistic model for the unique arrangement of the components that define the shape and volume of the target complex (Figure 1). Recently, this strategy was successfully used to determine the architecture of two large macromolecular complexes, the 2.5 MDa 26S proteasome (Lasker et al. 2012) and the 50 MDa nuclear pore complex (Alber et al. 2007).

### Design

Our primary targets would be *S. cerevisiae* spliceosomes trapped at every major point in the splicing cycle. We can acquire these trapped intermediates using a battery of spliceosomal mutants and an optimized pre-mRNA substrate developed in Christine Guthrie's lab (Krishnan et al. ; Abelson et al. 2010a; Abelson et al. 2010b). In particular I believe that the best targets for high resolution cryo-EM would be the post-catalytic but pre-ATPase acted upon spliceosomes. Anecdotally we know that Reinhard Luhrmann is pursuing the Bact spliceosome, ie. pre-Prp2. Ramya's smFRET data would suggest that this is a conformation wherein the reactive groups are far apart from one another and we know from the preliminary reports from Reinhard Luhrmann indicate that the species is very flexible and they have had to employ the use of chemical cross-linkers to decrease sample heterogeneity. As an alternative approach, what I propose is to focus on the catalytic state, ie. the C complex, or after Prp2 has acted and after Cwc25 has activated the first catalytic step, which is cleavage of the 5' splice site and formation of

the lariat intermediate. From Ramya and Staley we know that this species is one where the BP and the 5' splice site are stably associated with one another, thus this is possibly a more compact state. Additionally, it is striking that an ATPase is required to exit the catalytic state, but entry to the catalytic state can occur spontaneously with a low frequency in the absence of accessory factors, suggesting that it is easier to exit the silent state than the catalytic state. This suggestion may actually explain via John Abelson's model, why there is still first step activity with hairpin alleles which should bias towards the silent state. Please refer to chapter 3, figure 1, which shows that the ATPases Prp16 and Prp22, act to displace the products and the accessory stabilizing factors after each catalytic step and which highlights the compact nature of the catalytic state

We would purify spliceosome intermediates from splicing extracts made from these mutants and then analyze those intermediates using a wide range of approaches. Negative stain EM reconstructions would be used to define the shape and volume of the spliceosomal intermediates to be modeled and additionally to identify immunolabeled spliceosome components using a library of tagged spliceosomal factors (see below) (Hacker et al. 2008). To obtain information on the components of these complexes, native and chemically cross-linked samples would be analyzed by mass spectrometry to identify complex composition and protein interfaces (Fabrizio et al. 2009; Rappsilber 2011). In addition to identifying the interactions of spliceosomal proteins through chemical cross-linking, we can also employ the use of site-specific cross-linkers in the pre-mRNA and the snRNAs to identify direct interactions of spliceosomal proteins with sequence specific points in the RNAs, and determine how these cross-links change as

the splicing cycle progresses (Rhode et al. 2003). Finally, these data would also provide a stepping stone to pursue conditions that would yield samples suitable for the high resolution cryo-EM techniques established in Yifan Cheng's lab(Li et al. 2013).

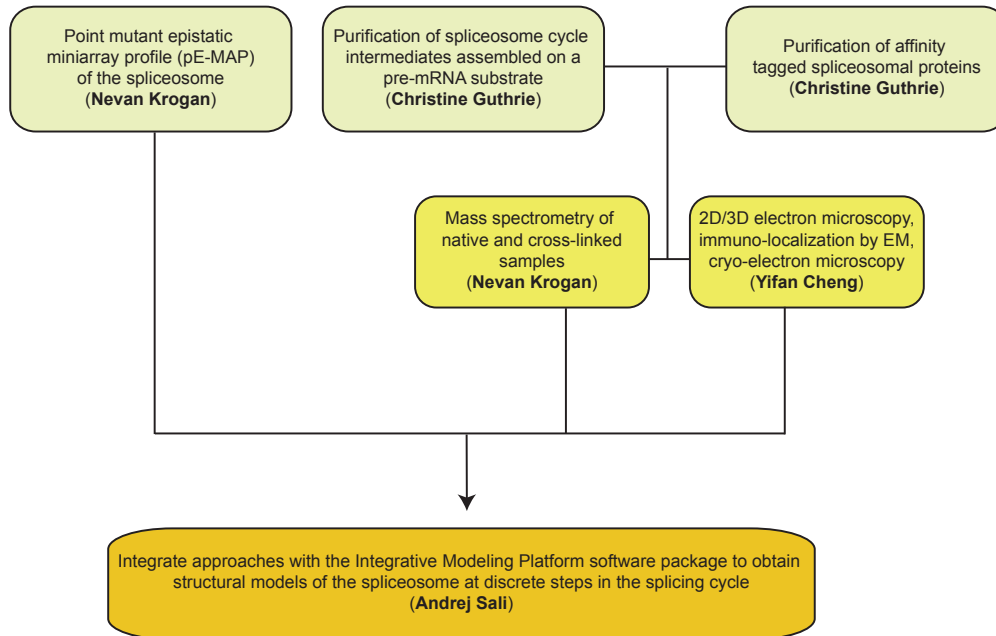
In addition to the interrogation of spliceosomal intermediates, we would employ high-throughput genetic and physical interaction studies of the spliceosome to provide additional information on the connectivity of spliceosomal components. Building on the recent methodological innovations in Nevan Krogan's lab, we would conduct point mutant epistatic miniarray profiles (pE-MAP) for key spliceosomal factors, which would contribute high resolution in vivo protein-protein interaction information (Braberg et al. 2013; Braberg et al. 2014). In addition we would construct fusion-tagged libraries of *S. cerevisiae*, *S. pombe*, and *H. sapiens* spliceosomal proteins (Volkel et al. 2010). Parallel affinity purifications of these spliceosomal factors would then be performed using high-throughput techniques, followed by native and cross-linked mass spectrometry analysis to identify co-purifying spliceosomal proteins. Purifying and analyzing every spliceosomal protein would allow us to identify multiple reciprocal co-purifying proteins, which would point to direct interactions between components, critical for providing neighbor constraints. The reciprocal co-purifications combined with the conserved interactions amongst the different species would also identify ideal sub-complexes whose architecture can be interrogated further with large-scale purifications both from the native species, and multi-protein co-expression strategies in heterologous protein expression systems for high-resolution cryo-EM and X-ray crystallography. Structures of splicing sub-complexes would not only contribute valuable high-resolution

information for integration into the overall structural model of the spliceosome, but would also be of general interest.



## Figures

Figure 1: An integrated approach towards the structure of the spliceosome



## References

- Abelson, J., Blanco, M., Ditzler, M.A., Fuller, F., Aravamudhan, P., Wood, M., Villa, T., Ryan, D.E., Pleiss, J.A., Maeder, C. et al. Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat Struct Mol Biol* **17**(4): 504-512.
- Abelson, J., Hadjivassiliou, H., and Guthrie, C. Preparation of fluorescent pre-mRNA substrates for an smFRET study of pre-mRNA splicing in yeast. *Methods Enzymol* **472**: 31-40.
- Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B.T. et al. 2007. The molecular architecture of the nuclear pore complex. *Nature* **450**(7170): 695-701.
- Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F. et al. 2013. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**(4): 775-788.
- Braberg, H., Moehle, E.A., Shales, M., Guthrie, C., and Krogan, N.J. 2014. Genetic interaction analysis of point mutations enables interrogation of gene function at a residue-level resolution: Exploring the applications of high-resolution genetic interaction mapping of point mutations. *Bioessays*.
- Chesnais, V., Kosmider, O., Damm, F., Itzykson, R., Bernard, O.A., Solary, E., and Fontenay, M. 2014. Spliceosome mutations in myelodysplastic syndromes and chronic myelomonocytic leukemia. *Oncotarget* **3**(11): 1284-1293.

- Ding, J.H., Xu, X., Yang, D., Chu, P.H., Dalton, N.D., Ye, Z., Yeakley, J.M., Cheng, H., Xiao, R.P., Ross, J. et al. 2004. Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* **23**(4): 885-896.
- Echeverria, G.V. and Cooper, T.A. 2014. Muscleblind-like 1 activates insulin receptor exon 11 inclusion by enhancing U2AF65 binding and splicing of the upstream intron. *Nucleic Acids Res* **42**(3): 1893-1903.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. 2009. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* **36**(4): 593-608.
- Hacker, I., Sander, B., Golas, M.M., Wolf, E., Karagoz, E., Kastner, B., Stark, H., Fabrizio, P., and Luhrmann, R. 2008. Localization of Prp8, Brr2, Snu114 and U4/U6 proteins in the yeast tri-snRNP by electron microscopy. *Nat Struct Mol Biol* **15**(11): 1206-1212.
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Luhrmann, R., and Stelzl, U. 2012. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* **45**(4): 567-580.
- Jurica, M.S. and Moore, M.J. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**(1): 5-14.
- Krishnan, R., Blanco, M.R., Kahlscheuer, M.L., Abelson, J., Guthrie, C., and Walter, N.G. Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing. *Nat Struct Mol Biol* **20**(12): 1450-1457.
- Lasker, K., Forster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. 2012. Molecular architecture of the

- 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci U S A* **109**(5): 1380-1387.
- Li, X., Zheng, S.Q., Egami, K., Agard, D.A., and Cheng, Y. 2013. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J Struct Biol* **184**(2): 251-260.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Luhrmann, R., and Wahl, M.C. 2014. Novel regulatory principles of the spliceosomal Brr2 RNA helicase and links to retinal disease in humans. *RNA Biol* **11**(4).
- Rappasilber, J. 2011. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **173**(3): 530-540.
- Rappasilber, J., Ryder, U., Lamond, A.I., and Mann, M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**(8): 1231-1245.
- Rhode, B.M., Hartmuth, K., Urlaub, H., and Luhrmann, R. 2003. Analysis of site-specific protein-RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *RNA* **9**(12): 1542-1551.
- Shea, J.E., Toyn, J.H., and Johnston, L.H. 1994. The budding yeast U5 snRNP Prp8 is a highly conserved protein which links RNA splicing with cell cycle progression. *Nucleic Acids Res* **22**(25): 5555-5564.
- Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**(3): 315-326.

- Volkel, P., Le Faou, P., and Angrand, P.O. 2010. Interaction proteomics: characterization of protein complexes using tandem affinity purification-mass spectrometry. *Biochem Soc Trans* **38**(4): 883-887.
- Wahl, M.C., Will, C.L., and Luhrmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**(4): 701-718.
- Webb, B., Lasker, K., Velazquez-Muriel, J., Schneidman-Duhovny, D., Pellarin, R., Bonomi, M., Greenberg, C., Raveh, B., Tjioe, E., Russel, D. et al. 2014. Modeling of proteins and their assemblies with the Integrative Modeling Platform. *Methods Mol Biol* **1091**: 277-295.

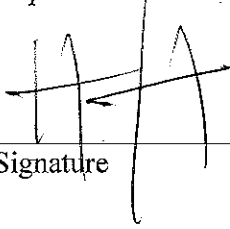
**Publishing Agreement**

*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

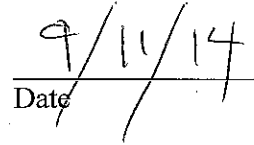
***Please sign the following statement:***

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*

Author Signature

A handwritten signature in black ink, consisting of several vertical strokes and a horizontal line crossing them.

Date

A handwritten date "9/11/14" in black ink, written above a horizontal line.