IDENTIFICATION OF CRITICAL RESIDUES IN SPLICEOSOMAL PROTEIN DIB1

by

Amber Lucas, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biochemistry
August 2014

Committee Members:

Corina Maeder, Chair

Steve Whitten

L. Kevin Lewis

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

I dedicate this work to my family who has supported me through this whole process.

ACKNOWLEDGEMENTS

My time at Texas State University has been a life changing experience. I am so grateful to all the people involved who have molded me into the person I am today. First and foremost, I would like to thank Corina Maeder who has inspired me to push myself and reach beyond my comfort zone to be the best I could possibly be. When I started this program, I just wanted to graduate and get a job, but now I'm continuing on with my education, pursuing my Ph.D., and have bigger dreams for myself. None of this would've been possible without her. All the days when I said I couldn't do it or that I didn't want to do it and she pushed me to anyway, those are the days I am so thankful for.

I am also extremely thankful to my lab mates and friends at Texas State who have supported me through this. All the long nights in lab and long days writing would've been much harder and much duller without everyone to help me get through it.

Finally, I would like to thank my mom and my stepdad, without whom none of this would be possible. I am so lucky to have an amazing set of parents behind me who have supported me both financially and mentally no matter what path I was on. It is thanks to their sacrifices and unconditional love that I have made it this far, and I will forever be grateful

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

In eukaryotes, when DNA is transcribed into pre-messenger RNA, multiple processing events must occur before translation into functional protein occurs. One important processing event is known as splicing. Splicing is the removal of non-protein coding regions, known as introns, from pre-messenger RNA and the ligation of the protein coding regions, known as exons, to form a mature mRNA. This process is extremely important. Recurring inaccuracies in splicing by just a single nucleotide can lead to the shift in a reading frame of a protein coding sequence, which then leads to nonfunctional protein or disease in some cases.

Splicing is facilitated by a large macromolecular complex known as the spliceosome. The spliceosome is composed of five snRNPs (U1, U2, U4, U5, and U6) and many associated proteins which help orient the pre-mRNA in such a way that catalyzes the splicing reactions. Of the associated proteins, Dib1 in *S. cerevisiae* is amongst those whose biochemical function still remains unknown.

Dib1 is a 15 kDa protein with a thioredoxin-like fold that is associated with U5 in the spliceosome. Dib1 is highly conserved from *S. cerevisiae* to *H. sapiens* and is necessary for cell viability. Previous studies have shown that Dib1 and the *S. pombe* homolog Dim1 are necessary for splicing, potentially have peptidase activity, and that many splicing factors interact with Dim1. In this study, the goal was to identify critical residues in Dib1 to help elucidate its biochemical function. In order to do this, mutational analysis, protein purification, and CD spectroscopy were performed.

CHAPTER 1

INTRODUCTION

Deoxyribonucleic acid (DNA) is the molecule that holds all of the genetic information for a single organism. It is composed of four deoxyribonucleotides (adenine, cytosine, guanidine, and thymine) and the arrangement of those four deoxyribonucleotides dictates the genetic information of an organism. In order for this genetic information to result inan observable phenotype, it must be translated into a functional protein. The cell accomplishes this by transcribing DNA into messenger ribonucleic acid (mRNA). mRNA is a messenger that links the genetic instructions in DNA to the expression of protein. RNA is similar in structure to DNA except that it has a 2' hydroxyl group and, where DNA would have a thymine, RNA instead has uracil. Though similar in structure, the reactive 2' hydroxyl group on RNA allows it to be much more versatile. This hydroxyl can act as a nucleophile to catalyze chemical reactions and it can form hydrogen bonds with other molecules or itself to form complex, stable structures. This allows RNA to participate in many biological processes in the cell, not just as mRNA.

Once transcribed, mRNA then is sent to the ribosome where its sequence is read and protein is produced. This flow of genetic information from DNA → RNA → protein is known in molecular biology as the central dogma (Figure 1.1). While these series of events seems simple and linear, in eukaryotes there are multiple processing events that must transpire after transcription before translation can occur. These include 5' capping, pre-messenger RNA (pre-mRNA) splicing, and export out of the nucleus. Of these processes,

pre-mRNA splicing is the most complex and arguably the most important for gene expression regulation.

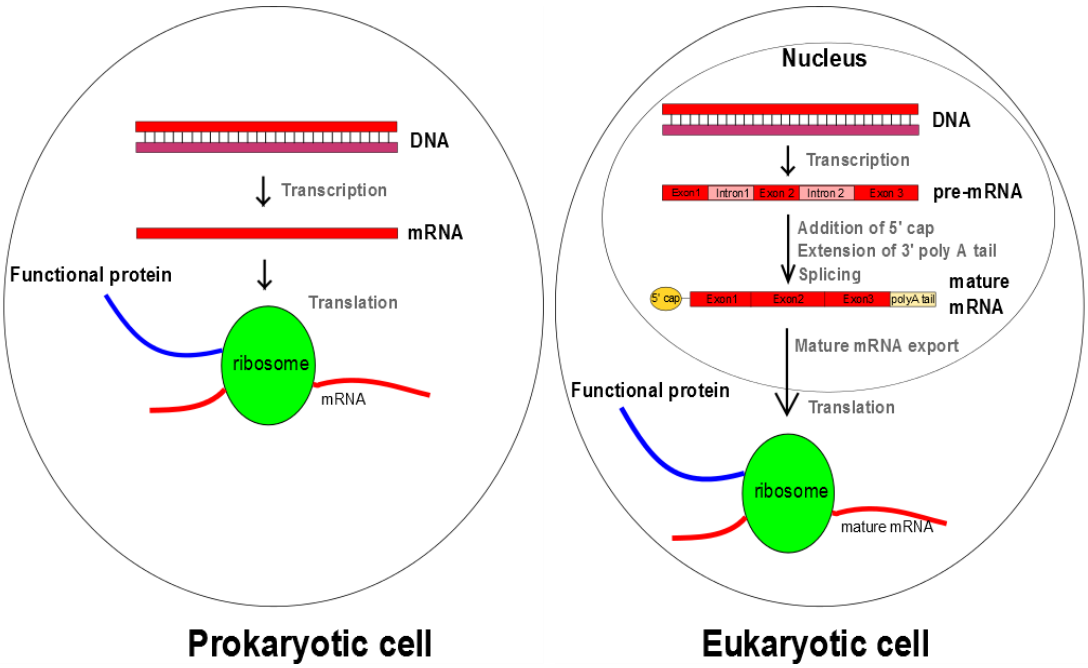

Figure 1.1. The central dogma of molecular biology. The central dogma of molecular biology states that DNA is transcribed into mRNA which is then translated into functional protein, resulting in the phenotypes we observe in organisms. This series of events is more complicated in eukaryotes where multiple processing events must occur between transcription and translation.

1.1 Splicing Overview

In prokaryotes, protein coding region of the mRNA is continuous, however in eukaryotes there are mRNA strands, known as pre-mRNA, that contain large non-protein coding regions that interrupt the protein coding regions. Upon transcription of DNA into pre-mRNA, the non-protein coding regions, known as introns, must be removed and the remaining mRNA fragments, known as exons, must be ligated into one continuous mRNA strand before translation can occur (Figure 1.2). This process is known as pre-mRNA splicing and is a crucial element of post-transcriptional modification and gene regulation.

Accuracy of pre-mRNA splicing must be to single nucleotide precision in order to maintain proper reading frame in a protein coding sequence.

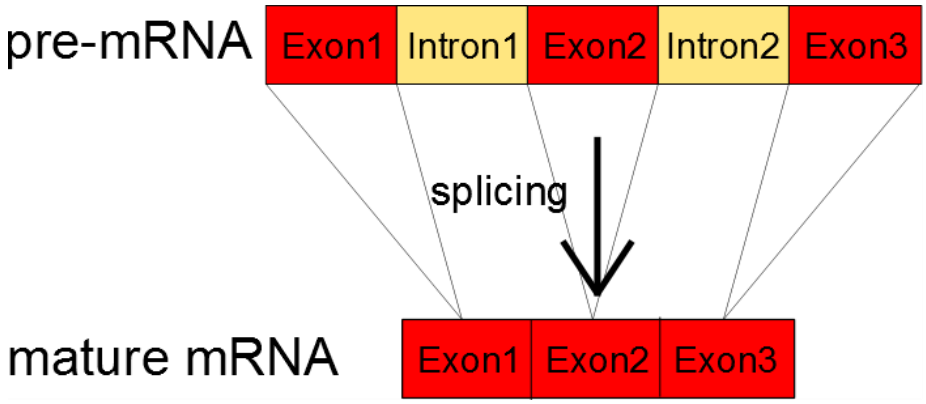

Figure 1.2: Process of pre-mRNA splicing. When pre-mRNA is transcribed from DNA, it contains non-protein coding regions, known as introns, which must be removed before translation can occur.

Pre-mRNA splicing was first seen in experiments using adenovirus-2 RNA. It was observed that the labeled nuclear mRNA of the adenovirus-2 was much longer than the nuclear mRNA. Originally it was thought that this was due to a cleaving event similar to RNase cleavage, producing shorter mRNA fragments that were then transferred to the nucleus.[1] However, after structural studies using an electron microscope showed different length mRNA transcripts in the nucleus composed of different sequence compositions, it was determined that this processing was not a purely a cleavage event, but rather a splicing event.[2,3] In this event, introns were being removed and exons were being ligated in various patterns to form different length products.

Splicing is conserved from *Saccharomyces cerevisiae* to *Homo sapiens,* though there are some large differences. In *H. sapiens* it has been shown through deep sequencing data that more than 95% of the genome contains introns.[4] Out of the entire human genome, exons only account for 1.1% of the genome while introns account for 24%.[5] Out of the more than

200,000 different introns, there is great variation in size, ranging from less than 100 nucleotides to greater than 700,000 nucleotides.[4] Exon lengths are much more consistent than introns, with an average of 123 nucleotides per exon.[4] Typical human genes contain between 7-9 introns, although some genes contain ten times this number.[4] The large size and the large amounts of introns in human genes puts a high importance not only on splicing itself, but also on splicing regulation and accuracy. A recurring splicing error that occurs by even one nucleotide can result in a shift in the reading frame that could result in a dysfunctional protein and, in some cases, disease.[6] Many diseases, such as retinitis pigmentosa, Frasier syndrome, atypical cystic fibrosis, spinal muscular atrophy, and myotonic dystrophy, have been linked to splicing errors.[6]

The genome of yeast species such as baker's yeast *S. cerevisiae* and fission yeast *Schizosaccharomyces pombe* do contain introns, but far fewer and of much smaller size. Out of the ~6,000 *S. cerevisiae* genes there are only 228 introns.[7] Of the 5% of genes that contain introns, most contain only one intron, although genes containing multiple introns have been observed.[7,8] The length of *S. cerevisiae* introns ranges from ~55 nucleotides to ~1000 nucleotides, with the majority of introns being close to 100 nucleotides in length. [7] *S. pombe* has a much higher intron content, with 4,730 introns in ~5,000 genes.[8] In addition to having a much higher intron percentage (43%) than *S. cerevisiae*, it is also more common for *S. pombe* genes to contain more than one intron per gene, although it is still far less common than in humans.[8] The average *S. pombe* intron length is 81 nucleotides, similar to what is seen in *S. cerevisiae.*[8] While there are distinct differences in intron length and distributions between *H. sapiens, S. cerevisiae,* and *S. pombe,* the splicing mechanism itself is conserved in each species.

Splicing of pre-mRNA occurs in two simple S_n2-type reactions in the nucleus (Figure 1.3). In the first reaction, the 2' hydroxyl group of an adenosine at the branch point of the intron acts as a nucleophile and attacks the phosphate group at the 5' splice site. This reaction produces a lariat intermediate consisting of the intron and 3' exon and produces a 5' exon that now has a free hydroxyl group. This free 3' hydroxyl group on the exon then acts as a nucleophile and attacks the phosphate at the 3' splice site, ligating the two exons together to produce a complete mature mRNA consisting of only exons which is exported into the cytoplasm and releasing the intron lariat.

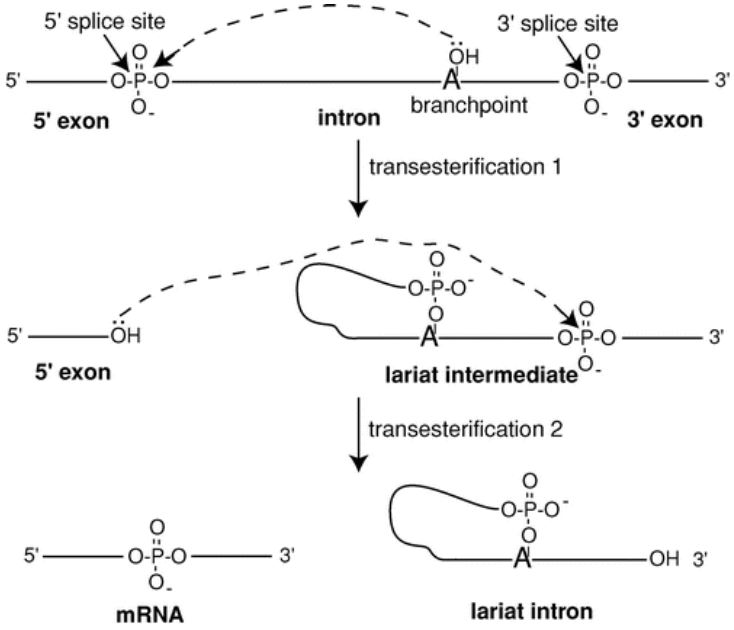

Figure 1.3. The two transesterification reactions in splicing. In the first step, the 2' hydroxyl group of the adenine at the branchpoint acts as a nucleophile and attacks the phosphate at the 5' splice site. This results in the formation of a lariat intermediate and a free hydroxyl group on the 3' end of the 5' exon. This hydroxyl group then acts as a nucleophile and attacks the phosphate group at the 3' splice site, ligating the exons together and releasing the intron. Figure reprinted from D. A. Brow.[9]

1.2 Overview of the Spliceosome

Although the chemistry involved in splicing is simple, it requires the pre-mRNA to be in a very specific orientation for the reaction to occur. The pre-mRNA transcripts can be very long and adopt complex secondary structures, so accomplishing this orientation

spontaneously can be very difficult and in some cases will never happen. Because of this, orientation of the pre-mRNA is facilitated by a large complex known as the spliceosome. The spliceosome is a highly conserved molecular machine composed of five small nuclear ribonucleoproteins (snRNPs) as well as more than 100 associated proteins.[4] Each of the five snRNPs (U1, U2, U4, U5, and U6) are complexes of proteins and small nuclear RNA (snRNA) and these complexes specifically interact with the pre-mRNA to orient it in the proper position for splicing to occur.

The biogenesis of the snRNPs in humans begins with transcription of each snRNA. U1, U2, U4, and U5 snRNAs are transcribed by RNA polymerase II and co-transcriptionally an N7-monomethylguanosine (m^7G) cap is added.[10] These snRNAs are then transferred to the cytosol where they associate with Sm proteins G, F, E, D1, D2, D3, B, and B' which bind the uridine rich region of the snRNA (Figure 1.4). Association of the snRNA with the Sm proteins forms the core of the snRNP and triggers hypermethylation of the m^7G cap to form a 5'-2,2,7-terminal trimethylguanosine (m^3G) cap by Tgs1 methyltransferase.[10,11] After this cap is added and the 3' end of the snRNP is processed, a transport protein known as Snuportin recognizes the snRNP core complex and transports it back into the nucleus.[12] After they enter the nucleus, these snRNPs aggregate in Cajal bodies where small Cajal body specific RNAs (scaRNAs) direct the pseudouridylation and 2'-O-methylation of U1, U2, U4, and U5 snRNAs.[13,14] Many of the snRNP associated proteins accumulate in these Cajal bodies, and it is hypothesized that many of these proteins associate with the core complex here before moving to nuclear speckles where they are stored.[11,15]

The biogenesis of U6 is very different than the other snRNPs. U6 snRNA is transcribed by RNA pol III, not RNA pol II as with the other snRNAs.[11] After synthesis, the uridine rich 3' end of the U6 snRNA binds the La protein, which is quickly replaced by a heteromer of proteins Lsm2, Lsm3, Lsm4, Lsm5, Lsm6, Lsm7, and Lsm8.[11,16] This heteromer first directs U6 snRNA to the nucleolus where pseudouridylation and 2'-O-methylation is facilitated by small nucleolar RNAs. [11,17] After modification, the U6 core complex is directed to the Cajal bodies, where it is believed that protein complex SART3/p110 coordinates the association of U4 and U6 snRNAs.[11,18] Once U4 and U6 associate, they then associate in nuclear speckles with the other snRNPs until splicing occurs.[11]

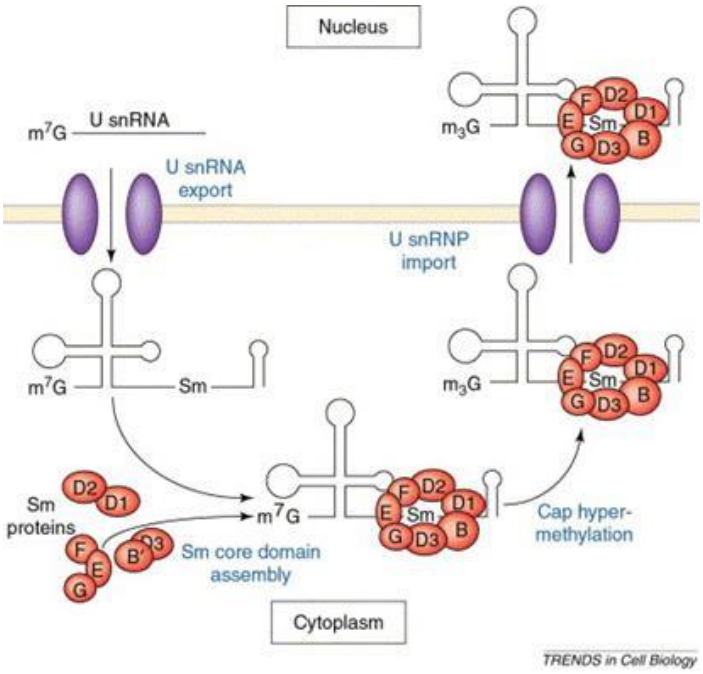

Figure 1.4. Post transcriptional modification of U snRNA. After transcription, U snRNA is exported into the cytoplasm where it associates with Sm proteins. After associate of these proteins, the m^7g cap is hypermethylated to an m^3g cap, and the complex is imported back into the nucleus where the snRNPs complete their biogenesis. Figure reprinted from Meister *et al.*.[19]

Once the snRNPs are assembled, they interact with one another to form the spliceosome. In the first step of spliceosome assembly, the 5' end of U1 snRNA base pairs with the 5' splice site by recognizing the consensus sequence at this location (Figure 1.5). In higher eukaryotes, the consensus sequence is AGGURAGU (where R designates any purine).[20] Other than the GU at the 5' end of the splice site which is almost 100% conserved, variation in this consensus sequence is tolerable for the most part.[21] In *S. cerevisiae*, the 5' consensus sequence GUAYGU is much more conserved, with very little allowance for variation.[21,22] Next, U2 snRNA base pairs with the branch point in the intron by recognizing the consensus sequence at this location. In higher eukaryotes, the branch point is designated by the sequence YUNAY (where Y is any pyrimidine and N is any nucleotide) and allows for great variation with the exception of the adenine.[23] In *S. cerevisiae* the consensus sequence at the branch site is UACUAAC and is much more strictly conserved than in higher eukaryotes.[23] The binding of U2 causes a conformational change in which the adenine at the branchpoint is bulged out, putting it in a more favorable orientation for nucleophilic attach.[24] This complex of U1, U2, and pre-mRNA is what is known as the pre-spliceosome. Binding of U1 and U2 in the formation of the pre-spliceosome is ATP independent and spontaneous.[25] The previously assembled triple snRNP consisting of U4/U6·U5 then joins the complex. The U4/U6 complex is unwound by the helicase Brr2 and the U1/5' splice site complex is unwound by the helicase Prp28.[26,27] The free U6 then displaces U1 at the 5' splice site, releasing U1 and U4 from the complex. This forms the catalytic spliceosome and at this point the pre-mRNA is in the proper orientation for splicing to occur. After the two transesterification reactions take place, the mature mRNA, the intron lariat, and snRNPs are released and recycled in an ATP dependent fashion.[28]

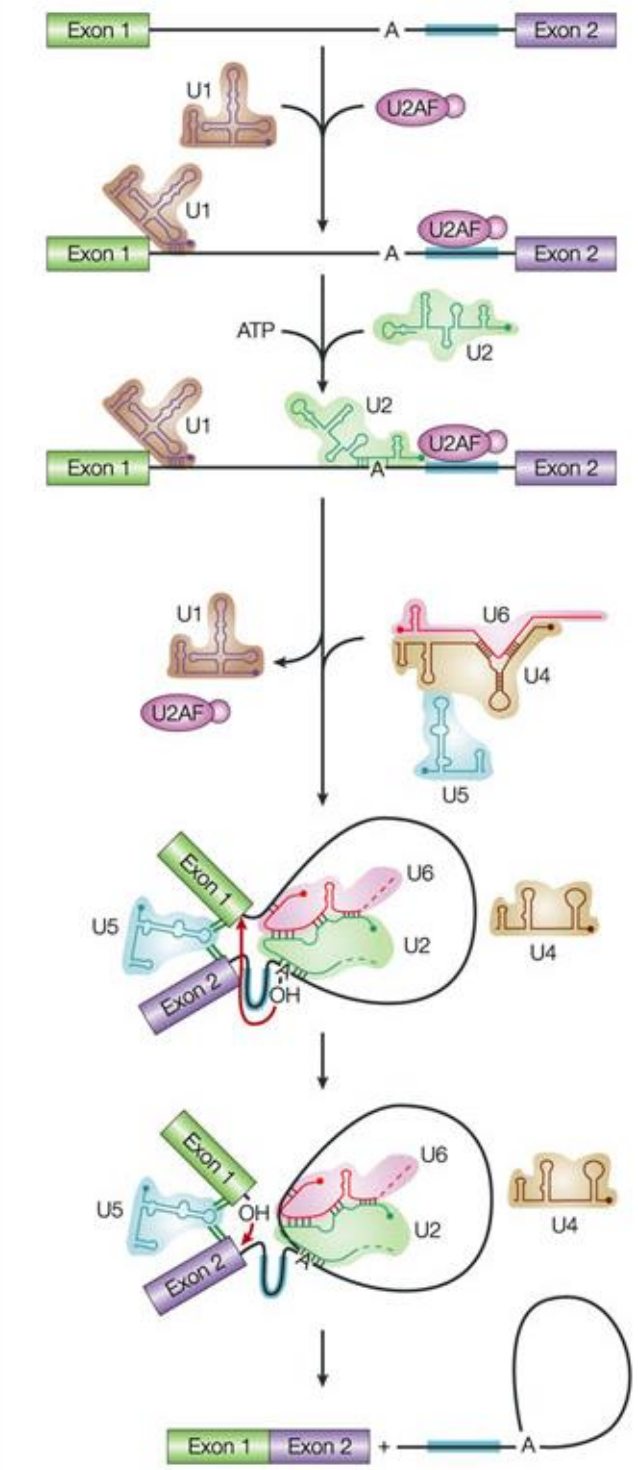

Figure 1.5. The splicing cycle in humans. In the first step, U1 base pairs to the 5' splice, followed by U2 base pairing at the branchpoint. Following this step, the tri-snRNP joins the complex followed by a rearrangement that releases U1 and U4 from the complex. This then forms the catalytically active spliceosome and the two transesterification reactions occur. The mature mRNA, introns, and spliceosome are then released. Figure reprinted from Schroeder *et al.*[29]

The correct and precise excision of introns in pre-mRNA splicing is critical for the cell. Small mistakes can have large effects on protein function and cell viability. The accuracy of intron removal and regulation of the splicing cycle is controlled not only by the accurate base pairing of U1 and U2 to the 5' splice site and branch point, respectively, but also by many splicing factors and regulatory proteins. While the snRNPs and their role in the splicing cycle are conserved throughout eukaryotes, splicing factors and regulatory proteins are much less conserved and more variable. In *S. cerevisiae*, the much stricter consensus sequences at the splice sites and branch points require less help from splicing factors and regulatory proteins to ensure accuracy, while in *S. pombe* and other higher eukaryotes the much more degenerate consensus sequences require more assistance to ensure accuracy.[30] In fact, many proteins like SR proteins and heterogeneous ribonucleoprotein particles (hnRNPs) that are useful in splicing regulation in humans are not conserved in *S. cerevisiae*.[31] Because 85% of the *S. cerevisiae* splicing elements are conserved in higher eukaryotes, it has been hypothesized that these splicing elements make up the catalytically conserved core of the spliceosome.[32] The fact that *S. cerevisiae* splicing elements are conserved, coupled with the fact that less than 5% of its genes contain introns and the introns are smaller in size, make it a simplistic model that is useful for studying the basic mechanism of splicing.[31]

Of all the snRNP complexes, the U4/U6·U5 triple-snRNP undergoes the most drastic conformational change upon catalytic activation of the spliceosome and contains many of the critical proteins needed for the rearrangement that produces the catalytically active spliceosome. Because of this, and its conservation between lower and higher eukaryotes, the U4/U6·U5 triple-snRNP, and more specifically U5, has been the focus of many research groups.

1.3 A closer look at the U4/U6·U5 triple-snRNP and U5

The U4/U6·U5 triple-snRNP is crucial to the splicing process. It is not until the U4/U6·U5 triple-snRNP arrives at the pre-catalytic spliceosome and rearrangement occurs to form the catalytically active spliceosome that splicing can occur. The U4/U6·U5 triple-snRNP is a 25S complex of snRNA and ~30 associated proteins that is assembled prior to its incorporation into the spliceosome (Figure 1.6).[33] All of the U4/U6·U5 triple-snRNP associated proteins are conserved from *S. cerevisiae* to humans except one yeast specific protein known as Spp381, a regulatory protein that was implicated to interact with Prp38 to promote U4/U6·U5 triple-snRNP activity.[32,34] In the U4/U6·U5 triple-snRNP the snRNAs of U4 and U6 are base paired together, while U5 is held to the U4/U6 complex by Snu66 and Prp6 interactions with proteins in U5 and U4/U6.[33,35] Once assembled, it has been suggested that the U4/U6·U5 triple-snRNP is recruited to the pre-mRNA strand via the internal loop 1 of the U5 snRNA. Once associated with the spliceosome, rearrangement can occur to form the catalytic spliceosome.[36]

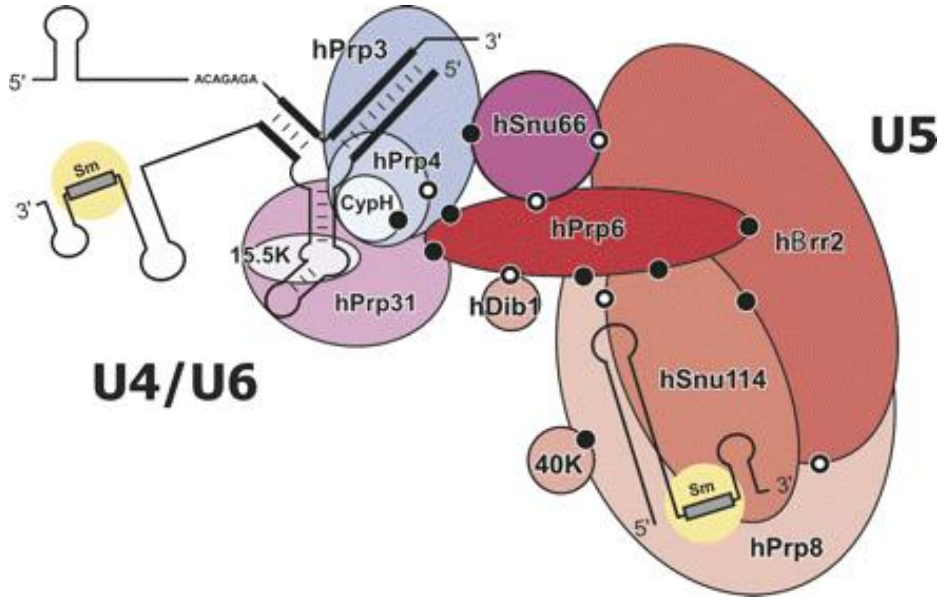

Figure 1.6. Cartoon representation of the human U4/U6·U5 triple-snRNP. A yeast two-hybrid study was done to elucidate the protein-protein interactions in the human U4/U6·U5 triple-snRNP. The tri-snRNP consists of U4 and U6, which are base paired together, U5, and many associated proteins. Figure reprinted from Liu *et al.*[33]

Within U4/U6, a 15.5 kDa protein (Snu13 in yeast), binds to the purine rich stem loop of U4 snRNA and acts as a scaffold for assembly of the remaining U4/U6 complex.[37] Another U4/U6 associated protein known as 61K (also known as Prp31) has been shown to contribute to the formation of the U4/U6·U5 triple-snRNP and mutation of this protein has been linked to autosomal dominant retinitis pigmentosa.[38,39] In humans, proteins 20K (CypH), 60K (Prp4), and 90K (Prp3) form a complex that also associates with U4/U6.[40] This complex has been implicated in the formation of U4/U6 snRNA secondary structure that is critical for splicing activity.[40] While 60K and 90K are conserved in yeast there is no known homolog for 20K.[32] Both the the 20K/60K/90K complex and 61K will only associate with U4/U6 once 15.5 kDa has associated with it, further supporting the idea that 15.5 kDa acts as a scaffold for the other U4/U6 proteins.[40]

While U4/U6 associated proteins as of now have only been shown to assist in accumulation of the U4/U6·U5 triple-snRNP and stabilization of U4/U6 snRNA, U5 snRNA and associated proteins have been shown to be more diverse in their functions and critical to the splicing process. The snRNA associated with U5 is critical for exon alignment and ligation, acting as a tether to hold both of the exons in the proper orientation for splicing to take place (Figure 1.7).[41,42] This snRNP also contains many proteins that are critical for splicing such as Brr2, Snu114, Prp8, Prp6, Prp28, and Dib1.[32]

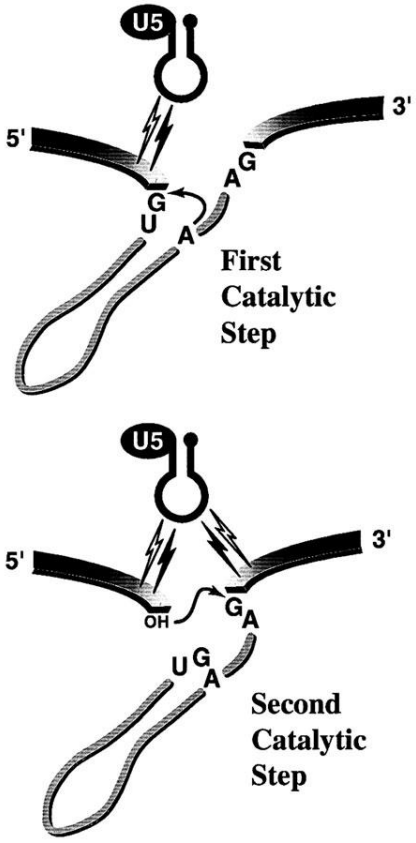

Figure 1.7. U5 snRNA as an exon tether. U5 snRNA acts as a tether, holding the 5' exon in proximity of the 3' exon to facilitate the second transesterification reaction that ligates the two exons together. Figure reprinted from A. J. Newman.[42]

Brr2, one of the more extensively studied of the U5 proteins, is a DExD/H box protein with ATPase activity that acts as a helicase and unwinds U4/U6.[26] This unwinding

event is crucial; without it, U6 and U2 cannot base pair, the catalytically active spliceosome cannot assemble, and splicing cannot occur. A U5 DEAD box protein known as Prp28 is the RNA helicase responsible for the unwinding of U1 from the 5' splice site, allowing for the critical rearrangement that must occur to form the catalytically active spliceosome.[27] The U5 protein Prp8, the largest spliceosomal protein, has been proposed to play the largest role in the spliceosome, being involved in multiple steps and regulation. Prp8 regulates the activities of Prp28 and Brr2, only allowing for the conformational changes that lead to the catalytically active spliceosome to occur once everything is assembled in the correct place.[43] It also has been shown that Prp8 mutants can either promote the 1st transesterification reaction by inhibiting the 2nd transesterification reaction, or it can promote the 2nd transesterification reaction by inhibiting the 1st transesterification reaction depending on where the mutation is in Prp8.[44,45] This led to a model that proposes that Prp8 is also involved in the transition between the 1st and 2nd transesterification reactions.[46] Another U5 protein, Snu114, is special because it is the only GTPase in the spliceosome and is critical for spliceosome activation. It has been proposed that Snu114 acts as a regulator of Prp8, controlling when the spliceosome becomes catalytically active.[47] All of these proteins coordinate together efficiently to ensure that splicing is only activated in the correct time and place, and deletion of any of these genes results in lethality.[48–51]

The U5 associated protein Prp6 is also necessary for cell viability and splicing for its use as a scaffold.[52] The U5 snRNP and the U4/U6 snRNP complex are not bound together directly; they must be held together by protein-protein interactions. Prp6 interacts with at least four U5 proteins and at least two U4/U6 proteins, acting as a molecular bridge holding the tri-snRNP together (Figure 1.6).[33,3933] Without Prp6, U4/U6 levels stay constant, but accumulation of the tri-snRNP diminishes significantly.[35]

Of all the U5 associated proteins, the least studied and most ambiguous is a protein known by many names: U5 specific 15 kDa protein/Dim1 in humans, Dim1 in *S. pombe*, and Dib1 in *S. cerevisiae*. This protein is 143 amino acids in *S. cerevisiae* and 142 in humans, with yeast having an extra alanine at the N terminus. The sequences are highly conserved (Figure 1.8). While it is known that the yeast homolog Dib1 is crucial for cell viability and the crystal structure of human Dim1 has been elucidated, the true biochemical function of this protein remains unknown.

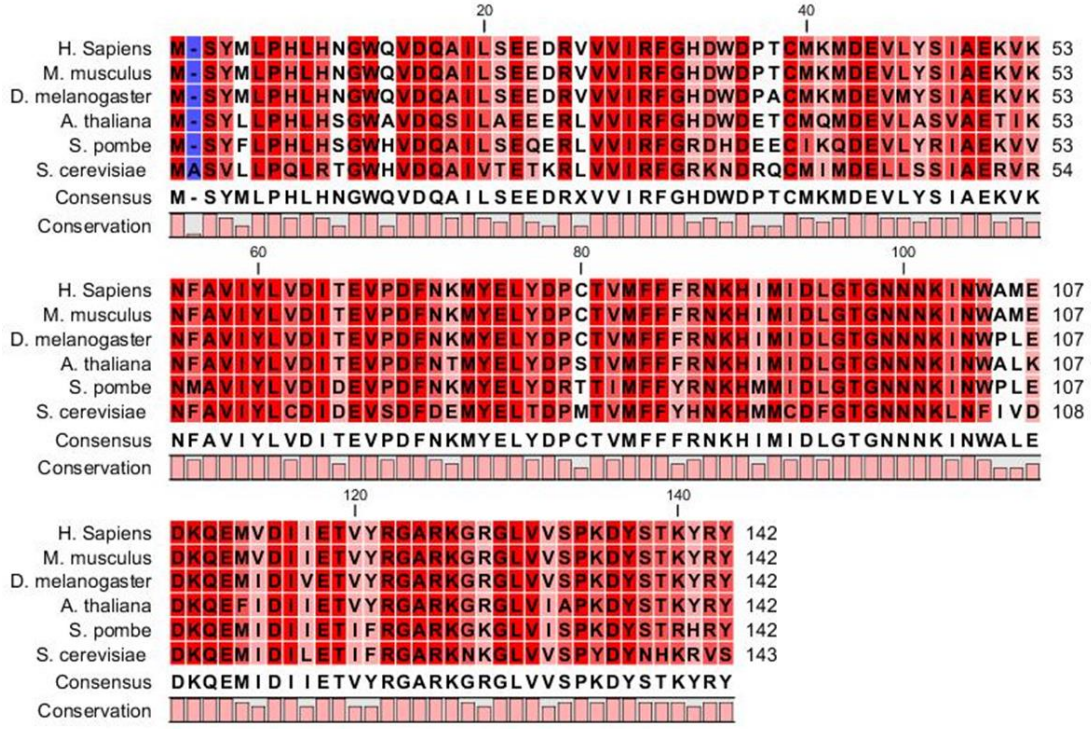

Figure 1.8. Sequence conservation of Dib1 across species. Dib1 is highly conserved from *H. sapiens* to *S. cerevisiae*, signifying its importance. Residues in darker red are highly conserved, lighter red residues are similar, and residues in white are not conserved. Figure created using CLC Sequence Viewer.

1.4 Structure of the Dib1 homolog Dim1

Human Dim1 has been crystallized and characterized in previous studies.[53–55] Human Dim1 has a thioredoxin-like fold, containing a core of four β-pleated sheets

surrounded by three α-helices and a fourteen amino acid C terminal tail (Figure 1.9).[53] Although other proteins with thioredoxin like folds have been shown to participate in oxidation-reduction reactions, human Dim1 has not been shown to have a similar activity.[55] In fact, the cysteine that would normally participate in oxidation reduction reactions is not conserved across species. This implies that Dim1 does not participate in oxidation-reduction reactions in the same way as other proteins with thioredoxin folds, but rather just maintains the common tertiary structure. The fact that this protein does not have the activity of other similarly structured proteins adds another layer of ambiguity as to its biochemical function.[56]

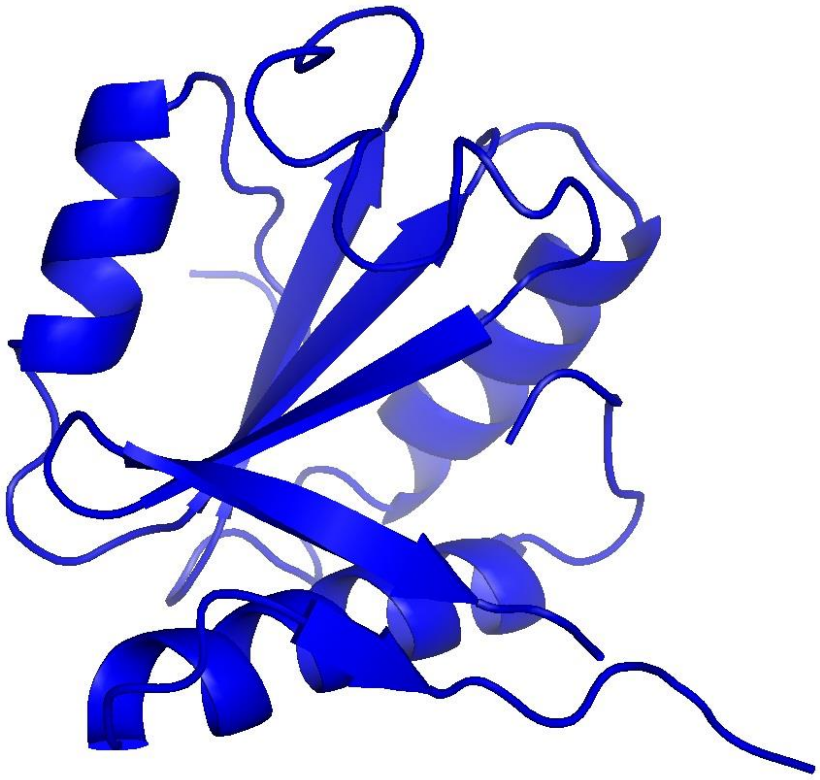

Figure 1.9. Crystal structure of human Dim1. Dim1 has a thioredoxin core consisting of four core beta sheets surrounded by three alpha helices and a fourteen amino acid C-terminal tail. Figure created using Pymol (PDB entry 1QGV[53]).

Yeast two-hybrid studies have identified proteins that interact with human Dim1, although where they interact on Dim1, with the exception of Pqbp1, remains unknown. A study by Zhang *et al.* identified a few interesting sites on Dim1 that could be possible protein-protein interaction points.[54] There are two hydrophobic grooves that are solvent exposed on Dim1: The first encompassing residues M72, M82, M91, I92, L94, I102, and W104 and the second encompassing residues W12, V14, I18, L19, F30, F69, and F84 (Figure 1.10).[53] Another interesting area on Dim1 is a basic patch consisting of residues R121, R124, K125, R127, R86, and K88 (Figure 1.10).[53] Mutation studies done with Dim1 where residues R86 and K88 were simultaneously mutated to alanine showed that these residues were critical for cell viability.[55] Further investigation into these three conserved sites on Dim1 could give insight into possible interaction points with other molecules.

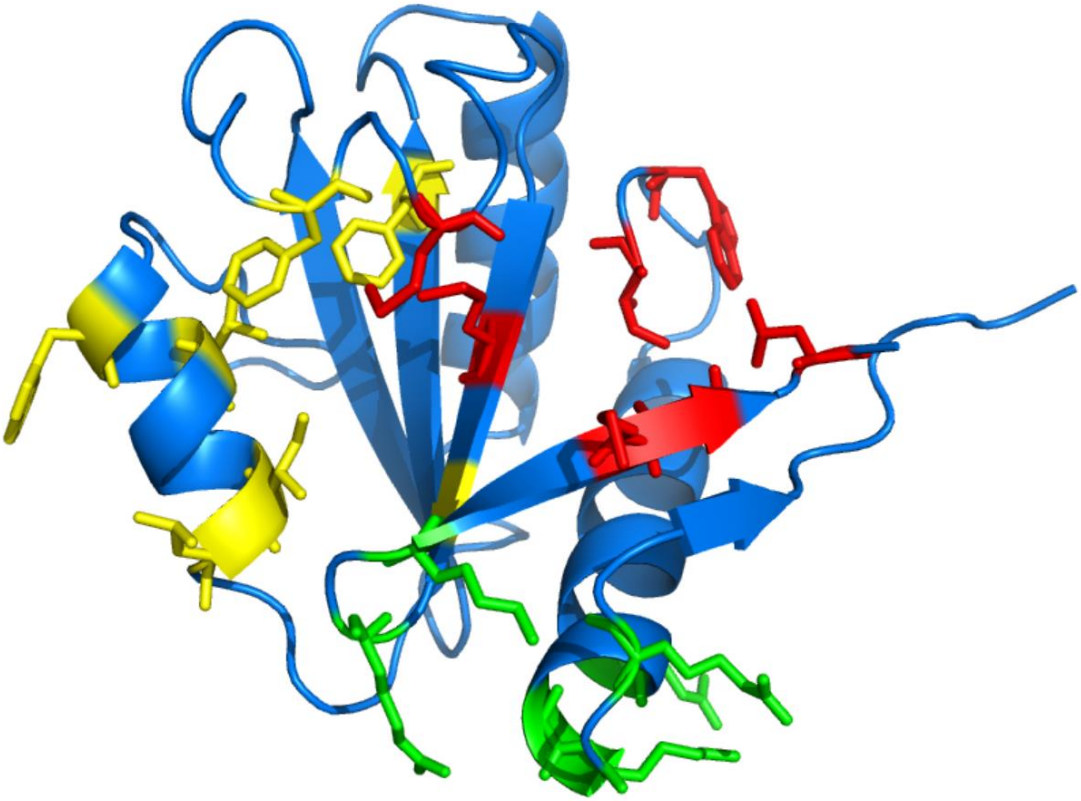

Figure 1.10. Crystal structure of Dim1 highlighting possible sites of interaction. Dim1 has two hydrophobic faces, one encompassing residues M72, M82, M91, I92, L94, I102, and W104 (red) and one encompassing residues W12, V14, I18, L19, F30, F69, and F84 (yellow). Dim1 also has a basic patch consisting of residues R121, R124, K125, R127, R86, and K88. These three sites have been proposed as possible sites of interaction with other proteins.[54] Figure created using Pymol (PDB entry 1QGV[53]).

In previous studies in *S. pombe,* it was shown that the fourteen amino acid C terminal tail is important for Dim1 function.[54,55] The C terminal tail contains the sequence LVVSPKDYSTKYRY and has both alpha helix and beta sheet properties, which could indicate that it acquires different secondary structure depending on its environment.[55] Removal of this tail causes a dominant temperature sensitive phenotype and significantly alters the secondary structure as seen by analysis using circular dichroism.[54,55] Specifically, it is thought that this change in secondary structure affects the orientation of the previously described basic patch on Dim1 that may be an interaction point for Dim1 with other molecules.[55]

1.5 Proposed functions of Dim1/Dib

Dim1 has been studied extensively since the late 1990's and several hypotheses have arisen as to its exact biochemical function including entry into mitosis, involvement in splicing, and peptidase activity.

1.5.1 Dim1 as a regulator of entry into mitosis and pre-mRNA export of Lid1

Dim1 was first proposed to be related to entry into mitosis after studies done in *S. pombe*.[57] When grown at 36.5 °C, a termperature higher than *S. pombe* normally grows at, it was found that Dim1 mutants displayed unequal division of DNA to its daughter cells, DNA oriented unnaturally along the mitotic spindle, or a undivided nucleus stuck in interphase.[57] In addition, it was found that cell viability significantly decreased halfway into the second cycle of cell division, indicating possible S phase defects due to the Dim1 mutation.[57] This was supported further by nitrogen starvation/release experiments and spore germination experiments.[57]

In a follow up study done by the same group, a synthetical lethal screen was performed with a temperature sensitive Dim1 mutant to identify any other proteins that interact with Dim1.[58] They identified one specific mutant, which they labeled *lid1-6 (*lethal in dim1-6*),* which showed a phenotype similar to previously identified *cut4* and *cut9* mutants, both genes that encode proteins that are part of the anaphase-promoting complex.[58] By putting the cDNA of the *lid1* gene under control of a *nmt41* promoter, thereby removing the need for splicing, they also showed that the phenotype which resulted in a decrease in Lid1 protein was not purely due to disregulation of splicing of the *lid1* gene [58] Interestingly, they showed that even this did not result in restoration of the wild type levels of Lid1.[58] In fact, *lid1* mRNA was shown to accumulate in the nucleus in the *dim1* mutant while total mRNA levels did not show a similar accumulation.[58] This led the group to hypothesize that Dim1

not only is associated with the spliceosome, but rather may serve multiple functions, including mRNA export of certain genes.[58]

These hypotheses are not widely supported by other research groups.[54,56,59,60] Other groups suggest that these phenotypes are simply due to disruption of pre-mRNA splicing of genes encoding proteins responsible for cell-cycle progression and mRNA export rather than Dim1's direct involvement.[6059] In fact, many of the genes mentioned in these papers involved in *S. pombe* cell cycle regulation (such as *cut9, cdc2, nuc2*) contain multiple introns.[8] The retention of these introns or mistakes in splicing are likely the cause of these phenotypes.

1.5.2 Dim1 as a peptidase

When Zhang *et al.* began looking into the structure and function of human Dim1, they found that Dim1 is cleaved after amino acid 128, removing the 14 amino acid tail but leaving the thioredoxin core intact.[54] This shortened Dim1 showed increased stability from the wild type but also displayed a dominant negative phenotype when expressed at restrictive temperatures in *S. pombe.*[54] In a follow up study by Jin *et al.*, it was shown that this cleavage activity is caused by Dim1 itself, indicating it has a kind of peptidase activity.[59] This was further shown by the synthesis of a Vhb-d20-GST fusion protein which contained the C-terminal twenty amino acids of Dim1 as the linker between Vhb-d20 and the GST tag.[59] The fusion protein was not cleaved until incubation with Dim1, showing that Dim1 recognizes this cleavage site whether internal or external.[59]

While these results are interesting, it raises more questions than it answers. Can we be sure that there were no contaminating peptidases in the Dim1 solution that led to the cleavage? They did not perform mass spectroscopy to verify the purity of their samples and, if there is a peptidase that has an affinity for Dim1, it is possible that they were copurified.[59]

It is also important to note that the Dim1ΔC14 proteinis still fully functional at permissive temperatures in *S. pombe* and does not show its dominant negative phenotype until incubated at restrictive temperatures.[54] This indicates that while these C terminal residues are likely important for stabilizing protein-protein interactions *in vivo*, they are not critical for activity.[61]

1.5.3 Dim1 as an essential part of the splicing machinery

Studies of protein-protein interactions and splicing *in vivo* have indicated that Dim1 likely plays a role in regulation of the activation of the spliceosome. The first such study was done in *S. pombe* with the *lid1* gene (now known as *cut20*). The *lid1* gene codes for a component of the APC/C ubiquitin ligase complex in *S. pombe* and contains four introns which must be removed in order to produce functional Lid1 protein.[8,62] It was shown that even at permissive temperatures, a *dim1* mutant resulted in a significant decrease in pre-mRNA splicing of the *lid1* pre-mRNA and thus a decrease in Lid1 protein expression.[62] In another study done using pre-U3 RNA in *S. cerevisiae,* Dib1 depletion was shown to coincide with the accumulation of unspliced transcripts.[53]

In addition to these studies, other protein-protein interaction studies have also given support to this hypothesis. In a yeast two-hybrid study done by Zhang *et al.* using a HeLa cDNA library to scan for protein-protein interactions with human Dim1, it was shown that Dim1 associates with proteins Prp6, Np/PQ, HEF1, Ubc9, hnRNP F, and hnRNP H^+ (Figure 1.10).[56] Np/PQ (also known as the polyglutamine tract-binding protein-1 (PQBP-1)) is a 265 amino acid protein containing 3 domains: a WW domain, a polar amino acid domain, and a C-terminal domain. PQBP1 binds to RNA polymerase II in a phosphorylated dependent manner via its WW domain.[63] PQBP1 binds Dim1 via an YxxPxxVL motif that recognizes a hydrophobic cleft on Dim1 (Figure 1.11).[64] It should be noted that this hydrophobic cleft does not correspond with the two hydrophobic faces that were suggested

previously as possible points of protein-protein interaction, but rather introduces a new hydrophobic cleft that was not previously examined. The interaction between PQBP1 and RNA polymerase II in addition to the interaction between PQBP1 and Dim1 indicates that PQBP1 and Dim1 may form a molecular bridge between transcription and splicing.[63,65] Frameshift mutants of PQBP-1 that have been shown to be associated with X-linked mental retardation (XLMR) lose this portion of the C-terminal domain to which Dim1 binds, signifying that disruption of this interaction may be linked to disease.[66]

hnRNP H^+ and hnRNP F are heteronuclear ribonucleoproteins linked to regulation of alternative splicing and tissue-specific splicing.[67–69] The fact that Dim1 also interacts with both of these splicing regulators in combination with the previously mentioned proteins and its central location within the tri-snRNP makes Dim1 a prime candidate to act as a hub to recruit many different splicing factors to the spliceosome (Figure 1.12). While the crystal structure of the Dim1/PQBP complex has been published, no other binding site on Dim1 has been found thus far. [64]

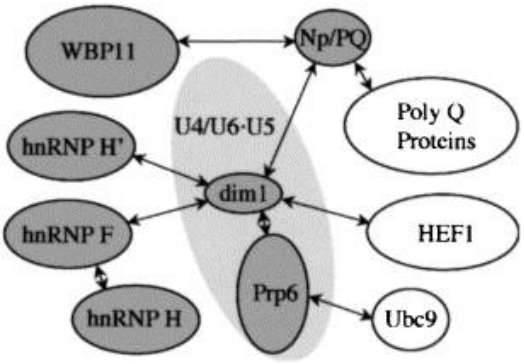

Figure 1.11. Crystal structure of the PQBP1/Dim1 complex. PQBP1 binds Dim1 via a YxxPxxVL domain on its C terminal tail (shown in blue) by recognizing a hydrophobic groove in Dim1 located on Dim1 (shown in green).[64] This figure was created using Pymol (PDB entry 4CD0[64]).

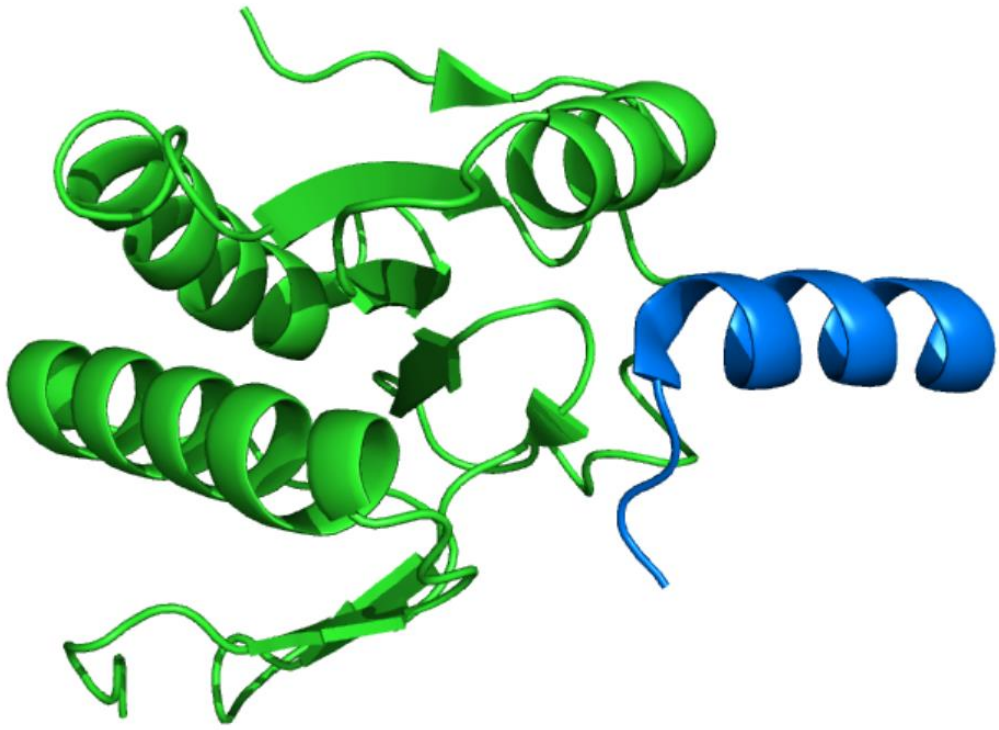

Figure 1.12. Interactions between Dim1 and other proteins from a yeast two-hybrid study. It was shown that, in addition to the tri-snRNP associated protein Prp6, Dim1 also interacts with Np/PQ, hnRNP H^+, hnRNP F, and HEF1 implying it may play a role in the transition between transcription and splicing as well as regulation of alternative or tissue specific splicing.[63,65,67–69] Figure reprinted from Zhang *et al.*[56]

1.5.4 Potential role of Dib1 in splicing

While many of these studies analyze U5-specific 15kDa/Dim1/Dib1 in very different environments, we hypothesize that this protein works in a very global way to allosterically keep the spliceosome inactive until all the necessary components are together and coordinated in a kind of lock and key method. This hypothesis is supported based on U5 specific 15kDa/Dim1/Dib1's multiple protein-protein interactions, its flexible C-terminal tail that can adopt different conformations, its central location in the U4/U6·U5 triple-snRNP, and the fact that it is present in the pre-catalytic spliceosome but dissociates once the spliceosome becomes catalytically active.[32,33,53–56,58,60,62,70] U5 specific 15kDa/Dim1/Dib1

in conjunction with Prp6 likely provides the scaffold that recruits many of the factors required for splicing, recruits the spliceosome to newly synthesize transcripts, and then activates the spliceosome with a conformational change that results in the release of U5 specific 15kDa/Dim1/Dib1 from the spliceosomal complex.[32,33,55,56,66]

1.6 Goals of this study

The ultimate goal of this research is to elucidate the biochemical function of Dib1 in the spliceosome, but the short term goals of this study were to identify amino acid residues critical for Dib1 activity, purify Dib1 protein and interesting Dib1 mutants, and use CD spectroscopy to analyze Dib1 and mutant Dib1 secondary structure. In order to do this, a *S. cerevisiae dib1*⊿ strain was created and used to analyze the effects of point mutations in *dib1* on cell viability. Once interesting mutants were identified, the corresponding Dib1 protein was purified using nickel affinity and ion exchange chromatography and CD spectroscopy was performed to analyze the effect of the mutations in *dib1* on Dib1 secondary structure.

CHAPTER 2

MATERIALS AND METHODS

2.1 Media preparation

A table of bacterial and yeast media recipes used in this study are listed in Table 2.1.

Table 2.1. List of media used in this study. All protocols were modified from Colowick, S. P. and Kaplan, N. O..[71]

Media	Recipe
YEPD broth	20 g Difco Bacto-peptone, 10 g Difco yeast extract, and 2% w/v glucose per liter *Prep:* Add Difco Bacto-peptone and Difco yeast and fill to 950 mL with water. Aliquot 95 mL into media bottles and autoclave. When ready to use, add 5 mL of autoclaved 40 % w/v glucose.
YEPD plates	20 g Bacto-peptone, 10 g Difco yeast extract, 15 g Bacto Agar, and 2% w/v glucose per liter *Prep:* Add Difco Bacto-peptone, Difco yeast extract, and Bacto agar and fill to 950 mL with water. Autoclave. Add 50 mL of autoclaved 40% w/v glucose and mix well. Pour into plates and allow to cool for at least 30 min until solid.
10 X amino acid dropout mix	0.3 g isoleucine, 1.5 g valine, 0.4 g adenine, 0.02 g arginine, 0.2 g histidine, 1.0 g leucine, 0.3 g lysine, 0.2 g methionine, 0.5 g phenylalanine, 2.0 g threonine, 0.4 g tryptophan, 0.3 g tyrosine, and 0.12 g uracil per liter *Prep:* Dissolve all amino acids except the dropout amino acid in 400 mL water and then bring the volume up to 1 L. Aliquot 100 mL into media bottles, autoclave, and store in a cool, dark place.

Table 2.1, continued.

Media	**Recipe**
SD dropout plates	6.7 g Difco Yeast Nitrogen Base w/o amino acids, 20 g Agar, 2 % w/v glucose, and 1X amino acid dropout mix per liter *Prep:* Combine agar, Difco Yeast Nitrogen Base w/o amino acids, and 850 mL water and autoclave. Add 50 mL of 40 % w/v glucose and 100 mL of 10X amino acid dropout mix and mix well. Pour media into plates and allow to cool for at least 30 min until solid.
Presporulation plates	10 g Bacto Yeast Extract, 30 g of Difco Bacto Nutrient Broth, 20 g Bacto Agar, and 5% w/v glucose per liter *Prep:* Combine all components except glucose and mix well. Autoclave. Add 125 mL of 40% w/v glucose and mix well. Pour into plates and allow to solidify.
Sporulation media	30 mM KOAc (Sigma Aldrich) and 0.4 mM raffinose (Sigma Aldrich) *Prep:* Combine all components and mix well. Autoclave.
5-fluoroorotic acid plates	6.7 g Difco Bacto Yeast Nitrogen Base w/o Amino Acids, 0.05 g dry uracil, 1X amino acid dropout mix without uracil, 1 g 5-FOA, 20 g Bacto Agar, and 2% w/v glucose per liter *Prep:* Dissolve agar in 500 mL of water and autoclave. While the agar is autoclaving, make and filter sterilize the 500 mL solution containing Difco Bacto Yeast Nitrogen Base w/o Amino Acids, uracil, amino acid dropout mix w/o uracil, dry uracil, 5-FOA, and glucose. After the afar has cooled slightly, combine agar and solution. Mix well and pour into plates to solidify

2.2 Preparation Top10 Competent *E. coli* cells

This protocol is modified from Chung and Miller (1988).[72] A freezer stock of Top10 competent cells (Life Technologies) was inoculated in 1 L of LB broth and allowed to grow overnight at 37 °C with vigorous aeration. The culture was then diluted 1/50 and was allowed to grow at 37 °C with vigorous aeration to an OD_{550} of 0.4. Cells were then spun at 3738 x g for 15 min in an Avanti® J30I centrifuge (rotor: J-LITE® JLA-9.1000). Following centrifugation, supernatant was removed and cells were put on ice. Cells were then resuspended in $1/20^{th}$ the volume of cold LB/PEG/DMSO solution (40 mL LB, 2.5 mM PEG4000, 25 mM $MgCl_2$, 25 mM $MgSO_4$, pH to 6.1, and 2.5 mL DMSO) and incubated on ice for 5 min. Cells were aliquotted in 300 μL amounts into 0.5 mL microfuge tubes on ice and stored in the -80 °C freezer until needed.

2.3 Preparation of pET15b-DIB1, pSE360-DIB1, and pSE362-DIB1 plasmids

2.3.1 Vectors used in this study

Vectors pSE360 and pSE362 were chosen for use in *DIB1* gene expression studies in *S. cerevisiae* and the vector pET15b was chosen for use in Dib1 protein expression and purification.[73,74] Features and the function of these plasmids can be seen in Table 2.2.

Table 2.2. Feature and functions of the plasmids used in this study

Plasmid	Features	Function
pSE360	Amp^R, *CEN4, ARS1, URA3*	Allows for expression of a specific gene in *S. cerevisiae,* selection on SD -URA media for *S. cerevisiae,* and selection on ampicillin containing media for *E. coli*
pSE362	Amp^R, *CEN4, ARS1, HIS3*	Allows for expression of a specific gene in *S. cerevisiae,* selection on SD -HIS media for *S. cerevisiae,* and selection on ampicillin containing media for *E. coli*
pET15b	Amp^R*, lacI*, *oriC, f1 ori*, *lacO*, T7 promoter, 6 histidine tag, thrombin cleavage site	Allows for overexpression of a specific protein in *E. coli* and utilization of a six histidine tag for protein purification. The *lacI* gene codes for an inhibitor that inhibits transcription of the T7 polymerase that is used to transcribe the gene of interest. Once a lactose derivative is added, the inhibitor no long inhibits T7 polymerase transcription, and transcription and translation of the gene of interest occurs very rapidly producing large amounts of protein.

2.3.2 Primer design for *DIB1* PCR

Primers for the full length *DIB1 S. cerevisiae* cassette containing ~600 bases upstream and downstream from *DIB1* coding sequence and the *DIB1* coding sequence were engineered with restriction digest sites that corresponded to the vector into which they would be cloned. Particular care was taken to ensure that the correct directionality would result and the oligonucleotides were ordered from Integrated DNA Technologies. The *DIB1* cassette forward primer contained a cut site for SacI and the reverse primer contained a cut site for XbaI to correspond to the XbaI and SacI restriction digest sites in vectors pSE362 and pSE360 (Table 2.3). The *DIB1* coding sequence forward primer contained a restriction site for NdeI and the reverse primer contained a restriction site for BglII

corresponding to the NdeI and BamHI restriction digest sites in pET15b (Table 2.3). It should be noted that *DIB1* has a BamHI site in the middle of the gene so BglII had to be used instead. Vectors pSE360 and pSE362 are *S. cerevisiae* expression vectors that allow the expression of specific genes in *S. cerevisiae* and the pET15b vector is an *E. coli* protein expression vector that allows for the overexpression of proteins with a six histidine tag and thrombin cleavage site for purification.[73,74]

Table 2.3. Primers used for PCR of *DIB1*

Primer number	Primer name	Sequence	Cut site
oCM118	*DIB1* cassette forward	GCGAGCTCGGCGCTATTCTTATTTCCACC	SacI
oCM134	*DIB1* cassette reverse	GATCTCTAGAGCCCATCCAACCTGAAGC	XbaI
oCM120	*DIB1* CDS forward	GGAATTCCATATGGCTAGTGTTTTGTTGCC	NdeI
oCM165	*DIB1* CDS reverse	GGGCAGATCTTTATGAAACACGCTTATGATTATAATCG	BglII

2.3.3 Genomic DNA purification from *S. cerevisiae* cells

Genomic DNA from *S. cerevisiae* was needed in order to perform PCR of *DIB1* genes. Purification of genomic DNA from *S. cerevisiae* strain yCM21 (Mat alpha, *his3*Δ*1, ura3*Δ*1, lys2*Δ*1, trp1*Δ*1, met15*Δ*1*, ATCC 200889[75]) was done as follows. yCM21 cells were streaked onto a YEPD plate and allowed to grow at 30 °C for two days. A single colony was then inoculated in 5 mL of YEPD broth and allowed to grow overnight at 30 °C with vigorous aeration. After overnight incubation, 1.5 mL of the culture was transferred to a microfuge tube and spun at full speed (16168 X g) in an Eppendorf Centrifuge 5145D for 15 s to pellet cells. The supernatant was then removed, and another 1.5 mL of culture was added to the microfuge tube. The spin and removal of the supernatant was repeated and the pellet was

resuspended in 300 μL of SET buffer (6% SDS, 10 mM EDTA, 30 mM Tris, pH 8). The cells were then incubated at 65 °C for 15 min and then transferred to ice for 5 min. After this incubation, 150 μL of cold 3 M NaOAc solution was added and mixed well. Cell suspension was then spun for 10 min at full speed (16168 X g) in an Eppendorf Centrifuge 5145D and the supernatant was transferred to a new 1.5 mL microfuge tube. Following this transfer, 50 μL of isopropanol was added and mixed thoroughly. This solution was then spun at full speed (16168 X g) in an Eppendorf Centrifuge 5145D for 2 min and the supernatant was removed and discarded. The DNA pellet was washed by slowly adding 500 μL of cold 70 % ethanol and incubating for 1 min. The supernatant was then removed and the pellet was dried by inverting on a sterile surface with the cap open for ~15 min. The DNA pellet was then resuspended in 50 μL of TE buffer (10 mM Tris, pH 8, 1 mM EDTA). Following resuspension, 1 μg of RNase A was added, mixed well, and incubated at 37 °C for 10 min. Genomic DNA was then stored at -20 °C until use.

2.3.4 PCR amplificationof *DIB1*

PCR was done on the *DIB1* cassette and *DIB1* coding sequence as follows. While on ice, 1X Bio-Rad iProof high fidelity buffer, 2.5 mM dNTPs (New England Biolabs), 50 ng of yeast genomic DNA, 0.5 μM of the forward primer, 0.5 μM of the reverse primer, and sterile ddH2O to a final volume of 20 μL were combined in a small, thin-walled PCR tube. In the final step, 0.4 U of iProof DNA polymerase (BioRad) was added and the tube was immediately put into the preheated PCR machine (BioRad T100 Thermal Cycler) with the following cycle parameters: 1 cycle for 3 minutes at 98 ºC, 30 cycles of 30 sec at 98 ºC, 30 sec at 65 ºC, and 1 min at 72 ºC, 1 cycle for 10 min at 72 ºC, and then the sample was held at 4 ºC overnight. In order to check for PCR product, 5 μL of sample was run on a 0.8 %

agarose gel, visualized with ethidium bromide, and imaged using the Alpha Innotech Red Imager.

2.3.5 Restriction digest of PCR products and vectors

Restriction digest of PCR products was done as follows. In a small microfuge tube, 1X NEB buffer (buffer 3 for *DIB1* coding sequence and buffer 4 for full *DIB1*), all of the PCR product, 5 μg of BSA (New England Biolabs), and 10 U of each restriction enzyme (Bgl1 and Nde1 for *DIB1* coding sequence and SacI and XbaI for full *DIB1*) were added to a final volume of 50 μL and incubated at 37 ºC overnight. Restriction digests of vectors pSE362 and pSE360 were done as stated above, using 1 μg of vector and restriction enzymes Xba1 and Sac1. Restriction digest of plasmid pET15b-*LDH* was done as stated above using 1 μg of plasmid and restriction enzymes BamhI and NdeI. Digests for vectors were confirmed by running 5 μL of digested product on a 0.8 % agarose gel and imaged using the Alpha Innotech Red Imager. This pET15b vector had an *LDH* gene insert and the digest removed this gene insert from the vector. This digest was run on a 0.8 % agarose gel and the bands corresponding to pET15b vector were excised and gel purified using a Qiagen QIAquick® Gel Extraction Kit (50) as stated by the Qiagen protocol.

2.3.6 Cloning of *DIB1* into vectors

The *DIB1* cassette was cloned into vectors pSE360 and pSE362 as follows. In a small microfuge tube, 1 μL of digested pSE360 or pSE362 was incubated with 16 μL of the digested full *DIB1,* 1X ligation buffer (New England Biolabs), and 400 U of T4 DNA ligase (New England Biolabs) were added to a final volume of 20 μL and the tube wasincubated at room temperature for 5 h. *DIB1* CDS was ligated into pET15b as stated above. After incubation, each ligated plasmid was transformed into Top10 *E. coli* competent cells. Top10 *E. coli* competent cells were thawed on ice for 10 min. The complete ligation mixture was

incubated with 100 µL of Top10 *E. coli* cells (prepared as in section 2.1) on ice for 5 min, placed in a 37 ºC water bath for 5 min, incubated for 5 min on ice, and then plated on LB plates containing ampicillin at a concentration of 0.1 mg/mL. These plates were incubated overnight at 37 ºC.

Each resulting colony was inoculated in 3 mL of LB with 0.3 µg ampicillin and allowed to grow overnight with vigorous aeration at 37°C. From these cultures, the plasmids were extracted using the Qiagen Qiaprep® Spin Miniprep Kit according to the protocol from Qiagen. Plasmid concentration was determined using Nanodrop 2000c spectrophotometer (Thermo Scientific). These plasmids were then digested as stated previously to check for the correct gene insert. Plasmids showing successful gene insertion were sent to Quintara Biolabs (San Francisco, CA) for sequencing.

2.3.7 Creation of pSE360-DIB1 plasmid

Because the ligation of pSE362 and *DIB1* cassette was successful and the sequence shown to be correct, the plasmid resulting from this ligation was once again digested with XbaI and SacI and run on a 0.8 % agarose gel. The *DIB1* cassette fragment was then gel extracted and purified using a Qiagen QIAquick® Gel Extraction Kit (50) according to the protocol from Qiagen. The vector pSE360 was digested again with Xba1 and Sac1 and ligated with the purified *DIB1* cassette. The ligated plasmid was transformed into *E. coli* and the plasmid extracted. The resulting plasmid was then sent to Quintara Biolabs for sequencing.

2.4 Creation of *dib1Δ* yeast strain

2.4.1 Preparation of single stranded salmon sperm

Single stranded salmon sperm DNA (ssssDNA) was prepared by dissolving high molecular weight salmon sperm DNA sodium salt (Sigma D126) in TE (pH 7.6) to a final

concentration of 10 mg/mL. The solution was mixed overnight at room temperature and the following day was sonicated for two 30 s intervals. The salmon sperm was then aliquoted into 100 μL portions and stored at -20 °C until use.

2.4.2 Transformation of pSE306-DIB1 into *DIB1/dib1Δ* yeast strain

Plasmid pSE360-*DIB1* was transformed into a *DIB1/dib1*⊿ yeast strain that was previously created by Steinmetz *et al.* in the Yeast Deletion Project.[76] YEPD broth (15 mL) was inoculated with yeast strain BY4743 (MAT a/α, *his3*⊿*1/his3*⊿*1, leu2*⊿*0/leu2*⊿*0, lys2*⊿*9/LYS2, MET15/met15*⊿*0, ura3*⊿*0/ura3*⊿*0, DIB1/dib1::KAN)* and incubated overnight at 30 ºC with vigorous aeration. The culture was the diluted to an A_{600} of 0.2 in 50 mL of YEPD broth and grown for 3 hours at 30 ºC with vigorous aeration. The culture was then centrifuged at 1806 X g for 5 min in a Beckman Coultier Allegra 25R tabletop centrifuge (rotor: TA-14-50). The supernatant was discarded and the pellet resuspended in 20 mL of sterile ddH_2O. The resuspended culture was then centrifuged again at 1806 X g for 5 min and the supernatant was removed. The resulting pellet was resuspended in 1 mL of 0.1 M LiOAc (Sigma Aldrich). Cells were spun at full speed (16168 X g) in an Eppendorf Centrifuge 5145D for 10 s to pellet cells and the supernatant was removed. Pellet was resuspended in 1 cell pellet volume of 0.1 M LiOAc and cells were incubated in water bath at 30 ºC for 15 min. A solution was made containing 240 μL of 50% PEG 3350 (Sigma Aldrich), 36 μL of 1.0 M LiOAc, 20 μL of sterile ddH_2O, 50 μg of ssssDNA that had been boiled for 10 min and cooled to room temp, 5 μL of pSE360-*DIB1* and 50 μL of the previously incubated cells. This mixture was then left to incubate overnight at room temperature. The next day, the mixture was incubated in a water bath at 30 ºC for 30 min. After incubation, 0.5 mL of sterile ddH_2O was added to the mixture, cells were pelleted, and the supernatant removed. Cells were then resuspended in 1.0 mL of sterile ddH_2O, the

cells pelleted again, and the supernatant removed. Cells were then resuspended in the residual ddH_2O, plated on SD dropout plates lacking uracil and incubated at 30 ºC for 48 h.

2.4.3 Isolation of *dib1Δ* haploid spores

A *dib1*⊿ haploid strain containing the plasmid pSE360-*DIB1* was isolated. A smear of yeast cells from the colonies resulting from the previous transformation were plated on presporulation plates and allowed to grow at 30 ºC for 8 h. Yeast cells were then inoculated in 3 mL of sporulation media and allowed to incubate at room temp for 7 days with slow aeration until tetrads were visible under a microscope. After tetrad visualization, 350 µL of the cells were centrifuged at 2320 x g for 30 s and the supernatant was discarded. The resulting pellet was resuspended in 20 µg of zymolase for 20 min. After incubation, cold sorbitol (Amresco) to a final concentration of 0.2 mM was added slowly. From this cell mixture, 8 µL was placed on edge of a YEPD plate and allowed to drip down to create a straight line of cells across the top on the plate (Figure 2.1). Using a dissection microscope (Nikon H550S), tetrads were picked up with a needle from the freshly streaked plate and each haploid cell of the tetrad was deposited in a column going down the plate. This was repeated until 10 columns of 4 haploid cells each were created. This plate was then incubated at 30 ºC for 2 days.

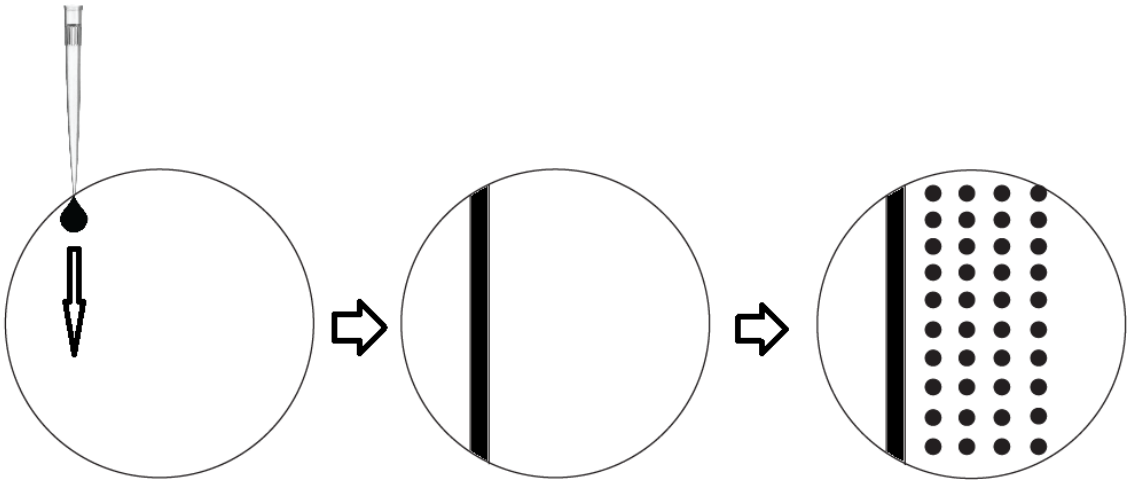

Figure 2.1. Schematic showing the process of tetrad picking of yeast cells. A small amount of sporulated cells are applied to the top of the plate and then the plate is tilted so that the cells drip down the top of the plate until a straight row is plated. Tetrad spores are then plated in 10 rows with 4 haploid spores in each row and allowed to grow.

The resulting colonies were replated on a YEPD plate labeled with a grid. This plate was allowed to grow for 2 days at 30 ºC. Each colony was then plated in the same grid fashion on a plate lacking uracil and a separate plate containing G418 to select for the plasmid which has a *URA3* gene that allows it to grow on plates lacking uracil, and the *DIB1* deletion which has a gene encoding G418 resistance. For colonies that grew on the SD -ura and G418 plates, PCR of the *DIB1* gene and *KAN1* were done to confirm these results.

2.5 Creation of *dib1* mutant plasmids using mutagenic PCR

Mutagenic PCR was performed to create *dib1* mutants. While on ice, 1X iProof HF buffer, 0.5 mM dNTPs, 0.5 mM $MgCl_2$, 1 μg pSE362-*DIB1* plasmid DNA, 0.2 mM forward and reverse primers were combined to a final volume of 50 μL in a 0.2 mL PCR tube. In the last step, 1 U iProof polymerase was added. Primers used can be seen in Table 2.4. PCR was performed using a Robocylcer Gradient 96 (Stratagene) with the following conditions: 1 cycle for 30 s at 98 ºC, 18 cycles with 10 s at 98 ºC, 30 s at 52 ºC and 10 min at 72 ºC, and 1 cycle for 30 min at 72 ºC. PCR products were run on a 0.8 % agarose gel for verification of

product. The resulting PCR product was digested with 20 U of Dpn1 and incubated overnight at 37 ºC. After digestion, the PCR product was transformed into Top10 *E. coli* cells (section 2.2) and candidate plasmid DNAs were isolated and purified using a Qiagen miniprep kit.

Table 2.4. Primers used for mutagenic PCR

Primer name	**Mutation**	**Primer sequence**
oCM146 (forward primer)	L76A, D78A	GATTTTGATGAGATGTACGAGGCTACGGCTCCCATGACAGTAATGTTTTTCTACCAC
oCM147 (reverse primer)	L76A, D78A	Reverse complement of oCM146
oCM148 (forward primer)	P134A	GGGCTGGTGGTCTCTGCTTACGATTATAATCATAAGCGTGTT
oCM149 (reverse primer)	P134A	Reverse complement of oCM148
oCM150 (forward primer)	C39A	GGTCGCAAAAATGATAGACAAGCTATGATCATGGACGAGTTGTTATCT
oCM151 (reverse primer)	C39A	Reverse complement of oCM150
oCM152 (forward primer)	D16A, Q17A	CGTACGGGATGGCATGTGGCTGCTGCTATTGTTACCGAGACT
oCM153 (reverse primer)	D16A, Q17A	Reverse complement of oCM152
oCM156 (forward primer)	D115A	GACGACAAACAAGAAATGATTGCTATATTAGAAACTATATTCAGAGGTGCC
oCM157 (reverse primer)	D115A	Reverse complement of oCM156

Table 2.4, continued.

Primer name	Mutation	Primer sequence
oCM158 (forward primer)	K102A	TGTGATTTTGGTACAGGGAACAACAACGCTTTGAACT TTATTGTAGACGAC
oCM159 (reverse primer)	K102A	Reverse complement of oCM156
oCM170 (forward primer)	D78A	AGAGGTGCCAGGAAGGGCAAAGGGCTGGTGGTC
oCM171 (reverse primer)	D78A	Reverse complement of oCM170
oCM172 (forward primer)	N127D	AGAGGTGCCAGGAAGGACAAAGGGCTGGTGGTC
oCM173 (reverse primer)	N127D	Reverse complement of oCM172
oCM195 (forward primer)	D63A	GAGTAAGGAACTTTGCAGTTATTTATTTATGTGCTAT AGATGAAGTTTCAG
oCM196 (reverse primer)	D63A	Reverse complement of oCM195
oCM197 (forward primer)	F85A	CCCATGACAGTAATGTTTGCCTACCACAATAAGCACAT GATG
oCM198 (reverse primer)	F85A	Reverse complement of oCM197
oCM199 (forward primer)	K89A	TCTACCACAATGCGCACATGATGTGTGATTTTGGTAC AGGGAAC
oCM200 (reverse primer)	K89A	Reverse complement of oCM199

2.6 Analysis of the effects of *dib1* mutants on cell viability

2.6.1 Transformation and plasmid swap of mutant *dib1* plasmids into *dib1Δ* yeast strain

Mutant *dib1* plasmids created by mutagenic PCR were transformed into the *dib1⊿* yeast strain as stated in section 2.4.2, except transformants were plated on SD -HIS plates instead of SD –URA plates. Once colonies were observed on the SD -HIS plates, the colonies were restreaked onto SD –HIS plates and allowed to grow for two days at 30 °C. The yeast cells were then transferred to 5-fluoroorotic acid (5-FOA) plates and allowed to grow for 32 h at 30 ºC. The 5-FOA plates select for yeast cells that do not contain the original pSE360-*DIB1* plasmid which has both *DIB1* and *URA3* genes.[77] It does this by causing any cell that contains the *URA3* gene to convert 5-FOA into the toxin 5-fluorouracil that results in cell death. This kills any cells that still retained the pSE360-*DIB1* wild type plasmid and leaves the cells that only contain the pSE362-*dib1* mutant plasmids. The yeast cells that grew on 5-FOA were then transferred to both SD -URA and SD -HIS plates and allowed to grow for 24 h at 30 ºC to verify loss of the pSE360-*DIB1* plasmid and maintenance of the pSE362-*DIB1* plasmid (Figure 2.2).

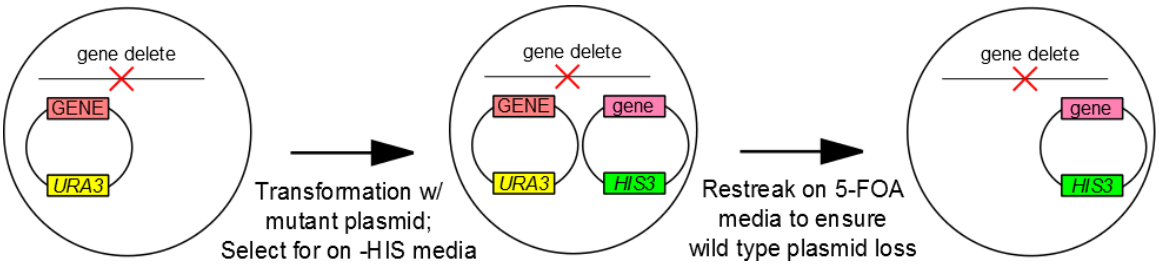

Figure 2.2. Plasmid swap in *S. cerevisiae*. Yeast cells containing the plasmid pSE360 are transformed with plasmid pSE362 and selected for on media lacking histidine. Cells are then plated on 5-FOA, further selecting for cells only containing pSE362. Cells are then plated on both media lacking uracil and media lacking histidine. If cells only grow on media lacking histidine, then the plasmid swap was a success.

2.6.2 Serial Dilution growth assay of *dib1* mutants

YEPD broth (5 mL) cultures were inoculated separately with mutant and wild type cells and allowed to grow overnight at 30 ºC with vigorous aeration. Each culture was diluted to an A_{600} of 0.2 and allowed to grow for 3 h. Each culture was again diluted to an A_{600} of 0.2 and placed in the first column of a 96 well microtiter dish. Six-fold serial dilutions were made down each row of the microtiter dish (Figure 2.3) and these dilutions were plated on 12 YEPD plates using a pronger. Three plates were incubated at 16 ºC, three at 25 ºC, three at 30 ºC, and three at 37 ºC and allowed to grow until the wild type strain showed significant growth. Each plate was analyzed for deviations in phenotype from the wild type strain.

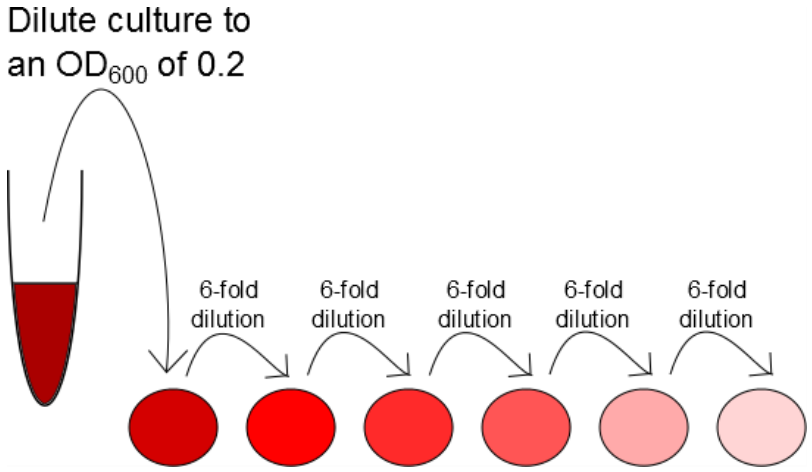

Figure 2.3. Serial dilution growth assay schematic. Cells are diluted to an OD_{600} of 0.2 with a final volume of 240 µL in the first well of a microtiter dish. From the first well, 40 µL of sample was removed and added to the second well. This was repeated for each succeeding well and then these diluted samples were plated on YEPD plates using a pronger.

2.7 Purification of Dib1

2.7.1 Small scale induction assay with Dib1

In order to verify expression of Dib1 protein from the pET15b-DIB1 plasmid, a small scale induction assay was performed (Figure 2.4). pET15b-*DIB1* plasmid was transformed

into One Shot BL21(DE3)pLysS cells (Invitrogen)as stated by the Invitrogen protocol, plated on LB plates containing ampicillin at a concentration of 0.1 mg/mL, and allowed to grow overnight at 37 °C. Three individual colonies from the transformation were separately incubated in 3 mL of LB with 3 μg of ampicillin and allowed to grow overnight at 37 °C with vigorous aeration. The following day, cells from each of the three tubes were then used to inoculate two separate new LB containing ampicillin cultures (one labeled + and one labeled -) at a starting OD_{600} of 0.1 and allowed to grow for 2 hours at 37 °C with vigorous aeration. Each + labeled culture was then induced using Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM. After induction, cultures were transferred to a 25 °C incubator and allowed to grow overnight with vigorous aeration. The following day, 30 μL of cells from each + and – labeled culture was incubated with 30 μL of 2X SDS and heated at 95 °C for 10 min.

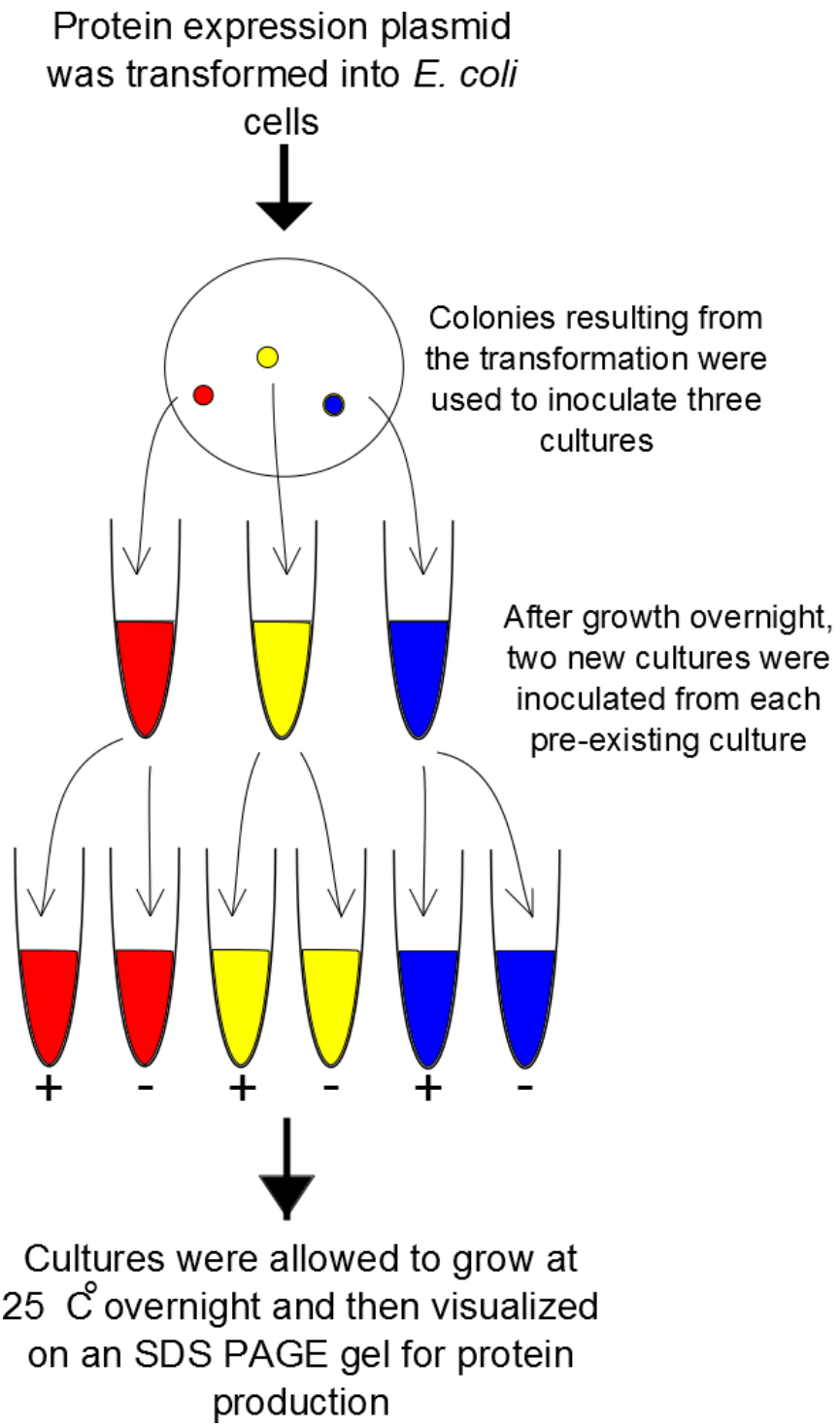

Figure 2.4. Small scale induction assay schematic.

2.7.2 Purification of Dib1 from *E. coli*

Dib1 protein was purified from an *E. coli* expression system. The plasmid pET15b-*DIB1* was transformed into One Shot BL21(DE3)pLysS cells as stated by the Invitrogen protocol. Individual colonies produced from the transformation were inoculated in 3 mL of LB with 0.3 μg of ampicillin and allowed to grow overnight at 37 ºC with vigorous aeration. The next day, that 3 mL culture was used to inoculate 1.0 L of LB with 1.0 mg of ampicillin and was allowed to grow to an A_{600} of ~ 0.3. Expression of *DIB1* was then induced by the addition of IPTG to a final concentration of 1.0 mM and the culture was incubated at 18 ºC overnight with vigorous aeration. After incubation, cells were spun at 3738 x g for 15 min in an Avanti® J30I centrifuge (rotor: J-LITE® JLA-9.1000) and the supernatant was discarded. The pellet was resuspended in 20 mL of lysing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 7) and transferred to an Oakridge tube. Cells were spun in the same centrifuge at 2506 x g (rotor: JA-17) for 15 min and supernatant discarded. The pellet was then stored at -80 ºC.

Probond Resin, a nickel-chelating resin (Invitrogen), was resuspended in 20% ethanol and 12 mL of 50% slurry was put into a column and allowed to settle. The column was then washed with 5 column volumes of ddH_2O followed by 5 column volumes of lysing buffer. The cell pellet was thawed on ice and resuspended in 20 mL of lysis buffer. The cells were lysed using a Brason Sonifier 450 for five 30 s cycles with an output control of 5 and a duty cycle of 80 with 30 s rests in between each cycle, keeping the cells on ice throughout. The cells were then spun in a pre-cooled JA-17 rotor at 25661 x g for 45 min. The supernatant was applied to the previously equilibrated column and allowed to flow through (Figure 2.5). The column was then washed with 15 mL of buffer identical to the lysis buffer with the exception of the imidazole concentration which was increased with each sequential wash

(started with 10 mM imidazole, followed by 25 mM, 50 mM, 100 mM, 250 mM, and 500 mM). The eluted samples were then run on a 20 % polyacrylamide gel to determine at what imidazole concentration the protein eluted. Elution samples corresponding to the 250 mM imidazole wash and the 500 mM imidazole wash were combined and dialyzed overnight against a buffer containing 50 mM 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), 100 mM NaCl, 1 mM dithiothreitol (DTT), and 5 % glycerol.

Dialyzed samples were digested with 10 U of human thrombin (Sigma) for 2 hours and dialyzed against 1.5 L x2 of buffer containing 20 mM Tris, 10 mM NaCl, 1 mM DTT, and 5 % glycerol, pH 8, switching out for fresh buffer after 1.5 h. During dialysis, 8 mL of Q Sepharose, an anion exchange resin, (Pharmacia Biotech) was degassed for 20 min and then loaded onto a column where it was allowed to settle for 90 min. The column was washed with 5 column volumes of ddH_2O followed by 5 column volumes of buffer containing 10 mM NaCl, 20 mM Tris, pH 8.0. The dialyzed protein was then applied to the column and allowed to flow through. The column was then washed with 15 mL of buffer containing 20 mM Tris, pH 8 but with each wash having an increasing concentration of NaCl (the first wash being with 10 mM NaCl, followed by 50 mM, 100 mM, 150 mM, and 200 mM). Elution fractions were run on a 20 % polyacrylamide gel to determine which fractions contained Dib1.

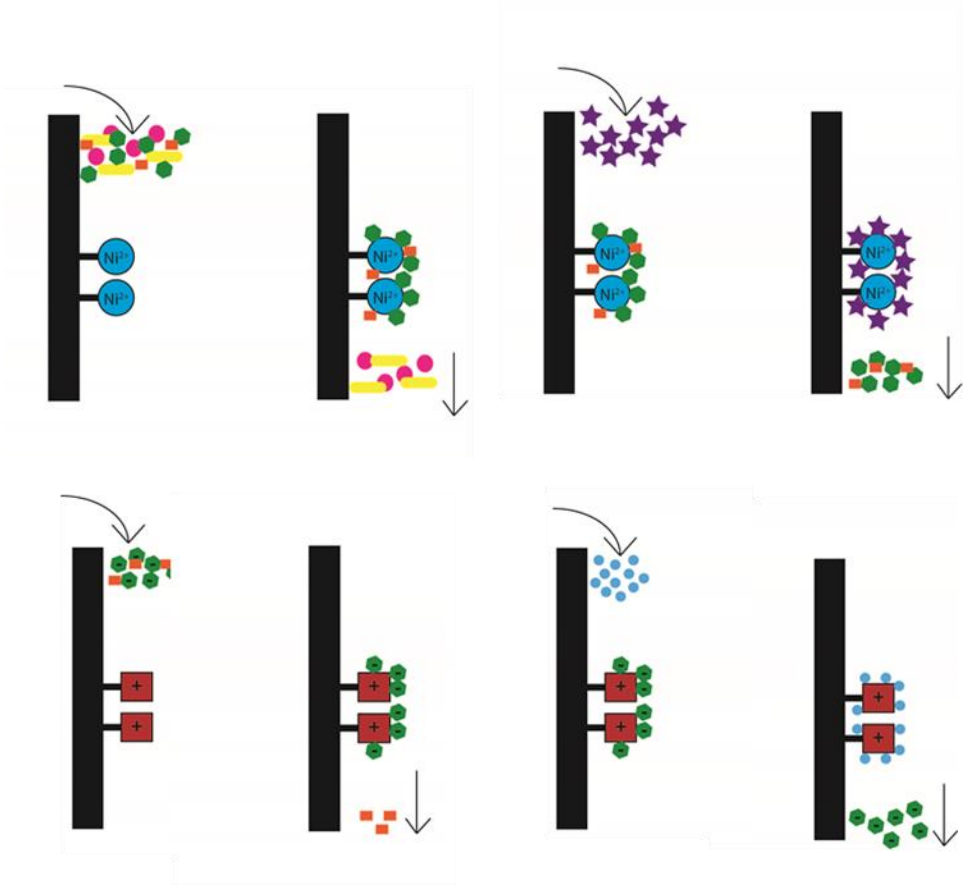

Figure 2.5. General schematic of Dib1 purification. The pink, yellow, and orange shapes represent contaminants, the purple stars represent imidazole, the blue circles represent NaCl, and the green hexagons represent Dib1. The protein containing supernatant was added to a nickel column and contaminants were washed away. Imidazole was added in excess to elute Dib1. Dib1 was then run over a Sepharose Q column, contaminants were washed away, and pure Dib1 was eluted using excess NaCl.

If contaminating bands were still present, a 25 mL of Superdex 75 (GE Healthcare Life Sciences) was added to a column and allowed to settle overnight. The column was equilibrated with 5 column volumes of water followed by 5 column volumes of 150 mM NaCl, 20 mM Tris, pH 8. A small amount of sample (~250 μL) was added to the column and the eluting sample was collecting in 0.5 mL fractions. These elution fractions were then run on a 20 % polyacrylamide gel to determine which fractions contained protein and protein concentration was determined using a Nanodrop 2000c spectrophotometer and Beer's law ($A = \varepsilon cl$ where A is absorption at 280 nm, ε is the extinction coefficient, c is the

concentration of protein, and l is the pathlength). The extinction coefficient of Dib1 is 13,075 $M^{-1}cm^{-1}$.

Purified protein was aliquoted into 250 µL fractions at a concentration of 0.3 mg/mL, flash frozen with liquid nitrogen, and stored at -80 ºC until ready to use. Mutant Dib1 protein was purified using this same protocol, except the beginning transformation into BL21(DE3)plysS cells was performed with the appropriate pET15b-*dib1* plasmid for that mutant.

2.8 Circular Dichroism Spectroscopy

CD spectroscopy of the wild type and mutant Dib1 protein was performed. An aliquot of protein was removed and thawed on ice. The protein at a concentration of 0.2 mg/mL was then dialyzed against a buffer containing 10 mM sodium phosphate, 100 mM NaCl, at pH 7 for a total of 3 h, switching out the buffer at the 1.5 h mark. A quartz cuvette was soaked in 10 M NaOH for 15 min, thoroughly rinsed with water and ethanol, and dried completely using a vacuum. After dialysis, 200 µL of protein sample was injecting into the clean cuvette and the cuvette was inserted into the Jasco J-715 CD spectropolarimeter. The CD spectrophotometer was purged with nitrogen gas according to the Jasco J-715 Spectropolarimeter protocol. Parameters were set in Spectra Manager as seen in Table 2.5 and a full length scan was done from 195 to 260 nm. After the full scan was performed, a thermal melt and cooling of the wild type and mutant Dib1 proteins was done by increasing the temperature by increments of 5 ºC and scanning at 222 nm at each temperature and then decreasing by increments of 5 ºC and repeating each measurement. All mdeg values were then converted to molar ellipticity using the equation $\frac{\left[(a)(0.1\ cm)\left(\frac{b}{c}\right)\right]}{[(0.1)(d)]}$ where a is mdeg, b is the molecular weight of the protein, c is the number of residues, and d is the concentration of

protein in mg/mL. The molecular weight of Dib1 is 16,776.3 g/mol and the number of residues Dib1 has is 143. For the full scans, these values were then plotted in Microsoft Excel with wavelength on the x-axis and molar ellipticity on the y-axis. For the thermal melt, the temperature was plotted on the x-axis and the molar ellipticity was plotted on the y-axis.

Table 2.5. Parameters used for CD spectroscopy

Data pitch	0.5 nm
Scanning mode	Continuous
Scanning speed	20 nm/min
Response	2 s
Bandwidth	10 nm
Accumulation	8 scans
Scan from	195 - 260 nm

2.9 Thermal stability of Dib1 in various buffers

Thermal stability studies of Dib1 were performed. Purified wild type Dib1 was thawed on ice and 250 μL was dialyzed for 1.5 h x2 against 3 L of the buffers listed in Table 2.6. After being dialyzed, the Dib1 solutions were heated at 95 °C and observed over a 10 min period. The time it took before precipitation of Dib1 was observed was documented.

Table 2.6. Buffers used in thermal stability study

Buffer 1	50 mM sodium phosphate, 50 mM NaCl
Buffer 2	50 mM sodium phosphate, 500 mM NaCl
Buffer 3	50 mM sodium phosphate, 50 mM KCl
Buffer 4	50 mM sodium phosphate, 500 mM KCl
Buffer 5	50 mM sodium phosphate, 100 mM NaCl, pH 7.1
Buffer 6	50 mM sodium phosphate, 100 mM NaCl, pH 8.7
Buffer 7	50 mM sodium phosphate, 100 mM NaCl, pH 5.8
Buffer 8	50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, pH 7.1
Buffer 9	50 mM sodium phosphate, 100 mM NaCl, 5 mM DTT, pH 7.1
Buffer 10	50 mM sodium phosphate, 100 mM NaCl, 1 mM β-mercaptoethanol, pH 7.1

CHAPTER 3

RESULTS AND ANALYSIS

Dib1 is a small, highly conserved 15 kDa protein located in the U4/U6·U5 tri-snRNP of the *S. cerevisiae* spliceosome that is necessary for cell viability. The human homolog, Dim1 has a thioredoxin-like fold and has been shown to interact with other proteins such as Prp6, hnRNP F, hnRNP H^+, and PQBP1. While its structure and interactions in the cell have been mapped, its biochemical function still remains elusive. The goal of this study was to create and analyze *dib1* mutants and their corresponding protein structures to identify amino acid residues in Dib1 critical for function, in the hopes that this would lend further understanding to its biochemical role and function in the cell.

3.1 PCR of *DIB1* and the *DIB1* cassette

In order to construct a protein expression plasmid for *DIB1* and a *DIB1* expression plasmid in *S. cerevisiae*, the *DIB1* gene was isolated using PCR. PCR was done as stated in section 2.1.1 in order to produce a *DIB1* coding strand (*DIB1* CDS) that only included the protein coding region to be used in the protein expression plasmid, and to produce a *DIB1* cassette which consisted of 1,500 bp and included regulatory elements upstream and downstream of *DIB1* as well as the *DIB1* gene (Figure 3.1). The results of these PCR experiments can be seen in Figure 3.2.

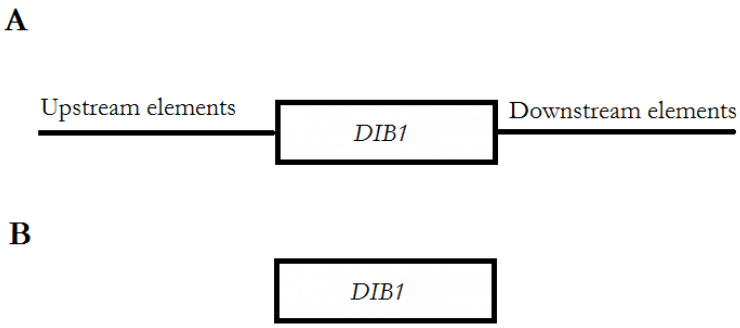

Figure 3.1. Schematic showing the *DIB1* fragments created using PCR. The *DIB1* cassette (A) contains both the upstream and downstream regulatory elements as well as the *DIB1* gene and the *DIB1* coding strand (B) consists of only the *DIB1* gene itself.

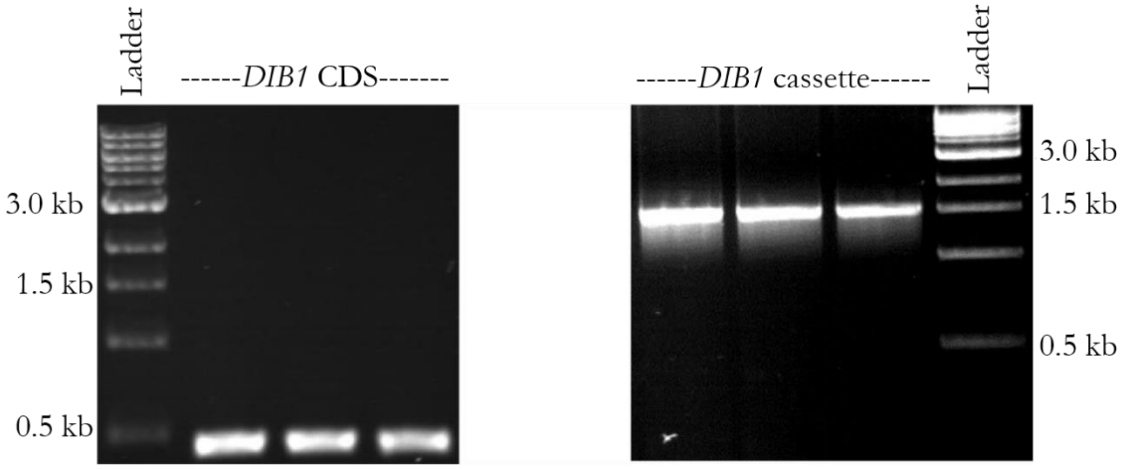

Figure 3.2. PCR of *DIB1* CDS and the *DIB1* cassette. PCR of coding *DIB1* resulted in a band at ~0.5 kb corresponding to the 432 bases in *DIB1* CDS. PCR of the *DIB1* cassette resulted in a band at ~1.5 kb corresponding to the 1.5 kb in the *DIB1* cassette. The ladder used was a 1 kb DNA ladder (New England Biolabs).

3.2 Restriction digest and ligation of PCR products and vectors

Once the *DIB1* fragments had been successfully amplified, they needed to be digested with restriction enzymes and ligated together with different vectors in order to be able to express mutant *dib1* in *S. cerevisiae* and purify Dib1 proteins from *E. coli.* The restriction digests were designed so that they would form sticky ends on the vectors that were complementary to sticky ends formed upon restriction digest of the *DIB1* PCR products. This allowed for the *DIB1* PCR products to be inserted into the appropriate vector. The gel

visualizing results from the restriction digest showed that the cut vector DNA did not run as far into the gel, indicating that the previously supercoiled vector DNA was now linear and the digest was successful (Figure 3.3). This same visualization cannot be done upon digestion of PCR products since they are already linear and only a few nucleotides are being removed. However, since we were using the same restriction enzymes to cut the PCR products as the vectors, we assumed that the restriction digest that worked for the vectors would work for the PCR products if the same parameters were used.

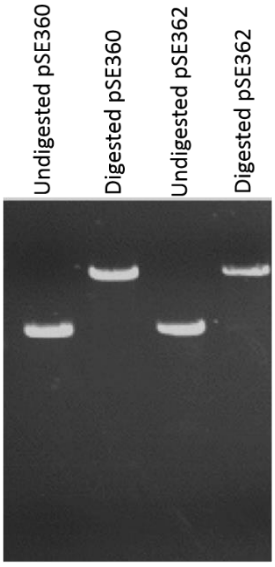

Figure 3.3. Restriction digest of vectors pSE360 and pSE360. Upon digestion by restriction enzymes, a significant apparent increase in molecular weight was seen, indicative of the change from supercoiled DNA to linear DNA upon digestion.

Once the vectors and PCR products were digested, the PCR products were cloned into their corresponding vector. Since pSE360 and pSE362 are expression vectors that allow for the expression of a specific gene in *S. cerevisiae*, the *DIB1* cassette was cloned into these vectors to ensure that the regulatory elements that were upstream or downstream of the actual coding portion of the gene would be included. Since pET15b is purely a protein purification vector and already contains a T7 promoter region upstream of the cloning site,

only the coding region of *DIB1* was needed and so *DIB1* CDS was cloned into this vector. Once cloning was complete and the newly created plasmids were purified from *E. coli*, we verified that the plasmids contained both the *DIB1* PCR product and the vector by digesting a small amount of plasmid with the same restriction enzymes that we used initially. This separates the *DIB1* insert and vector into two fragments that can be visualized on an agarose gel. Upon analysis of each plasmid, it was found that the *DIB1* genes were successfully inserted into each vector (Figure 3.4).

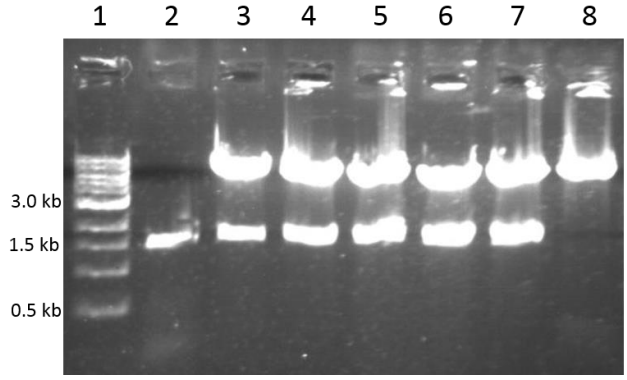

Figure 3.4. Confirmation of cloning of *DIB1* into pSE360. Lane 1: 1 kb DNA ladder; Lane 2: full length *DIB1* only; Lane 3-7: pSE360-*DIB1* candidates digested with Sac1 and Xba1; Lane 8: pSE360 vector only. Plasmids resulting from cloning were digested with restriction enzymes to separate the vector and *DIB1* insert and then visualized on a gel. If a band was seen corresponding to both *DIB1* and the corresponding vector, it was determined that cloning was successful and those plasmids were sent for sequencing.

3.3 Creation of mutant *dib1* expression plasmids

After creation of the two *DIB1* expression plasmids for *S. cerevisiae,* mutagenic PCR was performed on the pSE362-*DIB1* plasmid to create *dib1* mutants. This plasmid was chosen because it has a different selectable marker (*HIS3)* than the wild type pSE360-*DIB1* plasmid (*URA3)* and this marker can be utilized later in a plasmid swap. Mutagenic PCR was done as stated in section 2.3 to create mutants D16A & Q17A, C39A, D63A, L76A & D78A, F85A, K89A, K102A, D115A, N127G, N127D, and P134A (Figure 3.5)(Table 3.1).

Mutants E22A, R25G, D43A, D43A & E44A, Y74A & E75A, L76A, D78A, N88A, and Q111A were also attempted but primer dimers and primer secondary structure interfered and mutagenic PCR was unsuccessful. Further optimization will be necessary to isolate these mutants. All of these mutations were chosen based on conservation of the amino acid throughout species, its corresponding location in the human Dim1 structure, and sites of possible protein-protein interaction.[53,54]

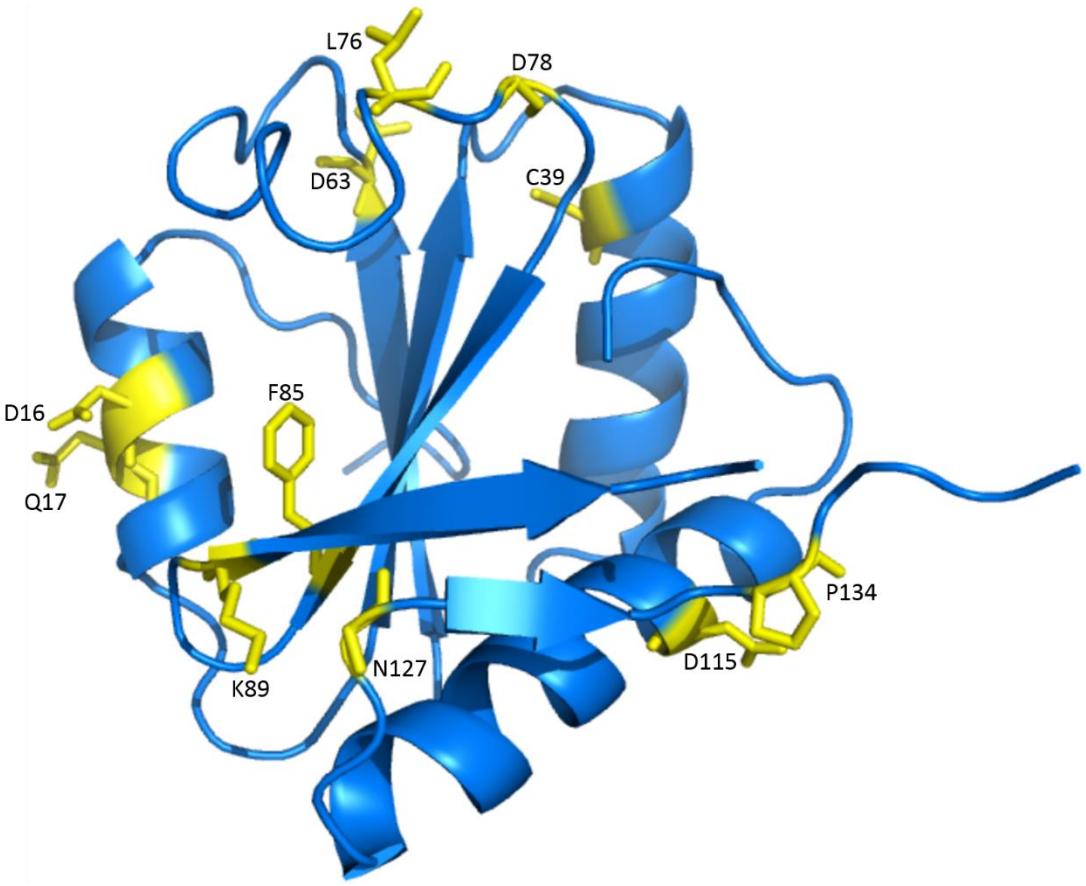

Figure 3.5. Crystal structure of Dim1 with corresponding mutations made in Dib. Amino acids highlighted in yellow were chose for mutational analysis and mutants were created using mutagenic PCR. Figure created using Pymol (PDB entry 1QGV[53]).

3.4 Creating a *dib1Δ* yeast strain

In order to analyze the effects of the mutations made in *dib1,* it was imperative to express them in a *S. cerevisiae* that does not contain its own copy of wild type *DIB1* in its

genome. This way, any phenotypes observed are solely from the mutation in *dib1* and not expression of any other copy of *DIB1* in the yeast's own genome. A *dib1*∠ yeast strain was created by transforming the pSE360-*DIB1* plasmid into a *DIB1*/*dib1*∠ diploid *S. cerevisiae* strain, stressing the resulting cells until they sporulated, separating the spores of the tetrads, and selecting for cells that lack *DIB1* in their genome and retained the pSE360-*DIB1* plasmid. The resulting strain was used in the serial dilution growth assay experiments.

3.5 Analysis of mutant *dib1* effects on cell viability

To analyze the effects of *dib1* mutations on cell viability in *S. cerevisiae,* transformation of the mutant pSE362-*dib1* plasmids, removal of the wild type *DIB1* plasmid, and serial dilution growth assays were performed. In the serial dilution growth assay, cells underwent a 6-fold serial dilution in a 96-well microtiter dish, were plated on YEPD plates using a pronger, and these plates were incubated at different temperatures to test for cold and temperature sensitivity. The results can be seen in Figures 3.6-3.8 and Table 3.1. Mutants L76A D78A, D16A Q17A, and F85A all show temperature sensitive phenotypes when grown at restrictive temperatures (37 °C). Mutant K89A was dominantly lethal immediately

upon transformation into the *dib1*Δ *S. cerevisiae* strain no matter what conditions or temperatures were used.

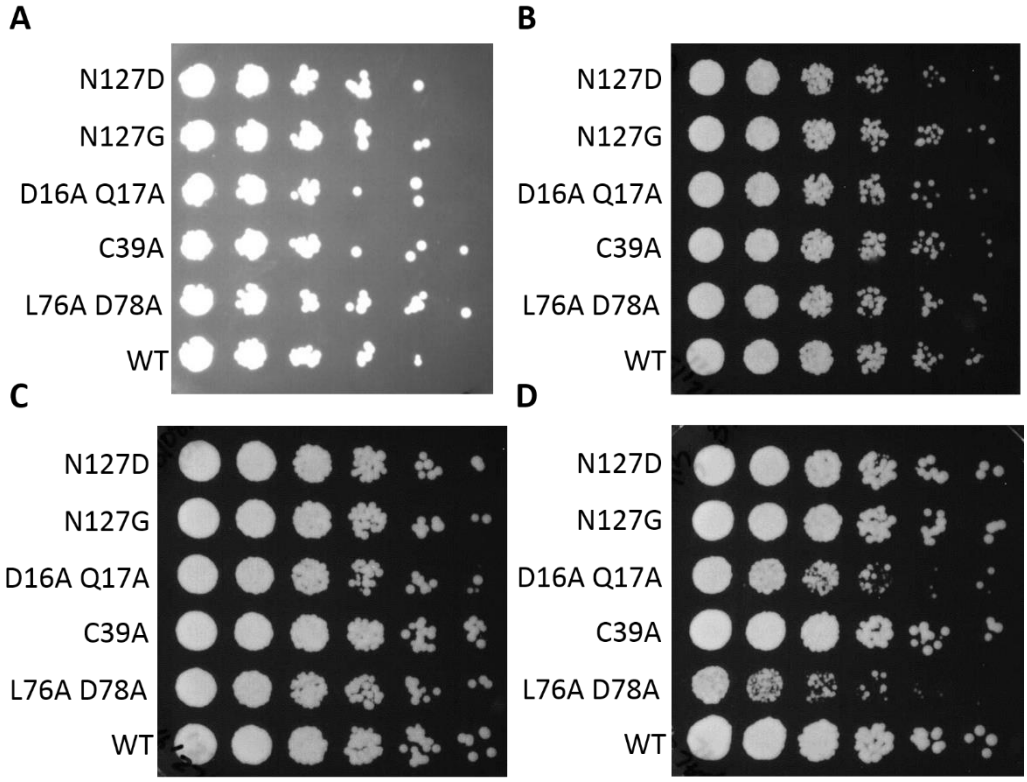

Figure 3.6. Serial dilution growth assay of mutants N127D, N127G, D16A Q17A, C39A, and L76A D78A. The left most column was plated at an A_{600} of 0.2 and each subsequent column is a 6-fold serial dilution. Plates were grown at 16 °C (A), 25 °C (B), 30 °C (C), and 37 °C (D). Mutants L76A D78A and D16A Q17A showed temperature sensitive phenotypes at restrictive temperatures.

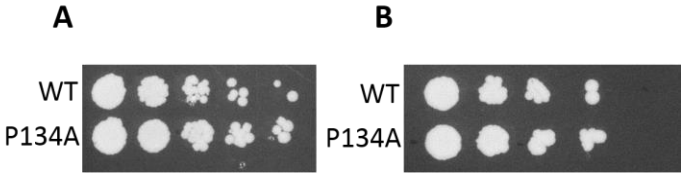

Figure 3.7. Serial dilution grow assay of *dib1* mutant P134A grown at 30 °C (A) and 37 °C (B). Mutation P134 in *dib1* does not result in any observable phenotype other than wild type when grown at a restrictive temperature.

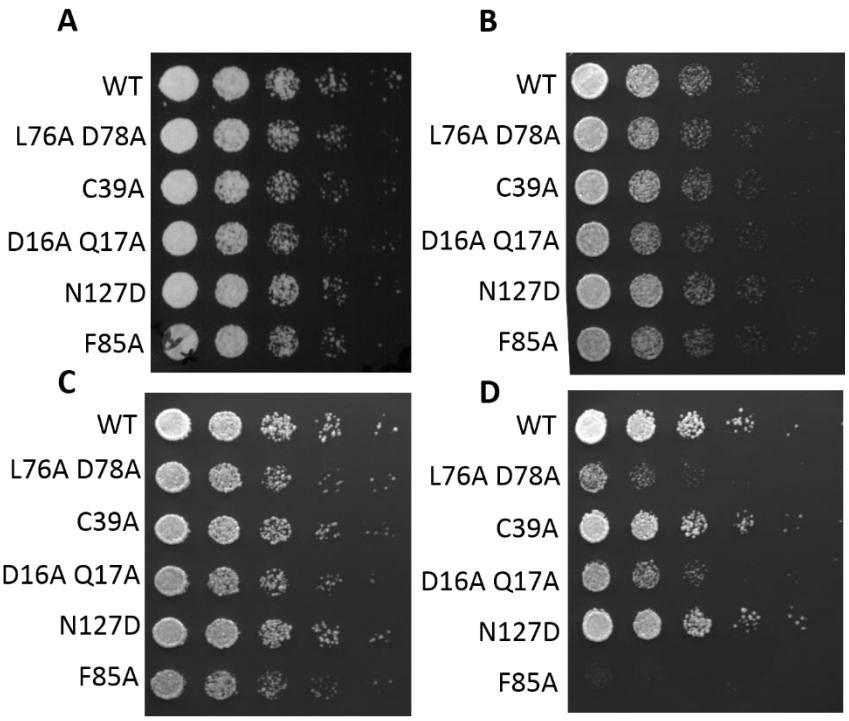

Figure 3.8. Serial dilution growth assay of mutants L76A D78A, C39A, D16 Q17A, N127D, and F85A. The left most column was plated at an A_{600} of 0.2 and each subsequent column is a 6-fold serial dilution. Plates were grown at 16 °C (A), 25 °C (B), 30 °C (C), and 37 °C (D). Mutant F85A showed slight temperature sensitivity at 30 °C and mutants L76A D78A, D16A Q17A, and F85A showed temperature sensitivity at 37 °C.

Table 3.1. Mutations made in *DIB1* and their resulting phenotypes

Nucleotide changes	**Amino acid change**	**Phenotype**
C226G, T227C, G228T, A233C	L76A, D78A	temperature sensitive
C400G, A402T	P134A	wild type
T115G, G116C	C39A	wild type
A47C, C49G, A50C, G51T	D16A, Q17A	temperature sensitive
A344C, C345T	D115A	wild type
A304G, A305C, G306T	K102A	wild type
A379G, A380G	N127G	wild type
A379G	N127D	wild type
A188C	D63A	wild type
T253G, T254C	F85A	temperature sensitive
A265F, A266C	K89A	dominant lethal

Mutation L76A & D78A was made in a conserved region of an extended loop region connecting β-sheets 2 and 3. Neither of these residues have been previously characterized. Mutation D16A Q17A was made in α-helix 1, in which ten of the eleven residues are conserved. In humans, D16 is also located in the region where PQBP1 has been shown to interact with Dim1 and could lead to the stability of this interaction.[64] These residues are also previously uncharacterized. Mutation F85A is in β-sheet 3, of which every residue is conserved. This residue is also part of a hydrophobic groove that is formed with W12, V14, I18, L19, F30, and F69 which could be another area of interest for molecular interactions (Figure 3.9).[53] Mutation K89A, which showed a dominant lethal phenotype at all temperatures, is located at the end of β-sheet 3. It is part of a basic patch consisting of residues R86, R121, R124, K125, and R127 and is highly conserved (Figure 3.10). In a previous study with human Dim1, it was shown that the double mutant R86A and K88A (corresponding to K89A in Dib1) was also dominant lethal.[55] Our study not only confirms their results, but also shows that the double mutant is not necessary for this phenotype; K89A alone is enough to confer dominant lethality.

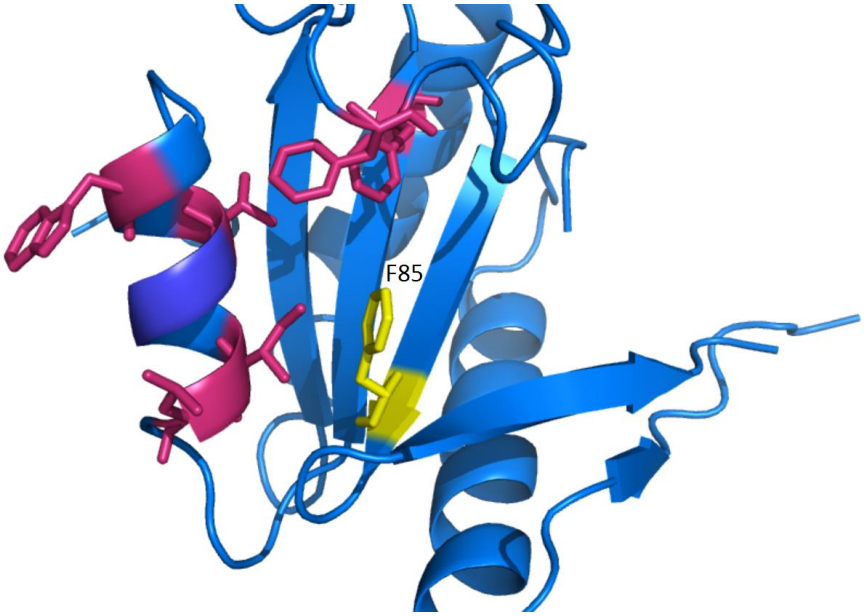

Figure 3.9. Crystal structure of Dim1 showing the location of F85. Residue F85 is shown here (yellow) in relation to the hydrophobic groove (magenta). Figure created using Pymol and PDB entry 1QGV.[53]

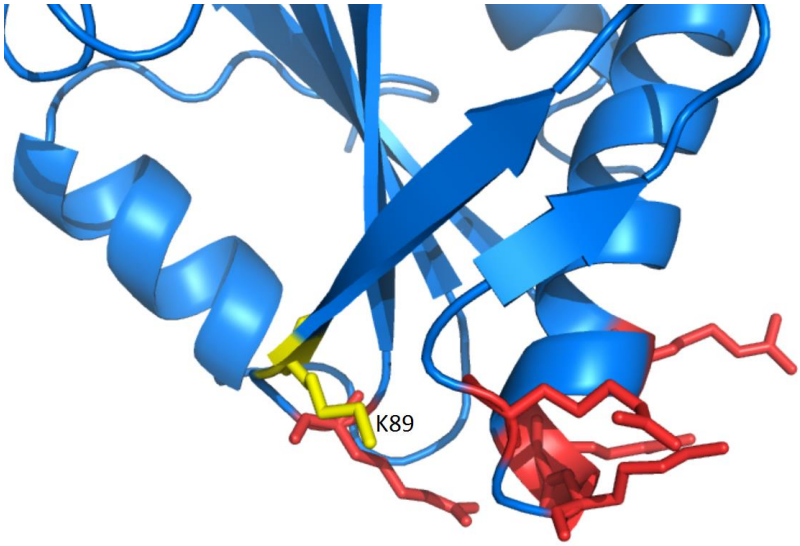

Figure 3.10. Crystal structure of Dim1 showing the location of K89. Residue K89 (yellow) is shown here in relation to the basic patch (red). Figure created using Pymol and PDB entry 1QGV.[53]

Amino acid G127 in humans is conserved throughout all species except *S. cerevisiae*, where this glycine is replaced by an asparagine. One of the first mutants studied in human Dim1 was G127D which results in temperature sensitivity.[57] The results shown in Figure 3.8 suggest that even though this residue may be critical in the human Dim1, it is not critical in *S. cerevisiae* Dib1. In fact, mutation to both the glycine found in Dim1 and to the aspartic acid found in the Dim1 mutant both had no effect.

Most proteins with thioredoxin folds exhibit oxidoreductase activity and have a cys-X-X-cys domain that is conserved and necessary for activity.[78] The two cysteines in the active site domain form a disulfide bond that is also necessary for activity.[78] In Dib1 this domain lacks the first cysteine in the Cys-x-x-Cys domain and instead has an aspartic acid (D36) in its place. It was hypothesized that Cys79, which is in close proximity to Cys39, may make up for the lack of the first cysteine in the Cys-x-x-Cys domain and form a disulfide bond to cys39 to form a thioredoxin-like active site. A previous studies by Zhang *et al.* showed that this disulfide bond formation is energetically unfavorable and that mutations in *S. pombe* Dim1 showed no effect on Dim1 function.[54,55] The results shown in Figure 3.6 and 3.8 lend support to the idea that C39 does not participate in any oxidation reduction reactions and is not required for Dib1 activity.

3.6 Purification of Dib1 and mutant Dib1 proteins

The observed effects on growth seen after mutagenesis could result from several molecular scenarios. The disruption of interactions between the protein and another molecule, disruption of a critical active site, or disruption of the 3-dimensional structure of the protein could all cause a biological phenotype. To rule out folding issues in the protein, Dib1 protein was purified and structural characterization was begun. Wild type Dib1, mutant K89A, and mutant F85A were all purified from an *E. coli* expression system.

The first step of the purification process was to make sure that the Dib1 expression plasmid was working correctly by performing a small scale induction assay. The vector that Dib1 was cloned into (pET15b) contains a gene that encodes T7 polymerase under the control of a lac operon, a *lacI* gene that codes for the repressor of the lac operon, and a T7 promoter directly upstream from a histidine tag followed by a thrombin cleavage site and the gene of interest. This promoter allows for LacI to inhibit production of T7 polymerase, which then in turn inhibits production of the gene of interest. IPTG binds LacI, preventing it from repressing the production of T7 prolymerase and thus resulting in efficient transcription of the gene of interest which can then be translated. The small scale induction assay was done to ensure that protein production was blocked in the absence of IPTG and that protein production was induced in the presence of IPTG. The induction assay showed that Dib1 was produced after induction with IPTG and the results can be seen in Figure 3.11.

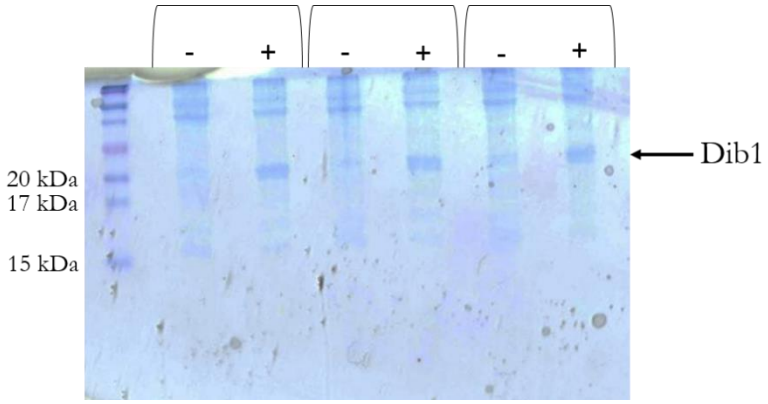

Figure 3.11. Small scale induction assay. Three separate colonies containing the Dib1 expression plasmid were inoculated and then split off into two groups: one that was induced with 25 mM IPTG (+) and one that was not (-). This figure shows that in those cultures that were induced with IPTG, Dib1 was successfully produced. Lane 1: ladder; Lane 2: Culture one without IPTG; Lane 3: Culture one with IPTG; Lane 4: Culture two without IPTG; Lane 5: Culture two with IPTG; Lane 6: Culture three without IPTG; Lane 7: Culture three with IPTG.

After it was shown that the expression plasmid was working correctly, the prepared cell lysate was run over a nickel affinity column. Since Dib1 was cloned into a pET15b vector which adds a histidine tag and a thrombin cleavage site at the N-terminus, it was able be purified from cell extracts using nickel affinity chromatography. The Ni-NTA resin used is made of agarose beads with nitrolotriacetic acid groups which chelates Ni^{2+} very strongly. The histidine tag on Dib1 interacts strongly with the chelated Ni^{2+} and sticks to the column while most other contaminants flow through. Dib1 can then be eluted using imidazole, another molecule with a high affinity for the chelated Ni^{2+}. Results from this first step of the purification process showed that if the pH of the wash buffers was not adjusted prior to washing of the column, Dib1 eluted at an imidazole concentration of 100 mM (Figure 3.12 B). After the pH of the buffers was kept at a constant pH of 7, Dib1 eluted at an imidazole concentration of 250 mM (Figure 3.12 A & C).

The pET15b vector includes a thrombin cleavage site immediately following the histidine tag. This allows for the removal of the tag, which could interfere with the three-dimensional structure of Dib1 if left attached. Dib1 was incubated with thrombin for 2 h at room temperature and removal of the histidine tag was successful, resulting in a change in molecular weight from ~17 kDa to ~15 kDa which was monitored using SDS-PAGE (Figure 3.13).

Dib1 has an estimated isoelectric point of 5.8 (calculated by ProtParam[79]) meaning when Dib1 is at a pH higher than 5.78, it is negatively charged. This allowed for the utilization of anion exchange chromatography to further purify Dib1. The Q-Sepharose beads that were used are composed of agarose beads with quaternary ammonium groups. These positively charged groups bind negatively charge molecules very tightly while the remaining contaminants flow through. Dib1 was then eluted using high concentrations of

NaCl. Results of this purification step can be seen in Figure 3.14. Dib1 was found to elute after the 150 mM NaCl wash with high yield and high purity.

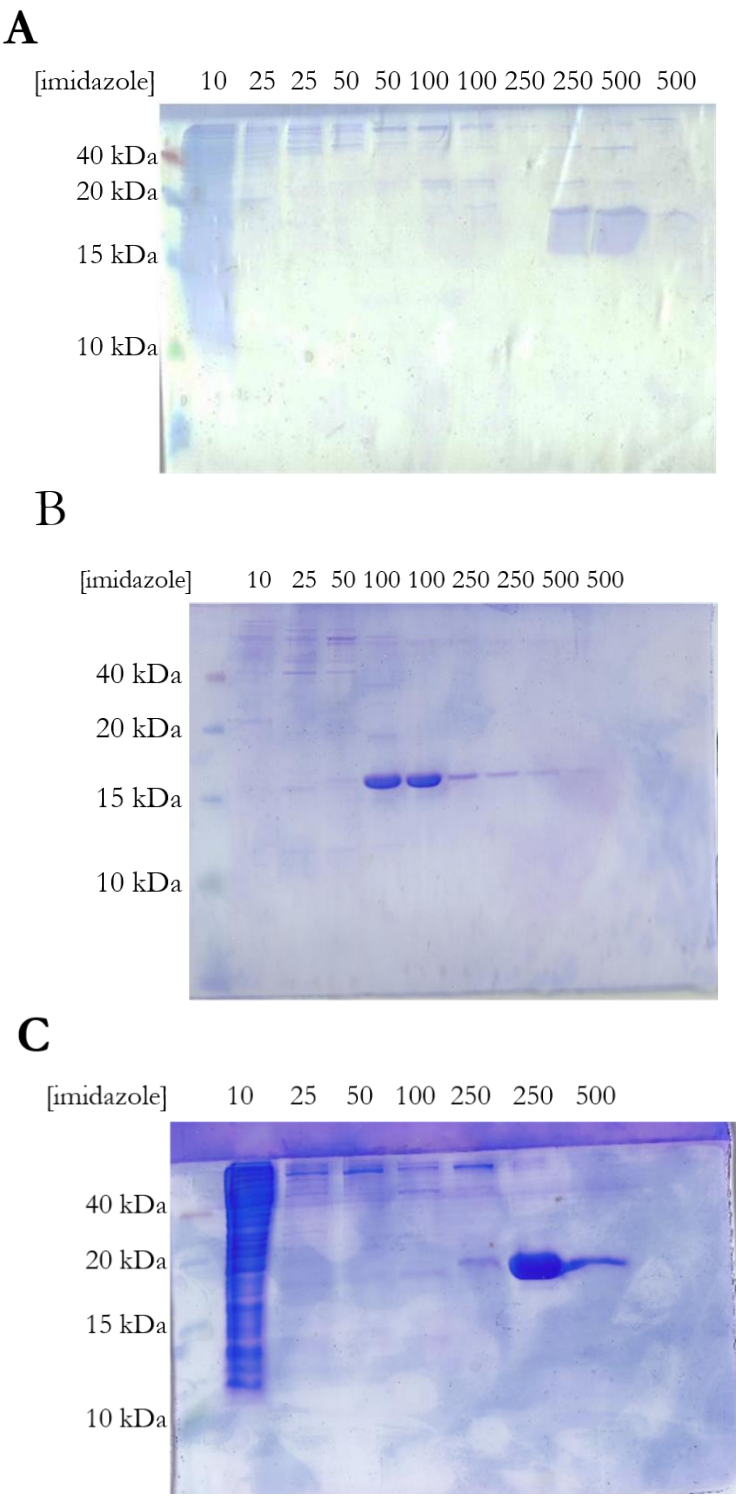

Figure 3.12. Nickel affinity chromatography of wild type Dib1 (A), Dib1 mutant F85A (B), and Dib1 mutant K89A (C). Without adjusting the pH of the elution buffers, Dib1 was found to elute after the 100 mM imidazole wash (B). When the pH was changed to a consistant pH of 7, Dib1 was found to elute after the 250 mM imidazole wash (A and C).

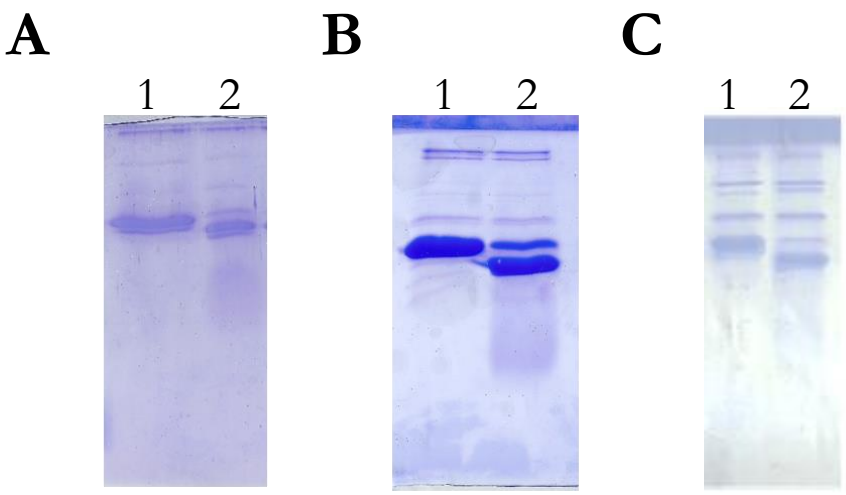

Figure 3.13. Results of the thrombin digest. Thrombin digests were done with wild type Dib1 (A), Dib1 mutant F85A (B), and Dib1 mutant K89A (C). Digestion with thrombin resulted in a shift to a lower molecular weight which was visible in each gel. Gels A, B, and C: Lane1: Dib1 with the histidine tag; Lane 2: Dib1 after removal of the histidine tag.

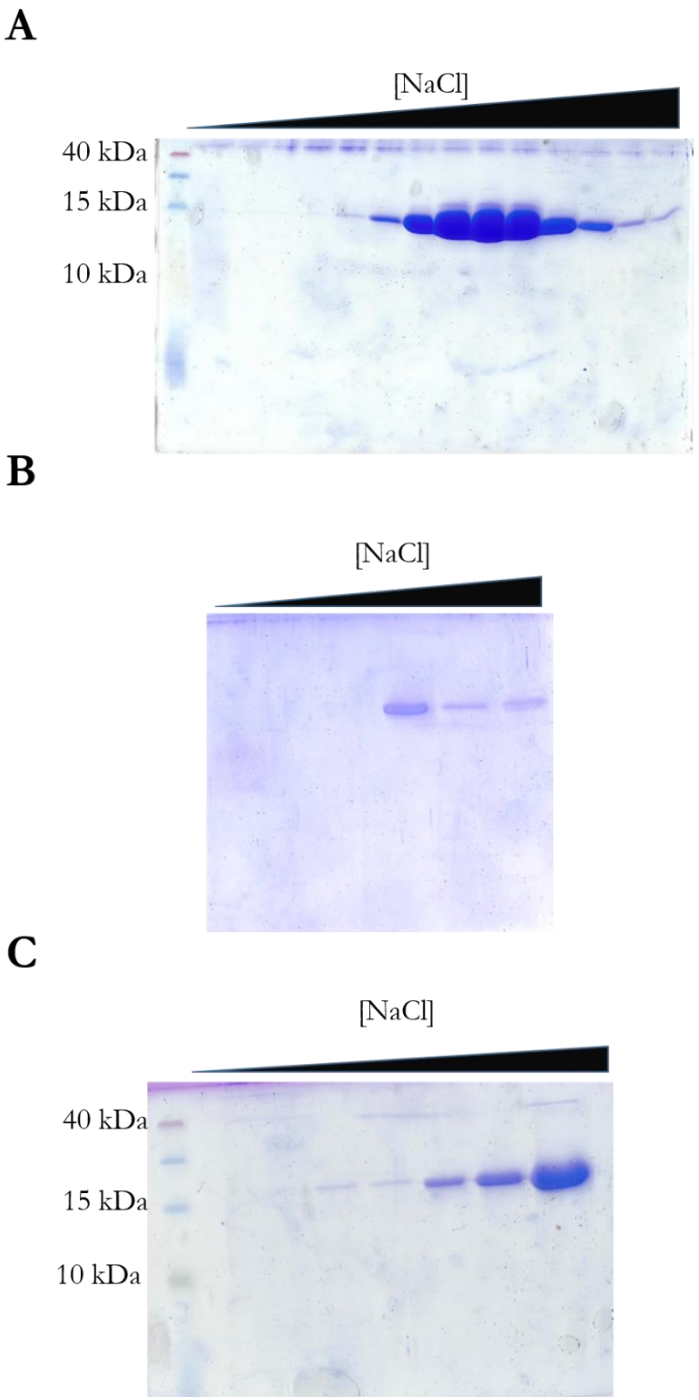

Figure 3.14. Ion exchange chromatography of wild type Dib1 (A), Dib1 mutant F85A (B), and Dib1 mutant K89A (C). Dib1 and mutants eluted after the 150 mM NaCl wash with very few contaminants. The band that can be seen around 40 kDa in gels A and C is a result of precipitation of Dib1 and can be removed via filtration.

3.7 CD Spectroscopy of Dib1 and Dib1 mutants

3.7.1 CD spectra of Dib1 and Dib1 mutants

To determine if Dib1 mutants still folded properly, the structure of the wild type and mutant Dib1 proteins were analyzed using circular dichroism spectroscopy. Every chiral molecule rotates right and left polarized light differently. Since every amino acid has a chiral center in its backbone, we can take advantage of this difference in light rotation to tell us about the secondary structure of the protein. Every protein has a distinct CD spectroscopy fingerprint depending on whether the protein displays α-helix characteristics, β-sheet characteristics, or random coil characteristics. If the protein is primarily α-helix, it will have a characteristically large peak at 190 nm and will have two dips: one at 210 nm and one at 220 nm. If the protein is primarily β-sheet, it will have a moderate peak at 195 nm and a single dip at 215 nm. If the protein is primarily random coil, it will have a dip at 195 nm and will rise to plateau out at 210 nm.

From the previous crystallography and CD spectroscopy studies in human Dim1, it was expected that *S. cerevisiae* Dib1 would have both α-helix and β-sheet characteristics.[53,55] Our results showed an almost identical pattern between the CD spectrum of human Dim1 from a previous study by Zhang, *et al.* and the *S. cerevisiae* Dib1 in this study (Figure 3.15). Both proteins show a high peak around 200 nm, a pronounced, low dip around 220 nm, and

a very slight dip around 210 nm. This pattern supports the idea that *H. sapiens* Dim1 and *S. cerevisiae* Dib1 are similar in structure and contain both α-helical and β-sheet elements.

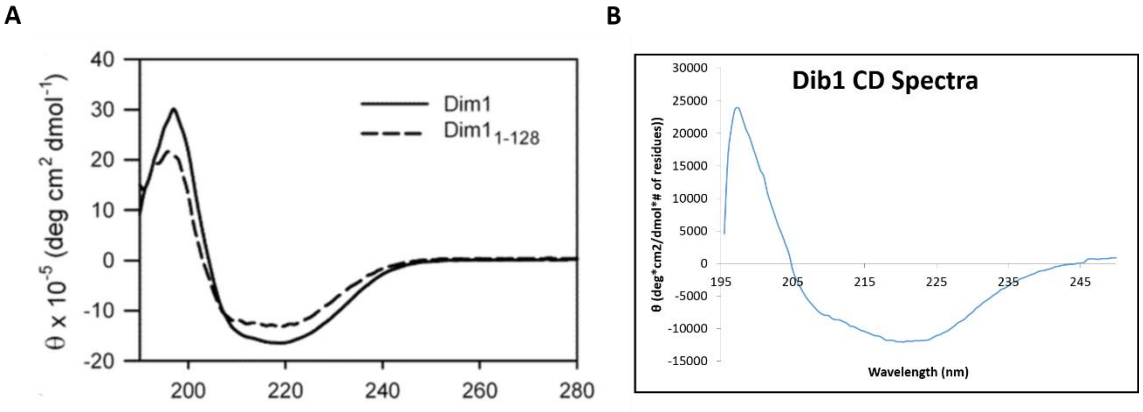

Figure 3.15. Comparison of the CD spectra of H. sapiens Dim1 and S. cerevisiae Dib1. The human Dim1 spectra was reprinted from a previous study by Zhang *et al.*.[55] In graph A, the solid line represents the wild type Dim1, while the dashed line represents a Dim1 mutant lacking the C-terminal tail. For the purposes of this comparison, the focus should be on only the wild type Dim1. Both wild type Dim1 (A) and Dib1 (B) contain a high peak at 200 nm, a trough with a low point at 220 nm, and a very slight dip at 210 nm.

A comparison of wild type and mutant Dib1 spectra might explain whether the phenotypes seen with the mutation are caused by disruption of Dib1's structure or not. Dib1 mutant F85A showed a great loss in secondary structure, determined by the large loss of the intensity of the peak at 200 nm and the trough at 220 nm (Figure 3.16). This suggests that residue F85 is critical for structure rather than function. When F85 is mutated to an alanine, as the temperature increases the Dib1 structure becomes unstable and unfolds, resulting in temperature sensitivity. Dib1 mutant K89A showed a small gain in β-sheet secondary structure, shown by its slight decrease in the peak at 200 nm and the slight increase in the trough at 220 nm (Figure 3.16). While it is possible this slight change in secondary structure could have important ramifications, it is unlikely that it is the change in structure that leads to this mutant's lethality. K89A's close resemblance to the wild type indicates this residue is likely in a critical area of molecular interactions involving Dib1 or is

in a critical active site. Further experimentation will have to be done to determine the exact nature of this residue.

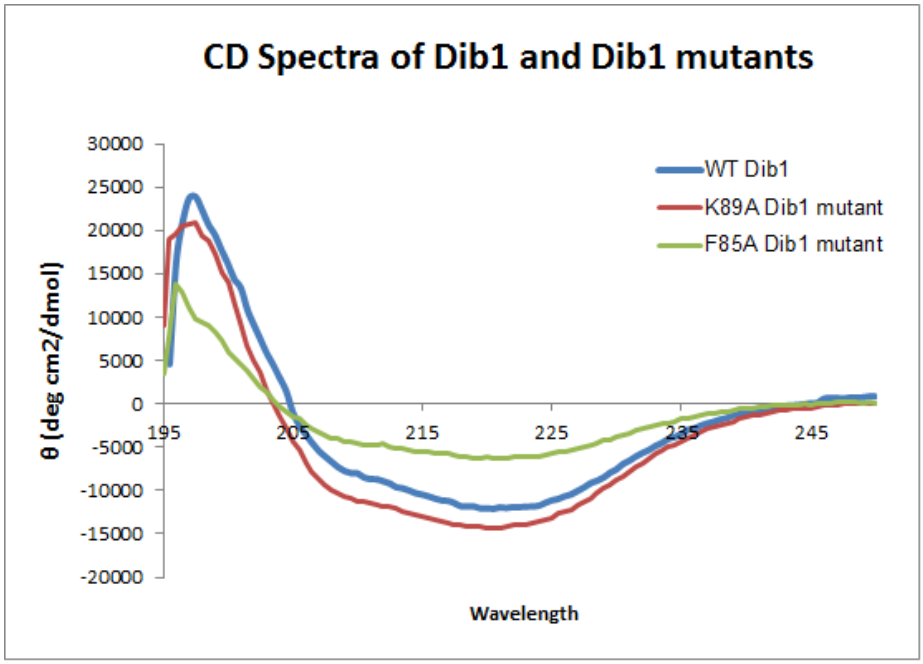

Figure 3.16. CD spectra of wild type Dib1 (blue) and Dib1 mutants K89A (red) and F85A (green). Dib1 mutant F85A showed a significant decrease in secondary structure. This indicates that residue F85 is critical for Dib1 structure stability. Dib1 mutant K89A showed a small gain in β-sheet secondary structure but this slight change is unlikely to be the reason for lethality in the K89A mutant.

3.7.2 Thermal melt studies of Dib1 and Dib1 mutants

It was observed that upon heating of Dib1, Dib1 precipitated out of solution. In order to do an accurate calculation of thermodynamic properties, the thermal denaturation of Dib1 must be reversible, meaning that upon cooling the protein returns to its original folded state. Upon obtaining the CD spectra of the thermal melt and cooling, it was found that refolding of the protein was not occurring (Figure 3.17). It seems that upon heat denaturation, the protein precipitates and falls out of solution and remains in this denatured state even after cooling back to room temperature. In order to attempt to prevent Dib1 precipitation, a thermal stability study was done using different buffers. Even after

changing the salt composition, concentrations, the pH, and components of the buffer Dib1 still precipitated. Because of this, the thermal denaturation studies can no longer be used, but rather this study will move forward using urea denaturation instead of thermal denaturation.

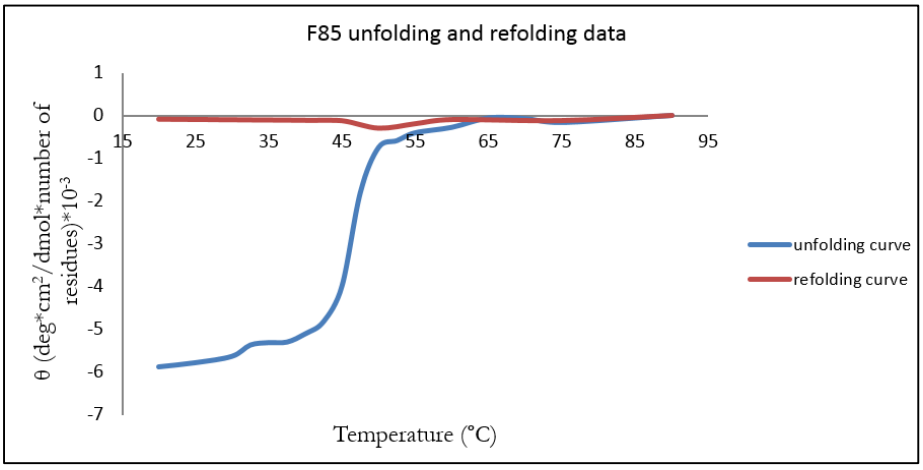

Figure 3.17. Unfolding and refolding data for Dib1 mutant F85A. Upon cooling to 20 °C the protein remained in an unfolded state, indicating that the thermal denaturation of Dib1 is not reversible. Because of its irreversibility, the data from the thermal unfolding curve cannot be used to calculate the thermodynamic properties of Dib1 and its mutants.

CHAPTER 4

CONCLUSIONS

Dib1 is a 15 kDa protein associated with U5 of the U4/U6·U5 triple snRNP of the spliceosome. It is conserved from *S. cerevisiae* to *H. sapiens* and is necessary for cell viability. Dim1, the human homolog of Dib1, has been crystallized and shown to have a thioredoxin-like fold.[53] This means it has a four β-sheet core surrounded by three α-helices and a 14 amino acid C-terminal tail with flexibility to adopt either β-sheet or α-helix properties depending on its environment. Despite having a conserved thioredoxin fold, it does not have oxidation reduction activity as shown by mutational studies by Reuter *et al.*.[53]

Very little is known about Dib1's biochemical function in the spliceosome. Many previous studies with human and *S. pombe* Dim1 have identified it as necessary for entry into mitosis, necessary for splicing, necessary for export of certain mRNAs from the nucleus, as having autopeptidase activity, and as an interactor with PQBP1, hnRNP F, hnRNP R^+, and Prp6.[54,54,56–59,62] We hypothesize that Dib1 acts as an allosteric inhibitor of spliceosome activation until all the components have been recruited and are in the correct orientation for splicing to occur. Once all the components are oriented properly, a conformational change occurs, Dib1 is released, and the spliceosome becomes catalytically active. This project described here is the first step of many to test this hypothesis. We chose the model organism *S. cerevisiae* to study Dib1, the homolog of the human Dim1 because this model organism provides a simpler model of the human spliceosome while maintaining high conservation. The goals of this study were to identify amino acid residues critical for Dib1

activity, purify Dib1 protein, and use CD spectroscopy to analyze Dib1 and mutant Dib1 secondary structure.

In order to identify amino acid residues critical for Dib1 activity, mutagenic PCR was used to create specific point mutations in *DIB1*. Once mutations were created, these mutant *dib1* genes were expressed in a *S. cerevisiae dib1⊿* strain and a serial dilution growth assay was performed to analyze any effects these mutations had on cell viability. Eleven *dib1* mutants were created, and from these four temperature sensitive mutants were identified: D16A Q17A, L76A D78A, and F85A. Another mutant, K89A, was lethal at all temperatures. The remaining mutants all showed wild type phenotypes and so it was concluded that they had no observable effect on Dib1 structure or activity. This includes mutant C39A which is the cysteine that is normally involved in redox reactions in proteins with thioredoxin folds and mutant N127D which was shown to result in temperature sensitivity in *S. pombe* Dim1 studies.[53,57] Our results show that *S. cerevisiae* Dib1 also does not have thioredoxin activity and that amino acid N127 is not necessary for Dib1 function like it is in Dim1.

In order to purify Dib1 from *E. coli*, *DIB1* and *dib1* mutants F85A and K89A were cloned into a pET15b vector. This vector allowed for production of Dib1 protein in *E. coli* with the addition of a histidine tag and a thrombin cleavage site at the N terminus of the protein. This histidine tag was utilized by first performing nickel affinity chromatography followed by cleavage and removal of the histidine tag and ion exchange chromatography to further purify Dib1.

In order to analyze the secondary structure of Dib1 and Dib1 mutants F85A and K89A, circular dichroism was utilized. Based on the CD spectra obtained, it was shown that Dib1 is almost identical in its secondary structure to its human homolog Dim1. Dib1 mutant F85A had a significant loss in secondary structure, indicating that this amino acid

residue is likely critical for protein stability and structure. Dib1 mutant K89A was very similar in structure to the wild type with a slight gain in β-sheet secondary structure. This indicates that it is unlikely that K89 is critical for structure and stability, but rather is more likely important for molecular interactions between Dib1 and other molecules or part of a Dib1 active site.

In the future, mutants L76A D78A and D16A Q17A should be separated into single mutations in order to determine if one of the two residues is truly causing the temperature sensitive phenotype. These two mutants should also be purified and undergo CD spectroscopy in order to compare them to the wild type as done with F85A and K89A in order to get insight into their structure. In addition, denaturation curves could be obtained using CD spectroscopy in order to further analyze the thermodynamic properties of the wild type Dib1 and its mutants. Splicing assays and splicing microarrays could also be done to analyze which step of splicing is being disrupted in Dib1 mutants. All of the results presented here and work proposed for future studies together narrow in on identifying amino acid residues critical for Dib1 function in the spliceosome.

REFERENCES

1. Bachenheimer, S. & Darnell, J. E. Adenovirus-2 mRNA is transcribed as part of a high-molecular weight precursor RNA. *Cell Biol.* **72,** 4445–4449 (1975).
2. Chow, L. T. & Broker, T. R. The spliced structures of adenovirus 2 fiber message and the other late mRNAs. *Cell* **15,** 497–510 (1978).
3. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12,** 1–8 (1977).
4. Hoskins, A. A. & Moore, M. J. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.* **37,** 179–188 (2012).
5. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
6. Faustino, N. A. & Cooper, T. A. pre-mRNA splicing and human disease. *Genes Dev.* **17,** 419–437 (2003).
7. Spingola, M., Grate, L., Haussler, D. & Ares, Jr., M. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisae. *RNA* **5,** 221–234 (1999).
8. Wood, V. *et al.* PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.* **40,** D695–699 (2011).
9. Brow, D. A. Allosteric cascade of spliceosome activation. *Annu. Rev. Gen.* **36,** 333–360 (2002).
10. Mattaj, I. W. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46,** 905–911 (1986).
11. Kiss, T. Biogenesis of small nuclear RNPs. *J. Cell Sci.* **117,** 5949–5951 (2004).
12. Huber, J. *et al.* Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO* **17,** 4114–4126 (1998).
13. Jady, B. E. *et al.* Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal bodies following import from the cytoplasm. *EMBO J.* **22,** 1878–1888 (2003).
14. Sleeman, J. E. & Lamond, A. I. Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* **9,** 1065–1074 (1999).
15. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4,** 605–612 (2003).
16. Achsel, T. *et al.* A doughnut-shaped heteromer of human Sm-like proteins binds to the 3' end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.* **18,** 5789–5802 (1999).
17. Ganot, P., Jady, B. E., Bortolin, M. L., Darzacq, X. & Kiss, T. Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol. Cell Biol.* **19,** 6906–6917 (1999).
18. Stanek, D., Rader, S. D., Klingauf, M. & Neugebauer, K. M. Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.* **160,** 505–516 (2003).
19. Meister, G., Eggert, C. & Fischer, U. SMN-mediated assembly of RNPs: a complex story. *Trends Cell Bio.* **12,** 472–478 (2002).

20. Kramer, A. Function of Proteins Involved in Mammalian pre-mRNA Splicing. *Annu. Rev. Biochem.* **65,** 367–409 (1996).
21. Green, M. R. pre-mRNA splicing. *Annu. Rev. Gen.* **20,** 671–708 (1986).
22. Langford, C. J. & Gallwitz, D. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* **33,** 519–527 (1983).
23. Gao, K., Masuda, A., Matsuura, T. & Ohno, K. Human branch point consesnsus sequence is yUnAy. *Nucleic Acids Res.* **36,** 2257–2267 (2008).
24. Berglund, J. A., Rosbash, M. & Schultz, S. C. Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. *RNA* **7,** 682–691 (2001).
25. Roscigno, R. F. & Garcia-Blanco, M. A. SR proteins escort the U4/U6.U5 tri-snRNP to the Spliceosome. *RNA* **1,** 692–706 (1995).
26. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current Biol.* **8,** 847–855 (1998).
27. Staley, J. P. & Guthrie, C. An RNA Switch at the 5' Splice Site Requires ATP and the DEAD Box Protein Prp28p. *Mol. Cell* **3,** 55–64 (1999).
28. Arenas, J. E. & Abelson, J. N. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *PNAS* **94,** 11798–11802 (1997).
29. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5,** 908–919 (2004).
30. Schwartz, S. *et al.* Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **18,** 88–103 (2008).
31. Johnson, T. L. & Vilardell, J. Regulated pre-mRNA splicing: the ghostwriter of the eukaryotic genome. *Biochem. Biophys. Acta.* **1819,** 538–545 (2012).
32. Fabrizio, P. *et al.* The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol. Cell* **36,** 593–608 (2009).
33. Liu, S., Rauhut, R., Vornlocher, H. P. & Luhrmann, R. The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. *RNA* **12,** 1418–1430 (2006).
34. Lybarger, S. *et al.* Elevated levels of a U4/U6.U5 snRNP-Associated Protein, Spp381p, Rescue a Mutant Defective in Spliceosome Maturation. *Mol. Cell Biol.* **19,** 577–584 (1999).
35. Galisson, F. & Legrain, P. The biochemical defects of prp4-1 and prp6-1 yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Neclic Acids Res.* **21,** 1555–1562 (1993).
36. Kim, R., Paschedag, J., Novikova, N. & Bellini, M. The recruitment of the U5 snRNP to nascent transcripts requires internal loop 1 of U5 snRNA. *Chromosome Res.* **20,** 943–51 (2012).
37. Nottrott, S. *et al.* Functional interaction of a novel 15.5 kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *EMBO J.* **18,** 6119–6133 (1999).
38. Vithana, E. N. *et al.* A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* **8,** 375–81 (2001).
39. Makarov, E. M., Makarov, O. V., Achsel, T. & Luhrmann, R. The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. *J. Mol. Biol.* **298,** 567–575 (2000).
40. Nottrott, S., Urlaub, H. & Luhrmann, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* **21,** 5527–5538 (2002).
41. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262,** 1989–1996 (1993).

42. Newman, A. J. The role of U5 snRNP in pre-mRNA splicing. *EMBO* **16,** 5797–6055 (1997).
43. Kuhn, A. N., Reichl, E. M. & Brow, D. A. Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. *PNAS* **99,** 9145–9149 (2002).
44. Liu, L., Query, C. C. & Konarska, M. M. Opposing classes of prp8 alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat. Struct. Mol. Biol.* **14,** 519–526 (2007).
45. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14,** 343–354 (2004).
46. Wachtel, C. & Manley, J. L. Splicing of mRNA precursors: The role of RNAs and proteins in catalysis. *Mol. Biosyst.* **5,** 311–316 (2009).
47. Brenner, T. J. & Guthrie, C. Genetic Analysis Reveals a Role for the C Terminus of the Saccharomyces cerevisae GTPase Snu114 During Spliceosome Activation. *Genetics* **170,** 1063–1080 (2005).
48. Lauber, J. *et al.* The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisae are members of the DEXH-box protein family of putative RNA helicases. *EMBO* **15,** 4001–4015 (1996).
49. Strauss, E. J. & Guthrie, C. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev.* **5,** 629–641 (1991).
50. Jackson, S. P., Lossky, M. & Beggs, J. D. Cloning of the RNA8 gene of Saccharomyces cerevisae, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. *Mol. Cell Biol.* **8,** 1067–1075
51. Vandenbol, M. & Fairhead, C. Mass-murder deletion of 19 ORFs from Saccharomyces cerevisae chromosome XI. *Gene* **247,** 45–52 (2000).
52. Legrain, P. *et al.* Cloning of the two essential yeast genes PRP6 and PRP9, and their rapid mapping, disruption and partial sequencing using a linker insertion strategy. *Mol. Gen. Genet.* **225,** 199–202 (1991).
53. Reuter, K., Nottrott, S., Fabrizio, P., Luhrmann, R. & Ficner, R. Identification, Characterization, and Crystal Structure Analysis of the Human Spliceosomal U5 snRNP-specific 15 kD Protein. **294,** 515–525 (1999).
54. Zhang, Y. Z. *et al.* The evolutionarily conserved Dim1 protein defines a novel branch of the thioredoxin fold superfamily. *Physiol. Genomics* **1,** 109–118 (1999).
55. Zhang, Y., Cheng, H., Gould, K. L., Golemis, E. A. & Roder, H. Structure, Stability, and Function of hDim1 Investigated by NMR, Circular Dichroism, and Mutational Analysis. **42,** 9609–9618 (2003).
56. Zhang, Y. *et al.* Evidence that Dim1 associates with proteins involved in pre-mRNA splicing, and delineation of residues essential for Dim1 interactions with hnRNP F and Npw38/PQBP-1. *Gene* **257,** (2000).
57. Berry, L. D. & Gould, K. L. Fission Yeast dim1+ Encodes a Functionally Conserved Polypeptide Essential for Mitosis. *J. Cell Biol.* **137,** 1337–1354 (19997).
58. Berry, L. D., Feoktistova, A., Wright, M. D. & Gould, K. L. The Schizosaccharomyces pombe dim1+ Gene Interacts with the Anaphase-Promoting Complex of Cyclosom (APC/C) Component lid1+ and Is Required for APC/C Function. *Mol. Cell Biol.* **19,** 2535 (1999).
59. Jin, T., Guo, F., Wang, Y. & Zhang, Y. Z. Identification of human dim1 as a peptidase with autocleavage activity. *Chem. Biol. Drug Des.* **68,** 266–72 (2006).

60. Simeoni, F. & Divita, G. The Dim protein family: from structure to splicing. *Cell Mol. Life Sci.* **64,** 2079–2089 (2007).
61. Chakshusmathi, G. *et al.* Desgin of temperature-sensitive mutants solely from amino acid sequence. *PNAS* **101,** 7925–7930 (2004).
62. Carnahan, R. H. *et al.* Dim1p Is Required for Efficient Splicing and Export of mRNA Encoding Lid1p, a Component of the Fission Yeast Anaphase-Promoting Complex. *Eukaryotic Cell* **4,** 577–587 (2005).
63. Okazawa, H. *et al.* Interaction between Mutant Ataxin-1 and PQBP-1 Affects Transcription and Cell Death. *Neuron* **34,** 701–713 (2002).
64. Mizuguchi, M. *et al.* Mutations in the PQBP1 gene prevent its interaction with the spliceosomal protein U5-15kD. *Nat. Commun.* **5,** 3822 (2014).
65. Waragai, M. *et al.* PQBP-1/Npw38, a Nuclear Protein Binding to the Polyglutamine Tract, Interacts with U5-15kD/dim1p via the Carboxyl-Terminal Domain. *Biochem. Biophys. Res. Commun.* **273,** 592–595 (2000).
66. Takahasi, M. *et al.* Polyglutamine tract-binding protein-1 binds to U5-15kD via a continuous 23-residue segment of the C-terminal domain. *Biochimica et Biophysica Acta* **1804,** 1500–1507 (2010).
67. Chen, C. D., Kobayashi, R. & Helfman, D. M. Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat B-tropomyosin gene. *Genes Dev.* **13,** 593–606 (1999).
68. Garneau, D., Revil, T., Fisette, J. & Chabot, B. Heterogeneous Nuclear Ribonucleoprotein F/H Proteins Modulate the Alternative Splicing of the Apoptotic Mediator Bcl-x. *J. Bio. Chem.* **280,** 22641–22650 (2005).
69. Min, H., Chan, R. C. & Black, D. L. The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.* **9,** 2659–2671 (1995).
70. Stevens, S. W. *et al.* Biochemical and genetic analyses of the U5, U6, and U$/U6 xU5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. *RNA* **7,** 1543–1553 (2001).
71. Colowick, S. P. & Kaplan, N. O. in *Methods Enzymol.* (eds. Guthrie, C. & Fink) **194,** (Elsevier Inc., 2004).
72. Chung, C. T. & Miller, R. H. A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Res.* **16,** 3580 (1988).
73. Elledge, S. J. & Davis, R. W. A family of versatile centromeric vectors designed for use in the sectoring-shuffle mutagenesis assay in Saccharomyces cerevisiae. *Gene* **70,** 303–312 (1988).
74. Mierendorf, R. C., Morris, B. B., Hammer, B. & Novy, R. E. Expression and Purification of Recombinant Proteins Using the pET System. *Methods Mol. Med.* **13,** 257–292 (1998).
75. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–32 (1998).
76. Steinmetz, L. M. *et al.* Systematic screen for human disease genes in yeast. *Nat. Genetics* **31,** 400–404 (2002).
77. Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* **154,** 164–175 (1987).
78. Eklund, H., Gleason, F. K. & Holmgren, A. Structural and functional relations among thioredoxins of different species. *Proteins* **11,** 13–28 (1991).
79. Gasteiger, E. *et al.* in *Proteomics Protoc. Handb.* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).