

THESIS

SPN1, A HIGHLY CONSERVED AND ESSENTIAL NODE  
OF RNA POLYMERASE II DEPENDENT  
FUNCTIONS

Submitted by:

Adam Almeida

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the degree of: Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2011

Master's Committee:

Advisor: Laurie Stargell

Karolin Luger

Robert Woody

Erica Suchman

## ABSTRACT

### SPN1, A HIGHLY CONSERVED AND ESSENTIAL NODE FOR RNA

### POLYMERASE II DEPENDENT FUNCTIONS

A multitude of proteins are responsible for regulating the activity of RNA Polymerase II (Pol II) in the nucleus of a eukaryotic cell. Two types of themes are used by these proteins to control transcription: recruitment-regulation and postrecruitment-regulation. The main difference between the two is the rate-limiting step for producing transcript. This rate-limiting step for the first mechanism is the recruitment of Pol II to the promoter. For the second mechanism, Pol II constitutively occupies the promoter, is “poised”, and an unknown rate-limiting postrecruitment step prevents transcription from commencing. The highly conserved and essential transcription factor Spn1 was identified as a protein that functions postrecruitment of Pol II and has been characterized for having a direct role at regulating the poised *CYCI* gene in *Saccharomyces cerevisiae*. This activity has been determined from mutations made within the most conserved portion of Spn1 made up of a highly folded central domain. Little is known about the functions of the N-and C-terminal regions flanking this central domain, which is the focus of the work done here. Genetic characterization indicates that these regions have physiologically relevant and important functions within the cell outside of optimum growth conditions, but do not involve significant regulation of the *CYCI* gene. A broader approach of experimentation is likely required to understand all of the Spn1 protein’s functions regarding transcription. This led to the observation that Spn1 is able to bind to nucleosomes in vitro and that this interaction is dependent on the N-and C-terminal regions of the protein. The possibility that Spn1 could affect nucleosome

dynamics in the cell is consistent with the physical and genetic interactions observed between Spn1 and the Spt6 and Swi/Snf histone chaperone and chromatin remodeling complexes. This result will provide several new avenues for future Spn1 research.

A genomic ChIP-chip experiment performed by two independent groups revealed that Spn1 is recruited to a majority of the genes in the yeast genome<sup>1;2</sup>. Evidence indicates that there are multiple, evolutionarily conserved pathways within the cell that are responsible for determining the rate at which an organism will age that include: ribosome biogenesis, protein translation, mitochondrial activity and function, heterochromatic stability, maintenance of the genome, and apoptosis<sup>3;4;5;6;7</sup>. The possibility that Spn1 regulates the genes involved in these pathways is highly suggestive that this protein could be an aging factor within the cell. Chronological aging assays revealed that the removal of the N- and C-terminal regions of the Spn1 protein dramatically increase the lifespan of the BY4741 strain of yeast. These results further verify the physiological importance of this protein and the need for further Spn1 research.

## ACKNOWLEDGEMENTS

Earning my master's degree in biochemistry at Colorado State University in Fort Collins has been the most difficult endeavor I have ever pursued. It has been a true test of my will, work ethic, discipline, indomitable spirit, and mental/emotional stability. Despite all of the challenges I have faced over the last few years, I am very glad that I chose this major and I will continue to work very hard to make a difference in the world with the knowledge and skills that I have obtained. This being said, there are a number of people that have contributed to my achievements and I owe them all a debt of gratitude for the work done here. I would first like to thank the entire Stargell lab for their constant support and guidance. They have all been amazing to work with and I will miss them very much. Dr. Xu Chen has always provided great scientific discussion for my project and has given me a number of good ideas for future experiments and directions. I also thank her greatly for the ChIP experiments that she performed for my project. Dr. Sarah Lee has been a great friend to me over the years as well. I have always been able to go to her if I needed to talk both professionally and about matters in my personal life. I thank her for all of her guidance and I wish her all of the success and happiness in the world. Fellow graduate student Marie Yearling is a great scientist and has been a great outlet for discussing my stresses and frustrations with my work, something that I have found is essential for success in any scientific field. Fellow research scientist Aaron Docter has become a great friend over the last year and a half. He has worked very hard with me on my project and I owe him so much for his contributions. I am very grateful to him and wish him all the best in his goal to obtain a Ph.D. in biochemistry. I want to give a special thank you to Dr. Cathy Radebaugh. She was my hands on

mentor when I first started in the Stargell lab and I cannot put into words how much I have learned about science from working with her. She has always looked out for me and I appreciate it more than she knows. The final member of the Stargell lab that I need to thank is Laurie Stargell. I could not have asked for a better teacher and mentor. Always supportive and slow to anger, Laurie has a lot of personal investment in her students and takes the time to get to know each one individually so that she can interact and manage each one in a way that best suits his/her personality. She has always supported and believed in me over the years. During the times when my project was in a rut and my confidence was at an all time low, she helped me work through it and I am very grateful that she never gave up on me.

Outside of the Stargell lab I would also like to thank Dr. Steve McBryant for his help with the AUC experiments I performed and for his guidance in general. Steve is an amazing teacher that truly enjoys helping others and I cannot thank him enough for his contributions to my project. He helped me analyze my AUC data even after leaving CSU and I am very grateful for all of this help. Dr. Mark van der Woerd is another individual that has contributed to my project, particularly with the SAXS experiments described in Appendix 2. He spent a considerable amount of time helping me analyze my data and I thank him very much for this. I would also like to thank Dr. Robert Woody for his help with the CD experiments that I performed. He has been a great resource for both the design and interpretation of my data and for scientific discussion in general. I would also like to thank Dr. Uma Muthajaran for her help with the nucleosome binding experiments performed here and graduate student Jerome Lee for the RNA binding experiments that were also described.

Last, but definitely not least, I must thank my family for their constant love and support. I have always been very close to my family and this has provided great motivation for me to succeed and make them proud. I have sacrificed so much to earn this degree and I am grateful for their understanding over the last six years. My parents have always taught me to put 100% effort

into anything I do. They are the reason that I have been capable of accomplishing such great things in my life. I love them very much and I dedicate the following thesis to them.

## TABLE OF CONTENTS:

Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	vii

### Chapter 1: Introduction

1.1 Prevalence and importance of RNA Polymerase II dependent transcription.....	1
1.2 Conserved mechanisms of RNA Polymerase II regulation.....	2
1.3 <i>Suppresses post-recruitment functions gene number 1</i> : transcription factor with postrecruitment related functions.....	6
1.4 Questions addressed and contributions made to the understanding of the Spn1/Pol II functional relationship.....	7

### Chapter 2: Materials and Methods

2.1 Sub-cloning/plasmids and yeast strains.....	12
2.2 Yeast media.....	12

2.3	Western blot analysis.....	12
2.4	Genetic analysis.....	16
2.5	Expression and purification of the full length and the central domain of Spn1 .....	17
2.6	Circular dichroism spectroscopy.....	18
2.7	Analytical ultra centrifugation: sedimentation velocity and equilibrium.....	18
2.8	Transcript analysis via S1 nuclease assay .....	19
2.9	Chromatin immunoprecipitation analysis.....	20
2.10	Chronological aging assay.....	21
2.11	Nucleosome binding experiments and gel shifts.....	22

**Chapter 3: Structural and functional characterization of the N-and C-terminal regions  
of the Spn1 protein**

3.1	Abstract.....	123
3.2	Results.....	24
3.2a	The N-and C-terminal regions of Spn1 are required outside optimum growth conditions.....	24

3.2b	The N-and C-terminal regions of Spn1 have partially redundant functions in the cell.....	25
3.2c	The N-and C-terminal regions of Spn1 modestly affect the activation of the <i>CYCI</i> gene.....	29
3.2d	Wild type Spn1 can bind to nucleosomes in vitro.....	35
3.2e	The N-and C-terminal regions of Spn1 are mostly intrinsically disordered in vitro.....	36
3.2f	Contributions of the N-and C-terminal regions on the hydrodynamic properties of Spn1.....	40
3.3	Discussion.....	45
3.3a	N-and C-terminal truncated Spn1, the new breed of Spn1 mutation.....	45
3.3b	Spn1, a node of RNA Pol II dependent functions.....	50
3.3c	Future directions.....	53
<b>Chapter 4: Spn1, a highly conserved protein with a connection to aging in</b>		
	<i>Saccharyomyces cerevisiae</i> .....	57
4.1	Abstract.....	57

4.2	Introduction.....	57
4.3	Results.....	63
4.3a	Removal of the N-and C-terminal regions of Spn1 increases the lifespan of the BY4741 strain of yeast.....	63
4.3b	Spn1 lifespan extension is not due to a change in mitochondrial activity or function.....	66
4.3c	The extension of lifespan seen for the 141-305 <i>spn1</i> allele is likely recessive.....	68
4.3d	Life-extension phenotype is not from perturbing the same pathways as caloric Restriction.....	69
4.3e	Increase in lifespan not seen in the SK1 genetic background.....	71
4.4	Discussion.....	71
4.4a	Future directions.....	74
<b>Appendix 1: Expression, purification, and characterization of wild type Spn1 and other <i>spn1</i> mutants.....</b>		<b>76</b>
A1.1	Over-expression and solubility of wild type Spn1 and various point mutants....	76

A1.2	Optimization of the wild type Spn1 purification scheme.....	79
A1.3	Purification of the K192N spn1 mutant.....	85
A1.4	Purification and in vivo characterization of the 141-305 + K192N double spn1 Mutant.....	90
A1.5	The K192A spn1 point mutant is temperature-sensitive in vivo in the BY4741 genetic background.....	96
<b>Appendix 2: The full length and central domain of Spn1 constructs are not good.....</b>		<b>99</b>
<b>Appendix 3: Spn1 can bind to single stranded RNA in vitro and has a preference for mRNA.....</b>		<b>110</b>
<b>References.....</b>		<b>114</b>

## **Chapter 1: Introduction**

### **1.1 Prevalence and importance of RNA Polymerase II transcription**

At the molecular level, proteins perform the majority of the functions that are required for cell survival and proliferation. DNA-dependent RNA Polymerase II (Pol II), is the central enzyme responsible for transcribing all of the protein-encoding genes within the eukaryotic genome. A high level of molecular regulation is required to spatiotemporally coordinate the activity of Pol II and allow for adaptation to many different environmental conditions and stresses. Cellular differentiation in multi-cellular organisms also requires the expression of specific combinations of proteins in a tightly defined temporal manner. To accomplish this regulation, the cell utilizes a number of different transcriptional programs and strategies that alter the transcriptome of the cell both qualitatively and quantitatively<sup>8</sup>. After an mRNA transcript is produced, it is extensively processed and then exported into the cytoplasm before being translated into a protein by the ribosome. This processing and the physical separation of the events of transcription and translation by the nucleus provide organisms with multiple steps for the regulation of overall protein expression<sup>9; 10</sup>. However, research has shown that the initial production of mRNA is a crucial step in the overall regulation of gene expression and is thereby the main driving force of the research done here<sup>11; 12</sup>. Modulation of the biochemical activity of Pol II is accomplished through the use of a number of transcription factors within the cell that each contribute to its functions at one or more steps throughout the transcriptional process. Mutations to any of the components of the transcriptional machinery can disrupt this regulation and result in the over/underproduction of various proteins at the wrong times and give rise to several disease states such as cancer, Parkinson's disease, or Alzheimer's disease<sup>13; 14; 15; 16</sup>.

Developing a clear mechanistic understanding of how all of the transcriptional factors function together with Pol II is essential for the successful development of drugs and other treatments that combat these diseases. Due to the drastic consequences of the mis-regulation of transcription, it is a vast and large area of research.

## 1.2 Conserved mechanisms of RNA Polymerase II regulation

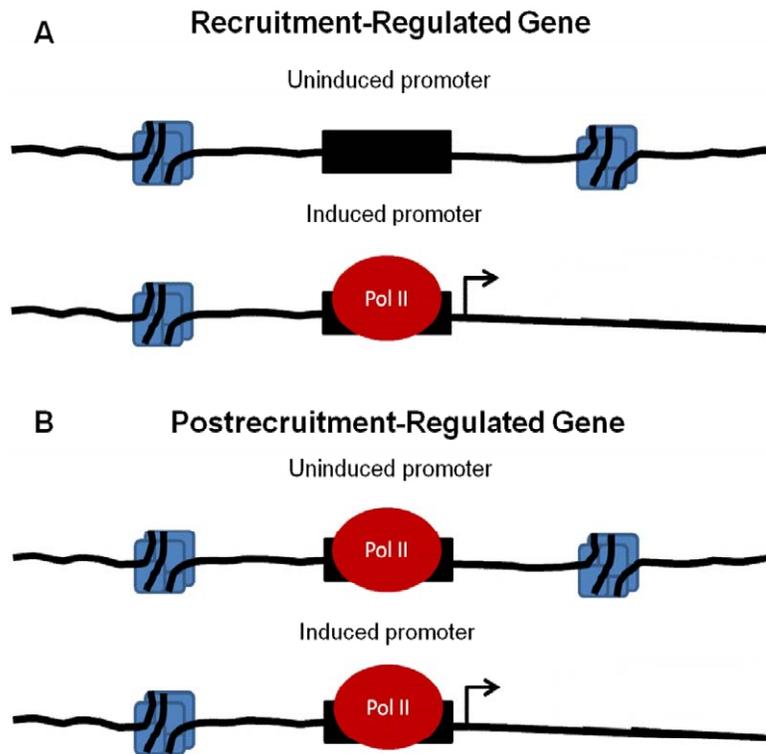
Despite subtle and dramatic variations in the types of Pol II-dependent transcriptional mechanisms reported for different organisms, billions of years of natural selection have given rise to a common set of protein orthologs and mechanistic principles that govern the regulation of protein encoding genes (reviewed by Hahn and colleagues)<sup>17</sup>. This is what makes research on transcription in a lower eukaryote such as yeast applicable to humans<sup>18</sup>. *Saccharyomyces cerevisiae* has been an invaluable model organism for studying cellular processes that include gene expression, chromatin regulation, meiosis, and even the cell cycle. With its short generation time of 90-120 minutes, ease of genetic manipulation, and well-annotated genome, *Saccharyomyces cerevisiae* provides a great starting place for studying various aspects of the process of transcription.

Although thermodynamically fluid, the process of transcription in the nucleus of a eukaryotic cell can be most easily understood and discussed when described as separate stages of Pol II activity that include: recruitment, initiation, promoter escape, elongation, and termination (reviewed by Hampsey and colleagues)<sup>19:20</sup>. Pol II can be regulated at any of these stages to increase or decrease the output of specific mRNA transcripts at a given time. Considering the large amount of DNA within the genome of eukaryotic organisms (~13 million base pairs in yeast and ~3.5 billion base pairs in humans)<sup>21</sup>, Pol II specificity for any given DNA sequence must be kept low to prevent physically tying up all of the available Pol II molecules at any given time. Therefore, rather than the specificity for promoter recognition lying within the structure of this protein, as is the case for bacterial polymerases, the specificity for recruitment and initiation at a

specific gene relies on a combination of other proteins called transcription factors<sup>17; 22; 23</sup>. As such, recruitment of Pol II to promoters typically begins with TFIID/TBP binding to a core promoter DNA sequence. Once bound to the DNA, TBP acts as a platform to nucleate several other general transcription factors that include: TFIIA, TFIIB, TFIIE, and TFIIH, all of which aid in the subsequent recruitment and stabilization of Pol II<sup>19; 24; 25; 26</sup>. In concert with one or more co-activators (TFIID, SAGA, Mediator) TFIIE and TFIIH will stabilize and catalyze site-specific DNA melting, allowing Pol II to initiate the biochemical production of RNA<sup>27; 28</sup>.

Before Pol II escapes from the promoter to form a productive elongation complex, a transition takes place where members of the pre-initiation complex dissociate and various elongation and mRNA processing factors bind to Pol II to aid in the subsequent steps of transcription. Some of these factors include Spt4 and 5 (which function in a complex called DSIF in humans), Spt6, TFIIIS, the mRNA-capping enzyme Ceg1, and various Pol II CTD kinases (reviewed by Saunders and colleagues)<sup>29</sup>. These proteins increase the processivity of the elongating Pol II and couple its catalytic activity with mRNA processing and export (reviewed by Sobennikova and colleagues)<sup>30; 31</sup>. An additional layer of transcriptional regulation exists in the cell that is accomplished by the compaction of DNA into higher order structures referred to as chromatin. The first level of chromatin compaction is accomplished by a complex of 8 histone proteins (2 copies of H2A, H2B, H3, and H4) that fold 147 base pairs of DNA into a nucleosome<sup>32</sup>. This folding of the DNA effectively secludes it from the transcriptional machinery and is used as a means to sterically inhibit the transcription of certain gene sequences<sup>33; 34; 35</sup>. Chromatin remodeling proteins and histone chaperones are therefore also recruited to Pol II-dependent genes in order to expose promoter sequences and open reading frames (ORFs) to the transcriptional machinery. Examples of proteins that perform these functions are Spt6 and the Swi/Snf chromatin remodeling family<sup>29; 36; 37</sup>.

Historically, the rate-limiting step for Pol II transcription was considered to be the initial recruitment of TBP to the promoter. The evidence supporting this view is that an increase in TBP occupancy strongly correlates with increases in the detection of RNA transcript at a significant number of genes<sup>38</sup>. However, a new mechanism has emerged in the last several years that utilizes a rate-limiting step postrecruitment of TBP and Pol II. In this scenario, Pol II occupies the promoter under all conditions even though transcript for that gene is not actively produced and is therefore considered to be “poised” (Figures 1.1a and 1b). Although the rate-limiting step for this type of transcriptional mechanism has remained elusive, the prevalence of this mechanism across the evolutionary spectrum is evident. Beginning with the *Drosophila* heat shock genes, evidence of postrecruitment-regulation has emerged from bacteria all the way up to humans with a particular emphasis at developmental and stress-induced genes (reviewed by Margaritis and Holstege)<sup>20</sup> (reviewed by Price)<sup>39</sup>. In budding yeast, several hundred growth-response genes were found to have a “poised” Pol II molecule at the promoter during stationary phase<sup>40</sup>. In higher eukaryotes, human stem cells also have a large majority of its promoters occupied by Pol II despite the lack of transcript being produced from these genes<sup>41</sup>. Despite the prevalence of this mechanism of Pol II-regulation, many questions remain regarding the molecular architecture of these promoters that keeps Pol II inactive, the molecular trigger for activation, and the consequences of the mis-regulation of these genes in the cell. Permanganate footprinting in *Drosophila* has shown that poised polymerases are initiated and stalled after 25-45 nucleotides have been synthesized suggesting that poised promoters are a form of elongation regulation<sup>42; 43; 44</sup>. In support of this theory, DSIF and NELF have been characterized for stalling Pol II after transcript initiation<sup>45; 46</sup>. In our attempt to better understand postrecruitment-regulation, we previously identified a transcription factor, Spn1, that is capable of suppressing a postrecruitment-defective TBP allele. *Suppresses Postrecruitment functions gene Number One*, has since been characterized as an essential and highly conserved protein from yeast all the way up to humans<sup>47</sup>.



**Fig. 1.1 Key differences between a recruitment regulated promoter and a post-recruitment regulated promoter. A)** Recruitment regulation: In the uninduced state, Pol II does not occupy the promoter and transcript cannot be detected. After induction, Pol II occupancy increases and significant levels of transcript can be detected. **B)** Post-recruitment regulation: In the uninduced state, Pol II occupies the promoter, but transcript cannot be detected. After induction, Pol II occupancy does not increase significantly although transcript levels do increase significantly.

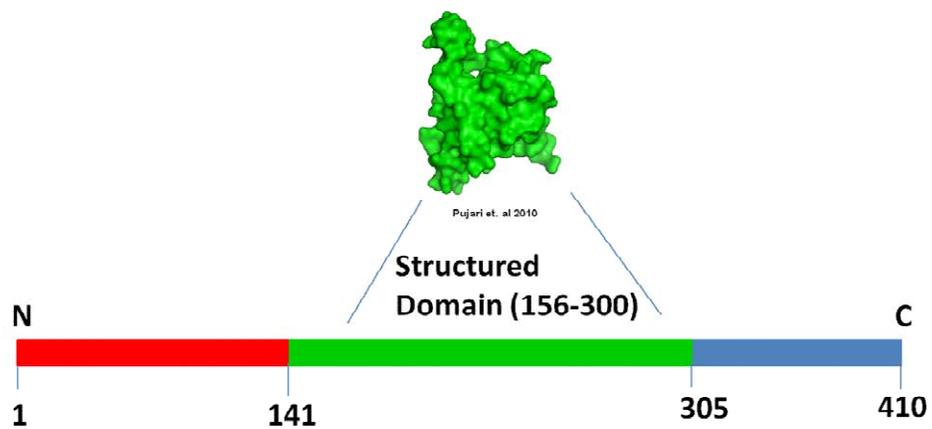
### 1.3 *Suppresses postrecruitment functions gene number 1: transcription factor with postrecruitment-related functions*

Previous research suggests that Spn1 plays a role at poised genes in the cell by affecting the postrecruitment-regulated *CYCI* gene<sup>47; 48</sup>. *CYCI* encodes the iso-1-cytochrome c electron transport protein that is expressed in aerobic growth conditions and in the presence of non-fermentable carbon sources, such as ethanol, that require mitochondrial respiration to be catabolized<sup>49; 50; 51; 52; 53</sup>. Due to the amenability of yeast for the study of transcription, *CYCI* has been one of the models used to study the molecular phenomenon of Pol II poising<sup>25; 48; 54</sup>. Although transcript is detected at only a basal level under partially repressive conditions (the presence of glucose), several members of the pre-initiation-complex (PIC) constitutively occupy the promoter that include: TBP, serine 5 phosphorylated Pol II, the SAGA complex, TFIIF, the Ceg1 mRNA processing factor, and Spn1<sup>48; 54; 55</sup>. Despite the high number of PIC factors present at the promoter, an unknown rate-limiting step/s must be overcome to activate the *CYCI* gene. Spn1 functions as a “poising” factor at least in part by directly interacting with Pol II and inhibiting the recruitment of the Swi/Snf chromatin remodeling complex. When cells are grown in ethanol, Spt6 is recruited almost immediately by interacting with the central domain of Spn1 and the subsequent increase in Swi/Snf occupancy correlates with an increase in transcript production<sup>48; 56; 57</sup>. When a lysine at position 192 out of 410 in Spn1 is changed to an asparagine (K192N), the interaction between Spn1 and Pol II is diminished and its occupancy at the promoter along with Spt6 is reduced in both repressive and activating conditions. Swi/Snf is also constitutively recruited to *CYCI* in the K192N background. Other molecular consequences of this spn1 mutant are faster activation kinetics and higher overall activation of the *CYCI* gene<sup>48; 58</sup>. This data has led to the model that Spt6 alleviates the repression of Swi/Snf when the *CYCI* gene product is required.

## 1.4 Questions addressed and contributions made to the understanding of the Spn1/Pol II functional relationship

Most of the data obtained for Spn1 function to date has been from the analysis of various point mutations made at different locations throughout the primary amino acid sequence or from siRNA knockdown of the protein<sup>47; 48; 56; 57; 58; 59; 60</sup>. The recent publication of the crystal structure of the conserved region of Spn1 (Figure 1.2) revealed that the tertiary orientation of these point mutations are all within a central globular domain<sup>58</sup>. Phenotypic data indicates that the central domain (amino acids 141-305) can robustly cover all of the essential functions of this protein under standard growth conditions and at elevated temperatures<sup>47; 58</sup>. This is all suggestive that the N- and C-terminal regions flanking the central domain of Spn1 are dispensable and do not perform any physiologically relevant functions within the cell. From an evolutionary standpoint, however, this seems unlikely. Low selective pressure from a lack of function should result in the removal of these regions of Spn1 over 2 billion years of evolution. Sequence analysis of higher eukaryotic Spn1 orthologs reveals that this protein does not get smaller as you move up the evolutionary ladder (Figure 1.3). A plausible explanation for this is that the N- and C-terminal regions of the Spn1 protein are required when cells are stressed or perturbed in some way. Typical growth conditions in the lab include an abundance of nutrients and constant temperature conditions. The growth conditions in nature are more stringent and the functions of the outer regions of Spn1 could be required outside optimal conditions.

To continue the characterization of Spn1, the work described in Ch. 3 includes a genetic analysis performed on various mutant strains that have one or both of the N- and C-terminal regions removed from the Spn1 protein (Figure 1.4). Despite the dramatic phenotypes that resulted, removing the N- and C-terminal regions of the protein has a more modest affect on the



**Fig. 1.2 Published crystal structure of the central domain of Spn1** Isolation and crystallization of the conserved domain of Spn1 revealed that it has a highly compact structure comprised entirely of alpha helices and unordered coil secondary structure. Only amino acids 156-300 had sufficiently reliable electron density to be mapped in the structure. The structure of the outer N-and C-terminal regions flanking this domain has not been determined.

```

S.cere. 121 PSSRQELEE--KLDRILKKPKV-----RRTRDED-DLEQYLDEKILRLKDEMNTAAQI
S. Pombe 153 LAAKKELDL--QMDAVLK-PTR-----T#KRSNED-NLEQMADEVLRLREQMRLAALR
Arabidopsis 200 AEEGEDED--EVNNLFKMGKKK-----KTERNPA-----EIALLVENVMAELEVTAAE
C. Elegans 200 DRHGRHFEW--DFDKMLAEKKAERKKKTRRGGKDGGIDI INDDDGTVSRLVERMKHAAKS
Drosophila 520 SNEPENSNFISDFDAMLMRRKKEEKRV--RRRKRDID--LINDNDDLIDQLIVSMKNASDD
Mouse 464 I KRGHMDFLSDFEMLQRKKSMCGK--RRNRRDGG--TFISDADDVSAMIVKMNEAAEE
human 517 I KRGHMDFLSDFEMLQRKKSMSGK--RRNRRDGG--TFISDADDVSAMIVKMNEAAEE

S.cere. 172 DIDTLNKRIETGDTSLIAMQVKLLPKVSVLSKANLADTILDNNLQSVRINWLEPLPDG
S. Pombe 204 DA-----ELNSEQLPATEKLMLPLDVAVLRKTHLYDTILDNNVLSVRMLEPLPDR
Arabidopsis 247 DA-----ELNRQQKPAINKLKLSLTDVLGKQLQTEFLDHGVLTLKNWLEPLPDG
C. Elegans 258 DR-----NANIERKPAFQKIKMLEVKAIMLRAGIVEVLIENGFMSALSEWLAPLPDK
Drosophila 576 DR-----QLNMIGQPATKKISMLQVMSQLIKKHLQLAFLEHNILNVLTDWLAPLPNK
Mouse 517 DR-----QLNNQKPALKLTLLPTVVMHLKQDLKETFIDSGVMSAIKEWLSPLPDR
human 570 DR-----QLNNQKPALKLTLLPTVVMHLKQDLKETFIDSGVMSAIKEWLSPLPDR

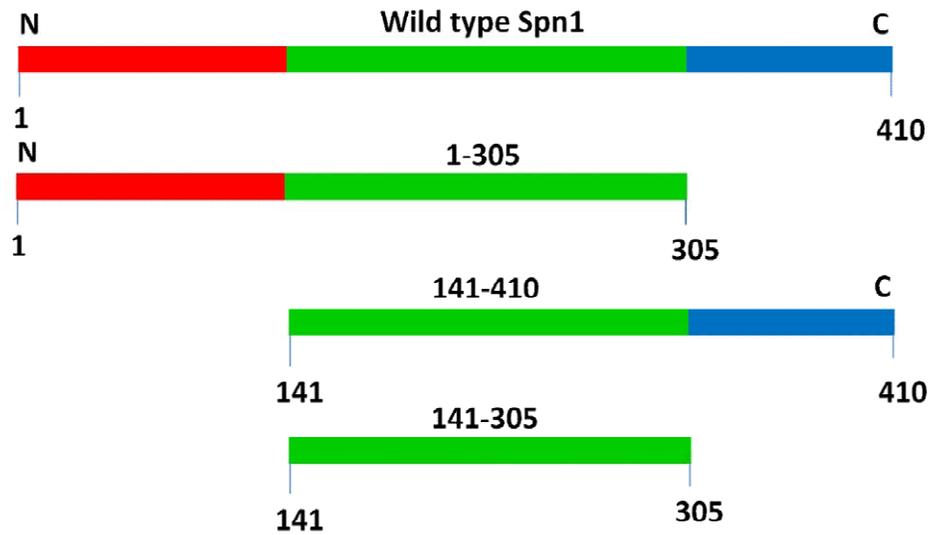
S.cere. 232 SLPSFEIOKSLFAALND--LP-VK-----TEHLKESGLGRVVIFYTKSKRVEAQLARLA
S. Pombe 257 SLPALNIQRSLMDILTK--LP-IQ-----TEHLRESKIGRIVLFYTISKKEPFFIKRIA
Arabidopsis 300 SLPNINIRAAILRVLTD--FP-IDLDQYDRREQLKSGLGKVIMLSKSDEETNSNRRLA
C. Elegans 311 CLPALDIRITVLKLLHNPRFWKLD-----RSTLKQSGLGAVMMLYKHPNETKENKGIA
Drosophila 629 SLPCLQIRESILKLLSD--FPTIE-----KGLLKQSGICKAVMYLYKHPQETKSNRDRA
Mouse 570 SLPALKIREELLKILQE--LPSVS-----QETLKHSGIGRAVMYLYKHPKESRSNKDMA
human 624 SLPALKIREELLKILQE--LPSVS-----QETLKHSGIGRAVMYLYKHPKESRSNKDMA

S.cere. 283 EKLIAEWTRPIGASDNYRDKRIMQLEFDSEKLRKSVMDSAKNRKKS--KSGEDPTS
S. Pombe 308 DNLVSEWSRPIKRSANYRDRAVGVASFNPE-----VFQTRRRRDLAA---AESND---
Arabidopsis 357 KDLVDKWSRPIFNKSTRFE-----DMRNLDEDRVPVRRPPVKKPSNK---ATMESRD-
C. Elegans 365 NLIGEWARPIVHLDTDYS-----TVSRQEREERDYSRMPEKKKINS--RDEEEPND
Drosophila 681 GRLISEWARPIFNVSCHFS-----AMSKERQERDLAQMSRHRHKSPDTEPSSSSKA-
Mouse 622 GKLINEWSRPIFGLTSNYK-----GMTREREQRDLEQMPQRRLSS---TGGQTPR-
human 676 GKLINEWSRPIFGLTSNYK-----GMTREREQRDLEQMPQRRMNS---TGGQTPR-

```

Spn1 ortholog	amino acid #
<i>S. cerevisiae</i>	410
<i>S. pombe</i>	428
<i>Arabidopsis</i>	502
<i>C. Elegans</i>	511
<i>Drosophila</i>	820
Mouse	766
Human	819

**Fig. 1.3 The N-and C-terminal regions flanking the central domain of Spn1 do not get smaller throughout evolution** Sequence alignment of the highly conserved central domain of various Spn1 orthologs and the corresponding total number of amino acids for each protein to show that natural selection has not deleted the additional amino acids flanking this domain. Alignments were performed using the Kalign multiple sequence alignment program. (<http://www.ebi.ac.uk/Tools/msa/kalign/>)



**Fig. 1.4 Spn1 protein derivatives of interest** Visualization of the main Spn1 proteins used for the research done here and the nomenclature that will be used to describe them. Spn1 is 410 amino acids and will be referred to as wild type when the full length protein is used. Amino acids 141-305 make up the central domain and amino acids 1-140 and 306-410 will be referred to as the N- and C-terminal regions of Spn1 respectively.

activation of the *CYCI* gene. Rather than playing a specific role in postrecruitment-regulated transcription, this data suggests that the N-and C-terminal regions of Spn1 have more general functions in the transcriptional process beyond regulation of *CYCI*. In support of this idea, these regions of the protein are able to bind to reconstituted mono-nucleosomes. The wild type Spn1 protein was biophysically characterized to better understand these interactions and the overall functions of Spn1 in general. Rather than folding into highly ordered domains similar to that of the central domain of Spn1, a high level of intrinsic disorder was detected in the N-and C-terminal regions of the protein. This data suggests that these regions of Spn1 are able to bind to a number of different proteins within the cell since they are not limited to a specific structure. The work described in Ch. 4 examines the Spn1 protein's role in the process of cellular aging. Removal of the N-and C-terminal regions of Spn1 dramatically increase the lifespan of *Saccharomyces cerevisiae*. The complexity of the process of aging supports the idea that Spn1 binds to numerous proteins in the cell and regulates multiple essential pathways required for viability. Additional research on the Spn1 protein can be found in three additional appendices. Appendix 1 is an in-depth description of the purification of the wild type Spn1 protein and various point mutants. Appendix 2 describes the small-angle X-ray scattering (SAXS) experiments that were attempted during the biophysical characterization of Spn1. Finally the work in Appendix 3 is preliminary data on the RNA-binding properties of Spn1. The work presented here strongly suggests a model where Spn1 is comprised of a highly globular central domain flanked by two intrinsically disordered regions and that the entire protein is required for viability and cellular proliferation.

## Chapter 2: Materials and Methods

### 2.1 Sub-cloning/Plasmids and Yeast Strains

See table 2.1 for a list of all of the plasmids used for the studies done here. For the plasmids not described previously the cloning was performed as follows. The wild type Spn1 promoter, terminator, and ORFs encoding the following Spn1 constructs were all isolated from the pJF244, pJF217, and pJF242 plasmids<sup>47</sup> and subcloned into the pRS313 (*HIS3*, CEN) backbone: 141-305 (central domain), 1-305 (N terminus + central domain), and 141-410 (central domain + C terminus) respectively. This was accomplished by performing a *SacI* and *HindIII* double-restriction enzyme digest. The *HindIII* 5' overhang was turned into a blunt end by filling it in using the NEB T4 DNA polymerase. *SacI* and *SmaI* restriction enzymes were used to cut the pRS313 plasmid. Following agarose gel purification, the Spn1 gene sequences were subsequently ligated into the plasmid backbone using NEB's T4 DNA ligase. For the in vivo experiments studying Spn1, these plasmids were transformed into the following strains using a method developed by Zhang and colleagues<sup>47; 48</sup> (see table 2.2 for all strains used in this study).

### 2.2 Cell Culturing Conditions

Media used for routine culture of yeast was prepared as described<sup>61</sup>. For ethanol induction, yeast cultures were grown overnight in YPD, then diluted and allowed to undergo two doublings in YPD. Cells were washed with YP three times and diluted into YP supplemented with 3% ethanol as the sole carbon source and were cultured at 30° with shaking for various times (30 minutes to 5 hours, as indicated). The synthetic dropout (SD) dropout medium used for the chronological aging assays was prepared as described<sup>62</sup>.

### 2.3 Western Blot Analysis

Yeast cells (10-20mL) were grown to an OD600 of ~0.8-1.0 in YPD medium at 30°C before cells were harvested, washed with double-distilled H<sub>2</sub>O, and resuspended in 100-200 µl of lysis buffer (25mM Tris Phosphate pH 6.7, 2mM PMSF). Whole-cell extracts were prepared by

**Table 2.1**

<b>Plasmid</b>	<b>Description</b>	<b>Source</b>
pJF211	Full length wild type Spn1, myc2-tagged, YCp22-based ( <i>TRP1</i> , CEN)	<sup>47</sup>
pJF212	Full length K192N Spn1, myc2-tagged, YCp22-based ( <i>TRP1</i> , CEN)	<sup>47</sup>
pJF217	N-terminus + central domain of Spn1 (1-305), myc2-tagged, YCp22-based ( <i>TRP1</i> , CEN)	<sup>47</sup>
pJF242	central domain + C-terminus of Spn1 (141-410), myc2-tagged, YCp22-based ( <i>TRP1</i> , CEN)	<sup>47</sup>
pJF244	central domain of Spn1 (141-305), myc2-tagged, YCp22-based ( <i>TRP1</i> , CEN)	<sup>47</sup>
pCR311	Full length wild type Spn1, myc2-tagged, pRS313-based ( <i>HIS3</i> , CEN)	This study
pCR611	Full length wild type Spn1, myc2-tagged, pRS313-based ( <i>URA3</i> , CEN)	This study
pCR312	Full length K192N Spn1, myc2-tagged, pRS313-based ( <i>HIS3</i> , CEN)	This study
pAA317	N-terminus + central domain of Spn1 (1-305), myc2-tagged, pRS313-based ( <i>HIS3</i> , CEN)	This study
pAA342	central domain + C-terminus of Spn1 (141-410), myc2-tagged, pRS313-based ( <i>HIS3</i> , CEN)	This study
pAA344	central domain of Spn1 (141-305), myc2-tagged, YCp22-based ( <i>HIS3</i> , CEN)	<sup>58</sup>
WT pET15b	Full length wild type Spn1, N-terminal His tag, ampicillin and chloramphenicol resistance	This Study
141-305 pET15b	central domain of Spn1 (141-305), N-terminal His tag, ampicillin and chloramphenicol resistance	<sup>58</sup>

**Table 2.2**

Strain	Description	Source
JF11	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + YCp50-SPN1 (URA3)</i>	<sup>47</sup>
JF15	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + pJF211</i>	<sup>47</sup>
JF16	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + pJF212</i>	<sup>47</sup>
AA1	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + pJF217</i>	This study
AA2	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + pJF242</i>	This study
AA3	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leuΔ2::PET56, spn1::LEU2 + pJF244</i>	This study
CR1	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2 + pCR311 (HIS3)</i>	This study
CR2	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2 + pCR312 (HIS3)</i>	This study
AA4	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2 + pAA317 (HIS3)</i>	This study
AA5	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2 + pAA342 (HIS3)</i>	This study
AA6	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2 + pAA344 (HIS3)</i>	This study
CR3	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, snf5Δ::Kan<sup>r</sup>, spn1::leu2 + pCR311</i>	This study
CR4	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, snf5Δ::Kan<sup>r</sup>, spn1::leu2 + pCR312</i>	This study
AA7	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, snf5Δ::Kan<sup>r</sup>, spn1::leu2 + pCR344</i>	This study
CR5	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, dst1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR311</i>	This study
CR6	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, dst1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR312</i>	This study
AA8	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, dst1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR317</i>	This study
AA9	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, dst1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR342</i>	This study
AA10	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, dst1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR344</i>	This study
CR7	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, rtf1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR311</i>	This study
CR8	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, rtf1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR312</i>	This study
AA11	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, rtf1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR317</i>	This study
AA12	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, rtf1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR342</i>	This study
AA13	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, rtf1Δ::Kan<sup>r</sup>, spn1::leu2 + pCR344</i>	This study

AA14	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2</i> + pCR311 ( <i>HIS3</i> ) + pCR611 ( <i>URA3</i> )	This study
AA15	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2</i> + pCR344 ( <i>HIS3</i> ) + pCR611 ( <i>URA3</i> )	This study
AA16	<i>MATa, his3Δ1, ura3Δ0, leu2Δ0, met15Δ0, spn1::leu2</i> + pCR344 ( <i>HIS3</i> ) + empty pRS316 ( <i>URA3</i> )	This study
AA17	<i>MATa, ura3-52, trp1Δ63, his3Δ200, leu2::PET56, spn1::LEU2</i> + YCp50- <i>SPN1</i> ( <i>URA3</i> ) + empty pRS313 ( <i>HIS3</i> )	This study

vigorous bead beating. Cellular debris was removed by spinning the extracts at 3,000 rpm at 4°C for 15 minutes. Protein concentrations were determined via Bradford assay (Sigma). Equal amounts of whole cell extracts (15-25µg) were separated on 10-15% SDS-PAGE gels and then transferred to a nitrocellulose membrane at 80 volts for 1 hour. The following antibodies were used at the given dilutions: anti-myc (Upstate Inc., 1:500), and polyclonal anti Spn1 or anti TBP (1:10,000). Horseradish peroxidase (HRP)-conjugated secondary antibodies were used at a 1:20,000 dilution and protein bands detected using ECL Plus reagents from Amersham Biosciences. Alternatively, secondary anti-mouse antibody conjugated to IRDye 680 or secondary anti-rabbit antibodies conjugated to IRDye 800CW (both from LI-COR Biosciences) were used and the blots were subsequently analyzed using the Odyssey scanning system. Emission was detected at 700nm and 800nm respectively. Odyssey imaging software was used to visualize the blots.

## 2.4 Genetic Analyses

To assay the genetic interactions of *SPN1*, the strains were grown under 17 different conditions. Phenotypic changes were scored by comparing the growth of strains covered by the K192N, 141-305, 1-305, and 141-410 mutant *spn1* alleles with wild type *SPN1* under the following conditions: 30°C, 38°C, and 16°C, 1 M NaCl, 1 M sorbitol, 2-4 mM H<sub>2</sub>O<sub>2</sub>, -Ino media, glucose versus galactose, raffinose, glycerol, or ethanol/glycerol carbon sources, 50 mM aminotriazole (AT), 10 mM caffeine, and 40 µg/ml MPA. Yeast media used to analyze phenotypic changes were made as described in the literature<sup>61</sup>. Five-fluoroorotic acid plates were made as described previously<sup>63</sup>. YPGal and YPEG plates were made by replacing the dextrose in YPD with 2% galactose, 3% ethanol, and 3% glycerol. Plates containing caffeine, sorbitol, NaCl, and H<sub>2</sub>O<sub>2</sub> were made by supplementing YPD medium with 10 mM caffeine (Sigma), 1 M sorbitol, 1 M NaCl, and 2-4 mM H<sub>2</sub>O<sub>2</sub>. Medium lacking inositol (-Ino) was made as described<sup>64</sup>.

Mycophenolic acid (MPA) plates were made by supplementing YPD plates with 40 µg/ml of MPA.

## **2.5 Expression and Purification of the Wild Type and 141-305 Spn1 Proteins**

Wild type Spn1 and the central domain (amino acids 141–305) were subcloned into a pET15b vector, which contains an N-terminal histidine tag, and they were each expressed using Rosetta 2 DE3 pLysS cells at 30 °C for two hours and at 16 °C overnight respectively. In brief, 1L of Luria Broth (LB) medium supplemented with ampicillin (100 µg/ml), chloramphenicol (34 µg/ml), and glucose (1%) was inoculated with cells from a 10ml overnight culture. For wild type Spn1, the cells were grown at 37 °C to an OD600 of ~0.4, transferred to 30 °C, and allowed to grow to an OD600 of ~0.6 before the cells were induced with 1mM IPTG. For the 141-305 mutant, the cells were grown at 37 °C to an OD600 of ~0.5 and then transferred to 16 °C for 30 minutes before inducing the cells with 1mM IPTG. After the appropriate induction period (2 hours for wild type, overnight for the 141-305 mutant), the cells were put on ice for 30 minutes and were then collected by centrifugation at 4000 rpm for 15 min and the pellets were stored at -80°C. For each construct a cell pellet from 1 L of medium was re-suspended in 40 ml of lysis buffer [100 mM Tris (pH 7.5), 1M NaCl, 10% glycerol, 50mM imidazole, and 500 µM PMSF]. The cells were lysed by sonication and the cellular debris was pelleted by centrifugation at 15,000 rpm for 40 min. Recombinant wild type Spn1 was purified on a 5mL Hi-Trap chelating column charged with NiSO<sub>4</sub> and was eluted using a linear gradient of imidazole from 50mM to 500mM over 20 column volumes. The appropriate fractions were pooled and concentrated to ~1mL using Millipore concentrators with a cutoff of no more than 30kDa. The concentrated pool was injected onto a Superdex 200 (16/60) size-exclusion chromatography (SEC) column at 0.5ml/min (25mM MES pH6.5, 200mM NaCl, 10% glycerol). A final purification step was performed by loading the sample onto a 5mL Hi-Trap SP cation exchange-column. Spn1 was eluted using a linear gradient of NaCl from 50mM to 600mM over 25 column volumes (25mM MES pH6.5 and 10% glycerol

still used). For the central domain of Spn1, the same purification scheme was used minus the final cation exchange step<sup>58</sup>. The histidine tag was left on both constructs for all experiments performed.

## 2.6 Circular Dichroism Spectroscopy

Circular Dichroism (CD) and Prediction Algorithms were performed on a Jasco-720 spectropolarimeter at 20 °C. Proteins were dialyzed extensively against 10 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 7.4 and 100 mM NaF buffer prior to obtaining measurements. One spectrum was collected for wild type Spn1 and the 141-305 mutant at 25μM and 30μM respectively. Each scan was obtained from averaging measurements taken at 10nm/min. with a response time of 16 seconds from 260nm to 180nm. Each spectrum was baseline subtracted from a similar scan performed with dialysis buffer. The molar ellipticity [θ] was obtained by normalization of the measured ellipticity (θ, millidegree), where  $[\theta] = (\theta \times 100) / (nlc)$ ,  $n$  is the number of residues,  $c$  is the total concentration (mM), and  $l$  is the cell pathlength (cm). The percentages of secondary structure types were determined from the spectra using the CONTINLL, SELCON3, and CDSSTR methods within CDPro analytical software<sup>65, 66</sup>. The SDP42 basis set was used to deconvolute the CD spectrum. The percentages of secondary structure reported are the averages of two independent biological replicates.

## 2.7 Analytical Ultra Centrifugation: Sedimentation Velocity and Equilibrium

Analytical ultracentrifugation sedimentation velocity experiments were performed with either a Beckman XL-I or a Beckman XL-A analytical ultracentrifuge using the absorbance optical system as described<sup>67</sup>. Wild type Spn1 and the 141-305 spn1 mutant proteins were extensively dialyzed against 20mM Tris-HCl (pH 7.5) + 150mM, 300mM, or 500mM NaCl buffer as indicated. Wild type Spn1 was spun at 45,000 rpm with a radial step size of .003cm at 1.2μM and 23.3μM (detected at 229nm and 280 nm respectively). The central domain of Spn1 was spun at 50,000 rpm with a radial step size of .003cm at 2.1μM and 64.8μM (detected at

229nm and 280nm respectively). Boundaries were analyzed using the method of Demeler *et al.*<sup>68; 69</sup> using Ultrascan (version 9.9). This analysis yields an integral distribution of sedimentation coefficients,  $g(s)$ . Sedimentation coefficients ( $s$ ) were corrected to that in water at 20 °C ( $s_{20,w}$ ). The solvent densities ( $\rho$ ) were calculated in Ultrascan. The partial specific volume for wild type Spn1 (.7239 cm<sup>3</sup>/g at 20 °C) and for the 141-305 spn1 mutant (.7418 cm<sup>3</sup>/g at 20 °C) were calculated from the primary amino acid sequence within Ultrascan. Modeling of hydrodynamic parameters (sedimentation coefficient, molecular weight,  $f/f_0$ , and RMSD values) was performed within Ultrascan. Experiments were all performed with two independent preps of protein to ensure the results were reproducible.

Sedimentation equilibrium experiments for wild type Spn1 were performed at 4 °C using charcoal-filled Epon 6-sector centerpieces following extensive dialysis into 20mM Tris-HCl (pH 7.5) + 150mM and 500mM NaCl. Scans were collected with the absorbance optical system at 229 and 280 nm with wild type Spn1 at 1.2 $\mu$ M to 28 $\mu$ M respectively by using an average of 20 scans collected at a radial step resolution of 0.001 cm. Overlays of successive scans taken four hours apart at four different speeds (18700, 26,500, 32,500 and 37,500 rpm) confirmed that the samples had reached equilibrium. The equilibrium concentration gradients were edited and globally fit to a variety of models within Ultrascan including a single-ideal species model and a two-component non-interacting model.

## 2.8 Transcript Analysis Via S1 Nuclease Assays

S1 nuclease assays were conducted as described previously<sup>70</sup>. For *CYCI* induction, cultures were grown to log phase for 4-5 hours (optical density of 0.8 to 1.0) in YP medium containing 2% glucose and were then washed three times in YP medium lacking glucose, diluted into YP medium containing 3% ethanol, and cultured at 30°C for the indicated time points. Uninduced samples were taken prior to the wash step. Total RNA was isolated by hot-phenol extraction and was quantitated using absorbance at 260nm. Hybridizations with excess probe

were done with 30 µg of RNA samples hybridized with excess <sup>32</sup>P-labeled probe overnight at 55°C. S1 nuclease digestion was performed on the hybridized samples for 30 minutes at 37°C. The samples were ethanol precipitated and resolved on 10% polyacrylamide gels. Gels were exposed to phosphor screens and band intensity was normalized to the intensity of the tRNA<sup>w</sup> band using Imagequant software. The wild type 5-hour time point was set to 10 within each biological replicate set and remaining data was normalized and averaged accordingly. The presented data is the average of four independent biological replicates.

## **2.9 Chromatin Immunoprecipitation Analysis**

Chromatin immunoprecipitations (ChIPs) were performed as described by Zhang and colleagues<sup>48</sup> with a few modifications. Cells (150 ml) were grown to an OD<sub>600</sub> of 0.8–1.0 in YP medium containing 2% glucose and were treated with formaldehyde (1%) for 15 min with swirling every 5 min. Glycine (125 mM) was added for 5 min to stop the crosslinking. Cells were collected and washed twice in ice-cold TBS buffer. Cells were then resuspended in FA-lysis buffer [50 mM Hepes, pH7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and a 13-protease inhibitor cocktail of PMSF, benzamidine, pepstatin, leupeptin, and chymostatin]. For ethanol induction, cells were grown to an OD<sub>600</sub> of 0.8-1.0 and were washed 3X with YP medium lacking glucose before inoculating into YP medium containing 3% ethanol for five hours.

Chromatin was sheared by sonication using a BransonW-350 model sonifier (10 times at 10 sec each on continuous pulse at a microtip power setting of 6). Ten percent of the chromatin material used for the immunoprecipitation was processed as the input after reversing the crosslinks and purifying the DNA. Chromatin material (500 µl) was incubated with 10 µl of anti-Myc (Upstate) antibodies or with 10 µl of monoclonal anti-RNA Pol II antibody (8WG16, Covance) rotating overnight at 4°C. Fifty microliters of protein-A sepharose beads were incubated with the chromatin material for 3 hours at room temperature. The beads were collected

by centrifugation, and the antigen–antibody complexes were recovered and treated with elution buffer (50 mM Tris, 10 mM EDTA, 1% SDS) for 15 min at 65°C to disrupt the complexes. Protein–DNA crosslinks were reversed by incubation overnight at 65°C and the DNA was purified by phenol–chloroform extraction and ethanol precipitation. The DNA was subsequently analyzed via quantitative PCR analysis. Quantitative PCR reactions were carried out in a volume of 25 µl using a BioRad iCycler and SYBR fluorescein mix. Standard curves were generated using 10-fold serial dilutions of Input DNA. PCR efficiencies ranged from 90 to 100%, with a correlation coefficient of 0.95 or greater. Threshold cycle data were quantified relative to the input as described <sup>71</sup>. Primers were designed for the promoter region of the *CYCI* gene (-151 to +36) and the ORF of the *CYCI* gene (+181 to +325). Spn1 occupancy at *CYCI* is reported as fold over background. This was calculated by dividing the % input for each mutant by the signal obtained for an untagged strain. Data for Pol II occupancy at *CYCI* is represented as % input. All experiments were performed in biological triplicate.

## **2.10 Chronological Aging Assay**

Starter cultures (10 ml) were inoculated with newly revived colonies and were grown overnight rotating at 30°C in synthetic dropout (SD) medium supplemented with the appropriate amino acids. These cultures were then used to inoculate 100 ml of fresh SD medium in 500-ml flasks to an optical density at 600 nm (OD<sub>600</sub>) of 0.1. The cultures were grown with shaking (220 rpm) at 30°C. After 72 hours, these cultures were considered as being completely in stationary phase, and in all experiments described this time point represents 100% viability. Viability was determined over the course of the experiment using quantitative plating experiments. Quantitative plating experiments were carried out every other day starting with day three by making tenfold serial dilutions up to 10<sup>5</sup> and plating 100µl of the final dilution onto YPD plates. Colony numbers were counted and were divided by day three's counts to calculate the percent viability of the population of each strain over the course of the experiment. All

percentages were averages of 2-5 biological replicates. The dilutions for each strain were also performed in triplicate each day of the assay and the results were then averaged for each time point within each replicate set. The index of respiratory competence (IRC) and erythromycin resistance assays were performed as described<sup>62</sup>.

For the chronological aging experiments testing whether the central domain life-extension allele phenotype is recessive or dominant, the same procedure was followed with a few modifications. Four new strains were created (Table 2.2) for the experiment. The strain encoding the 141-305 *spn1* allele was transformed with the pCR611 plasmid encoding the wild type *SPN1* allele. Three other control strains were also made, one encoding two copies of wild type *SPN1*, one with the 141-305 *spn1* mutant in the pRS313 vector and an empty pRS316 vector, and one with wild type *SPN1* and an empty pRS313 vector. These controls were to ensure that any phenotypes observed were not the result of the plasmids themselves. After transforming the plasmids into these strains, the only difference in the performance of the lifespan assay was that the medium was lacking both histidine and uracil in order to maintain both plasmids simultaneously.

### **2.11 Electrophoretic Mobility Shift Assay: Nucleosome binding experiments with Spn1**

Both 147-base pair and 165-base pair reconstituted nucleosomes were made using recombinant *Xenopus* histones and Widom DNA as described<sup>72; 73</sup>. Purified wild type Spn1 and the 141-305 *spn1* mutant were prepared as described here. The nucleosomes [5 $\mu$ M] and each Spn1 protein [molar ratio to nucleosomes as indicated] were incubated in 10 mM Tris 7.5, 1mM EDTA, and 50 mM KCl buffer at room temperature for 30 minutes prior to loading onto a pre-run 5% native polyacrylamide gel at 150V for 60 minutes. The gels were subsequently visualized using ethidium bromide and coomassie brilliant blue stain. All experiments were repeated with two independent protein preps of Spn1 and nucleosomes to ensure that the results were reproducible.

### **Chapter 3: Structural and functional characterization of the N-and C-terminal regions of the Spn1 protein**

This chapter contains the description of the characterization of the N-and C-terminal regions of the Spn1 protein. The physiological requirements of these regions (comprised of amino acids 1-140 and 306-410 of Spn1) were analyzed using a growth assay under various stress-inducing conditions and in different genetic backgrounds. The functions of these regions at the postrecruitment-regulated *CYCI* gene were also tested. The ability of the Spn1 protein to interact with nucleosomes in vitro will be discussed along with the biophysical characterization of the N-and C-terminal regions of Spn1. I wrote this chapter and contributed most of the figures. Dr. Xu Chen performed the chromatin immunoprecipitation experiments and contributed figures 3.5a and 3.5b. Dr. Uma Muthurajun was responsible for preparing the nucleosomes and for performing Spn1/nucleosome binding experiments. She contributed figures 3.7, 3.8, and 3.9. Dr. Steve McBryant played a significant role in the calculations and interpretations of the AUC data described here.

#### **Abstract**

A multitude of proteins are responsible for regulating the activity of RNA Polymerase II (Pol II) in the nucleus of a eukaryotic cell. Two types of themes are used by these proteins to control transcription: recruitment-regulation and postrecruitment-regulation. The main difference is the rate-limiting step for producing transcript. The rate-limiting step for the first mechanism is the recruitment of Pol II to the promoter. For the second mechanism, Pol II constitutively occupies the promoter, is “poised”, and an unknown rate-limiting postrecruitment step prevents transcription from commencing. The highly conserved and essential transcription factor Spn1 was identified as a protein that functions postrecruitment of Pol II and further experimentation revealed that it plays a direct role at regulating the poised *CYCI* gene in *Saccharomyces*

*cerevisiae*. These functions were identified from the analysis of multiple point mutations within the most conserved portion of Spn1 comprised of a highly folded central domain. Little is known about the functions of the N-and C-terminal regions flanking this domain (amino acids 1-140 and 306-410) and is the focus of the work done here. Genetic experimentation indicates that these regions have physiologically important functions within the cell outside of optimum growth conditions, but only play a minor role in the regulation of the *CYCI* gene. A broader approach of experimentation is required to understand all of the Spn1 protein's functions regarding transcription. This led to the observation that Spn1 is able to bind to nucleosomes in vitro and that this interaction is dependent on the N-and C-terminal regions of the protein. The possibility that Spn1 could affect nucleosome dynamics in the cell is consistent with the physical and genetic interactions observed between Spn1 and the Spt6 and Swi/Snf histone chaperone and chromatin remodeling complexes. This result will provide several new avenues for future Spn1 research.

## **Results**

### **3.2a The N-and C-terminal regions of Spn1 are required outside optimum growth conditions**

A genetic screen was performed by Zhang and colleagues to analyze the in vivo functions of the Spn1 protein by combining a K192N point mutant allele with a variety of genetic backgrounds comprised of single nonessential gene deletions for factors involved in RNA Pol II-mediated transcription<sup>48</sup>. In brief, the *spn1* mutant allele was introduced into each strain after the genomic copy was knocked out using homologous recombination and was tested for growth under fourteen different growth conditions. Three of the factors that the K192N mutant genetically interacted with was: *RTF1*, *DST1* (TFIIS), and the Swi/Snf chromatin remodeling complex<sup>48</sup>. Using this as a framework, two parental strains and three deletion strains (SK1, BY4741,  $\Delta snf5$ ,  $\Delta dst1$ (TFIIS), and  $\Delta rtf1$ ) were transformed with the 141-305 *spn1* mutant allele. The mutant strains were grown under seventeen different growth conditions, all meant to perturb

the transcriptional machinery and were assessed for growth relative to cells expressing wild type Spn1 (see Materials and Methods).

The effects of the removal of the N-and C-terminal regions of Spn1 were very specific, producing six dramatic growth defects (Table 3.1). In the SK1, BY4741, and  $\Delta snf5$  strains, the central domain was capable of robustly covering the essential functions of the Spn1 protein under all of the conditions tested. However, in the  $\Delta dst1$  strain, the 141-305 *spn1* allele was deficient at 15°C and almost lethal at 38°C or when the growth medium was supplemented with caffeine. This mutation was also almost lethal in the *Artf1* strain under all growth conditions (Figure 3.1). Since these phenotypes do not manifest without deleting these factors and mutating Spn1, they are classified as synthetic. Synthetic genetic interactions strongly suggest that the factors of interest are intimately involved in performing functions that are very important within the cell. Therefore the N-and C-terminal regions of the Spn1 protein perform both important and specific physiologically relevant functions alongside the Dst1 and Rtf1 proteins. It is also worth noting that these phenotypes were not the same as observed for the K192N mutation. This provides evidence that these two mutations affect different functions of Spn1 during transcription.

### **3.2b The N-and C-terminal regions of Spn1 have partially redundant functions in the cell**

After observing the physiological necessity for the N-and C-terminal regions of the Spn1 protein in vivo, a curious follow-up question was to determine the functional relationship of the N-and C-terminal regions relative to one another using the same growth assays. The *spn1* mutant alleles encompassing the central domain (141-305) and either the N-or C-terminal regions (amino acids 1-305 and 141-410 respectively) (Figure 1B) were transformed into the BY4741,  $\Delta dst1$ , and *Artf1* strains. These strains were assayed under the conditions from the above screen that produced the most significant phenotypes (Table 3.2). In the  $\Delta dst1$  background, either region is able to suppress the growth deficiencies of the central domain except for growth on caffeine. Under this condition, amino acids 141-410 are required for wild type growth while amino acids

**Table 3.1**

				H <sub>2</sub> O <sub>2</sub>													
Growth				2.5	3.5	4.5	>4.5										
media	15°C	30°C	38°C	mM	mM	mM	mM	YPEG	YPG	YPgal	YPraff	NaCl	caff.	-ino	sorb	AT	MPA

**SK1**

K192N	N	N	Ex	N	N	-	nt	N	N	N	N	N	N	N	N	N	nt	nt
141-305	N	N	mEx	N	N	-	nt	N	N	N	N	N	mEx	N	N	N	nt	nt

**BY4741**

K192N	N	N	Ex	N	N	S	S	N	N	N	N	N	N	N	N	N	N	N
141-305	N	N	N	N	N	N	mS	N	N	N	N	N	mEx	N	N	N	N	N

***Δsnf5***

K192N	N	N	S	S	S	S	-	S	S	S	N	S	N	N	N	mS	N
141-305	N	N	N	N	N	N	-	N	N	mS	N	N	mEx	N	N	mS	N

***Δdst1***

K192N	N	N	Ex	N	N	Ex	S	N	N	N	N	N	N	N	N	N	N	N
141-305	mEx	N	Ex	N	N	N	N	N	N	N	N	N	Ex	N	N	N	N	N

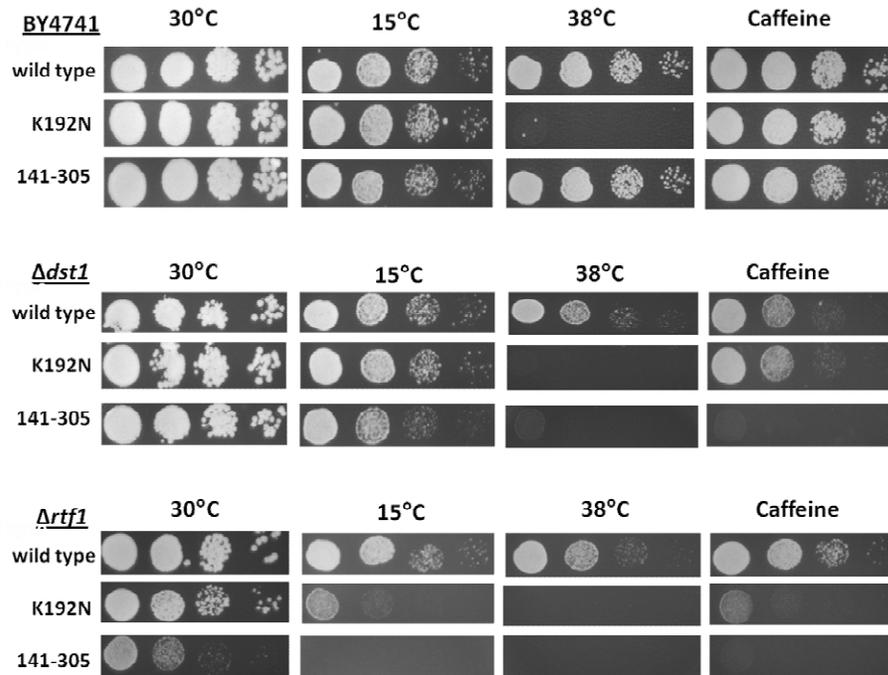
***Δrtf1***

K192N	Ex	Ex	-	Ex	Ex	Ex	-	Ex										
141-305	-	Ex	-	Ex	Ex	-	-	-	-	Ex								

**Legend:**

N = No change in phenotype from that observed in the parental or knockout strain  
mEx = Mild exacerbation of the phenotype observed in the parental or knockout strain  
Ex = Exacerbation of the phenotype observed in the parental or knockout strain  
mS = Mild suppression of the phenotype observed in the parental or knockout strain  
S = Suppression of the phenotype observed in the parental or knockout strain  
- = Strain is dead under indicated growth conditions  
nt = Strain has not been tested under indicated growth conditions

**Results of the genetic analysis** A plasmid construct containing the central domain of Spn1 (141-305) was transformed into the SK1, BY4741, *Δsnf5*, *Δdst1*, and *Δrtf1* genetic backgrounds and these strains were assayed for growth under 17 different conditions that challenge the general transcription machinery. With the exception of a couple of mild growth phenotypes, the central domain of Spn1 is capable of robustly covering the Spn1 protein's functions in the SK1 and BY4741 genetic backgrounds. This screen also revealed that the N- and C-terminal regions of Spn1 only mildly interact with the Snf5 subunit of the Swi/Snf chromatin remodeling complex. The deletion of *Dst1* and *Rtf1*, however, produced multiple drastic slow-growth phenotypes and provide strong evidence that the N- and C-terminal regions of Spn1 have physiologically relevant functions.



**Fig. 3.1 Central domain of Spn1 allele interacts strongly with *DST1* and *RTF1*** Similar amounts of yeast cells were diluted serially and plated onto YPD medium with and without 10mM caffeine. The strains are indicated to the left. Pictures were taken after growing cells for 2-5 days. Figure shows that the mild growth defects of  $\Delta$ *dst1* and  $\Delta$ *rtf1* are greatly exacerbated by removing the N- and C-terminal regions of Spn1.

**Table 3.2**

Growth media	15°C	30°C	38°C	Caffeine
<b>SK1</b>				
1-305	N	N	S	S
141-410	N	N	S	S
<b>BY4741</b>				
1-305	N	N	N	S
141-410	N	N	N	S
<b><i>Δdst1</i></b>				
1-305	S	N	S	mS
141-410	S	N	S	S
<b><i>Δrtf1</i></b>				
1-305	mS	S	mS	S
141-410	S	S	S	S

**Legend:**

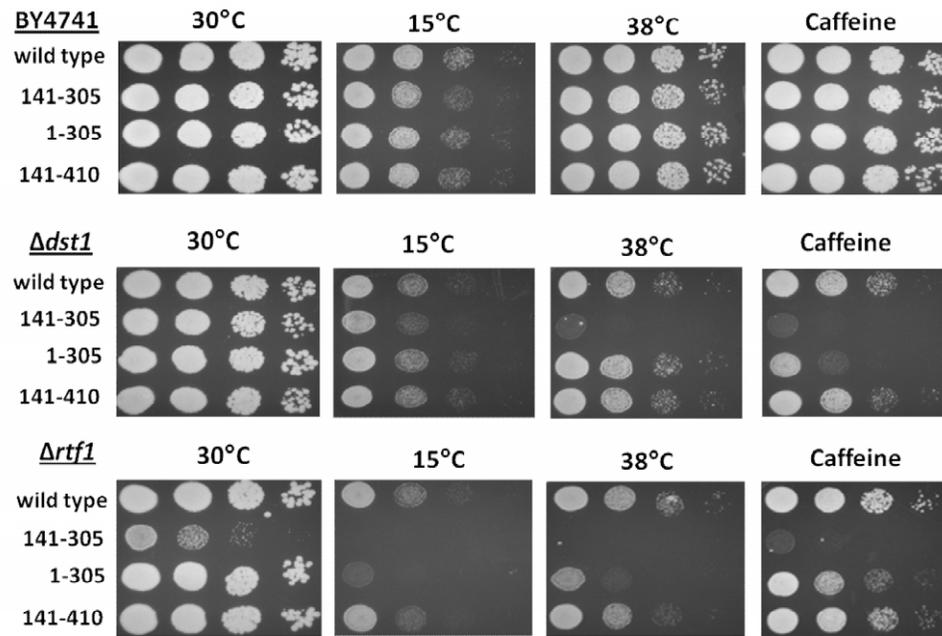
- N = No change in phenotype from the 141-305 *spn1* strain
- mEx = Mild exacerbation of the phenotype for the 141-305 *spn1* strain
- Ex = Exacerbation of the phenotype for the 141-305 *spn1* strain
- mS = Mild suppression of the phenotype for the 141-305 *spn1* strain
- S = Suppression of the phenotype for the 141-305 *spn1* strain
- = Strain is dead under indicated growth conditions
- nt = Strain has not been tested under indicated growth conditions

**Results of the genetic analysis** The 1-305 and 141-410 *spn1* alleles were transformed into the SK1, BY4741, *Δdst1*, and *Δrtf1* genetic backgrounds and were then assayed for growth under the conditions that produced dramatic phenotypes with the 141-305 *spn1* allele. In all instances, the C-terminal region of Spn1 was able to suppress these phenotypes while the N-terminal region only suppressed some of them. This would suggest that the C-terminus has a more important role within the cell.

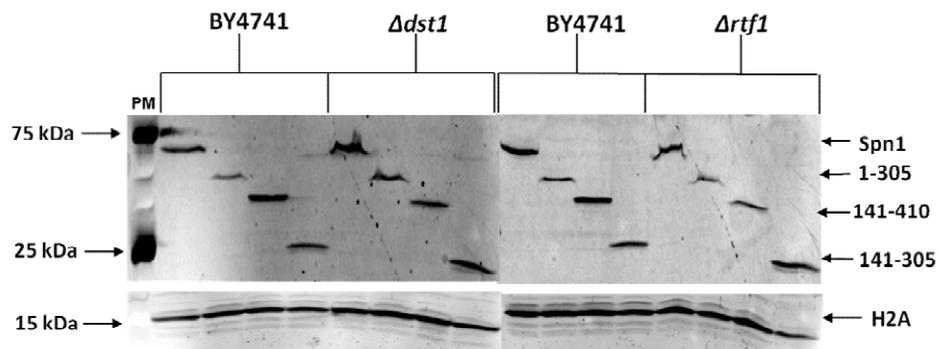
1-305 of Spn1 result in a rate of growth in between that of wild type and 141-305 (Figure 3.2). In the *Arf1* genetic background the addition of either region is again sufficient to suppress the synthetic slow-growth phenotype observed for the central domain under standard growth conditions. However, these strains were then tested at 38°C, 15°C, and on medium supplemented with caffeine. Under these stress conditions the addition of either region resulted in wild type growth on caffeine while amino acids 141-410 are required for robust growth at the altered temperatures. The conclusion from these experiments is that the interaction network between all of these factors is very complex. The N- and C-terminal regions of Spn1 have both separable and redundant functions within the cell involving Dst1 and Rtf1. To test whether any of these genetic interactions were the result of poor Spn1 expression, protein extracts were analyzed via western blot (Figure 3.3). The results suggest that Spn1 expression is fairly robust in these strains.

### **3.2c The N- and C-terminal regions of Spn1 modestly affect the activation of the *CYCI* gene**

Previous work with Spn1 at *CYCI* indicates that it is constitutively recruited to the promoter by directly interacting with Pol II. The K192N, D172G, and L218P point mutations within the central domain all diminish the Spn1 protein's interaction with Pol II resulting in faster activation kinetics and higher peak activation of the gene upon switching from glucose to a non-fermentable carbon source<sup>47; 48; 58</sup>. One hypothesis is that Spn1 is inhibiting transcription at this gene by blocking or stalling the constitutively recruited Pol II molecule. To determine whether the N- or C-terminal regions of Spn1 are responsible for this inhibition, the effect of the removal of these regions of the protein on the activation of the *CYCI* gene was tested. Total RNA extracted from cells after 0, 0.5, 1, 2, 3, 4, and 5-hours of growth in ethanol were analyzed using an S1 nuclease assay. This was performed on the wild type, K192N, 141-305, 1-305, and 141-410 *SPN1* strains. The *CYCI* signal was normalized to tRNA<sup>w</sup> (see Material and Methods). The



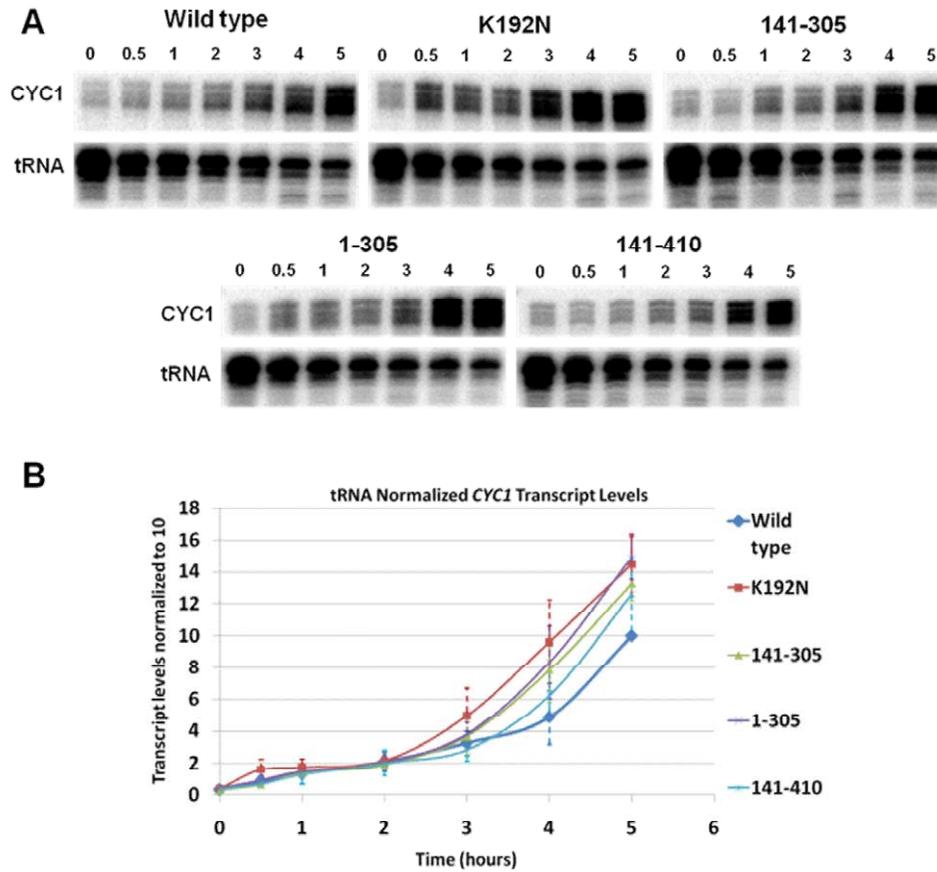
**Fig. 3.2 N- and C-terminal regions of Spn1 are functionally redundant** Similar amounts of yeast cells were diluted serially and plated onto YPD medium with and without 10mM caffeine. The strains are indicated to the left. Pictures were taken after growing cells for 2-5 days. This figure shows that the growth defects of  $\Delta dst1$  and  $\Delta rtf1$  are exacerbated by removing the N- and C-terminal regions of Spn1.



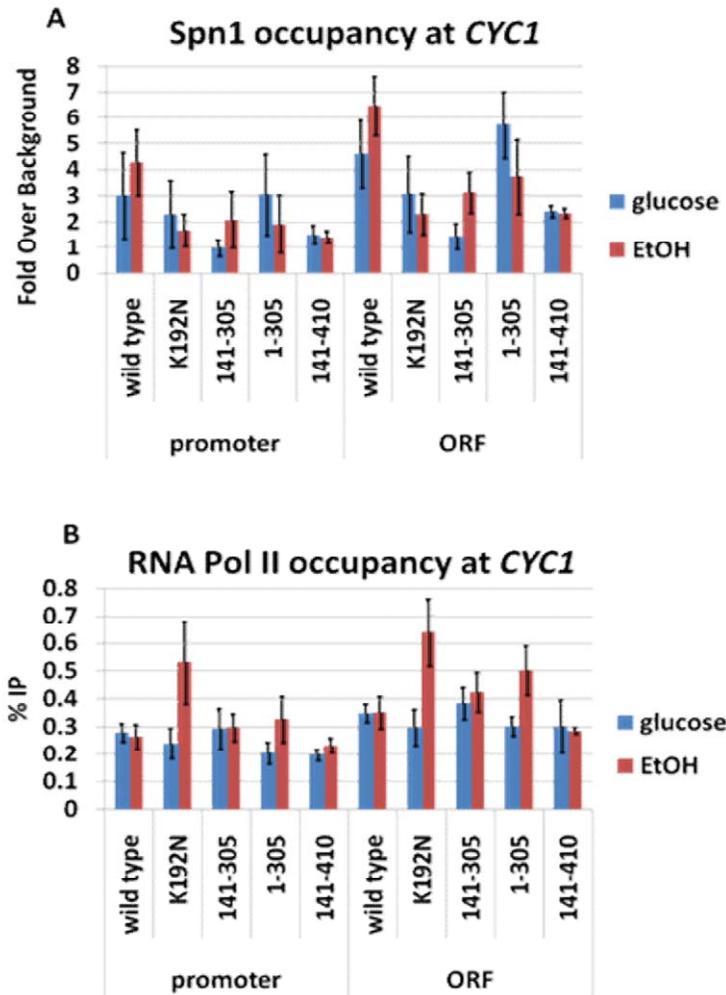
**Fig. 3.3 Spn1 protein levels are similar in the BY4741,  $\Delta dst1$ , and  $\Delta rtf1$  strains** Protein extracts prepared from the indicated strains were analyzed via western blot analysis using anti-myc and anti-H2A antibodies. Spn1 is stably expressed in all of these strains. H2A levels in these strains were used as a protein loading control. This verifies that the phenotypes observed in the genetic screen are not due to unstable Spn1 expression.

results were averaged after tRNA<sup>w</sup> normalization by setting the wild type 5-hour peak time point to 10 (Figures 3.4 a and b) for each biological replicate set individually. The K192N mutant strain was used here as a control and showed altered activation of *CYCI* similar to what has been observed previously. Although modest over-activation was detected when the N-and C-terminal regions are removed, there was not a dramatic change in the activation of the *CYCI* gene in these strains. The 5-hour time point was the only time point for the 141-305 and 1-305 strains for which modest statistically significant over-activation could be detected. The 141-410 strain behaves like wild type throughout the time course. As was seen in the genetic analyses discussed previously, these mutations behaved differently from the K192N mutation. This data suggests that the N-and C-terminal regions of Spn1 perform functions that are not relevant to post-recruitment regulated transcription.

The occupancy of these mutants at the *CYCI* gene was determined along with the effects on transcription using chromatin immunoprecipitation (ChIP). The protein/DNA complexes from samples grown in glucose and after 5-hours in ethanol were crosslinked with formaldehyde and immunoprecipitated using myc antibody and a monoclonal antibody to the Rpb1 subunit of Pol II. The DNA from each IP was amplified using real-time PCR with oligos designed to both the promoter and to the ORF of the *CYCI* gene. Spn1 occupancy is represented here as fold over background (comparison with an untagged strain) (Figure 3.5a). The K192N mutant did not show a decrease in occupancy in glucose as previously reported<sup>48</sup>. However, the expected deficiency in ethanol was still observed. It is possible that this could be the result of the two different types of PCR used (linear PCR previously, real-time PCR here) and the different primer sets used. Previous experiments with linear PCR utilized an oligo pair that spanned the *CYCI* promoter from -234 to +79 while the experiments performed here utilized real-time PCR with an oligo pair spanning a smaller region from -151 to +36. It seems plausible that the original deficiency for K192N recruitment in glucose could be due to the lower level of sensitivity of linear PCR and



**Fig. 3.4 Removal of the N- and C-terminal regions of Spn1 modestly affects *CYC1* activation** **A)** Transcript levels were analyzed with RNA prepared from the wild type, K192N, 141-305, 1-305, and 141-410 Spn1 strains in an ethanol time course at the indicated time points in hours. A tRNA<sup>w</sup> probe was used as a loading control in the S1 nuclease protections assays. Representative gel is shown. **B)** Quantification of the *CYC1* transcript levels in each strain. Level in the wild type strain after 5 hours was set to 10 for each individual biological replicate and all values were subsequently averaged. Bars represent the average of four biological replicates  $\pm$ SD.



**Fig. 3.5 Removal of the N- and C-terminal regions of Spn1 modestly reduces Spn1 occupancy at *CYC1* and does not affect Pol II occupancy.** **A)** Occupancy of the wild type, K192N, 141-305, 1-305, and 141-410 Spn1 proteins at the *CYC1* promoter and ORF were determined using a ChIP assay with anti-myc antibody. The values are represented as fold over background. Bars represent the average of three biological replicates  $\pm$ SD. **B)** Occupancy of Pol II at the *CYC1* promoter and ORF were determined using antibody to the Rpb1 subunit of Pol II (Material and Methods) in these Spn1 mutant backgrounds. The values are represented as % Input. Bars represent the average of three biological replicates  $\pm$ SD.

because the original oligo pair might not have been as efficient. The decreased occupancy for K192N in ethanol, however, is still consistent with the faster activation kinetics that is a consequence of this mutation at *CYCI*. Also about a two-fold increase in Pol II occupancy was detected in the K192N mutant strain in ethanol that supports higher levels of transcription. Pol II occupancy was similar to wild type for all of the other mutants (figure 3.5b).

Occupancy for the 141-305 *spn1* mutant was lower than for K192N in glucose, but shows an increase in ethanol. The 1-305 *spn1* mutant has a similar occupancy to wild type in both carbon sources and the 141-410 *spn1* mutant appears to have a reduced occupancy under both conditions. The pattern of occupancy for all of the strains at the promoter and the ORF are similar although the total level of signal is higher for the ORF. Considering the small size of the *CYCI* gene (~300 nucleotides) and the similar trends observed for the two primer sets, it is likely that the signal is overlapping and that the higher occupancies in the ORF could be attributed to higher primer efficiency in the PCR reactions. Since the 1-305 mutant does not have a decreased occupancy at this gene, it is possible that the N-terminal myc tag for these constructs is not accessible when it starts at amino acid 141. This is a commonality between the 141-305 and 141-410 *spn1* mutants and would explain why the occupancy is different for these three mutants while the transcription data is not. The orientation of Spn1 at the *CYCI* promoter could occlude the tag and prevent the myc antibody from binding as well during the IP.

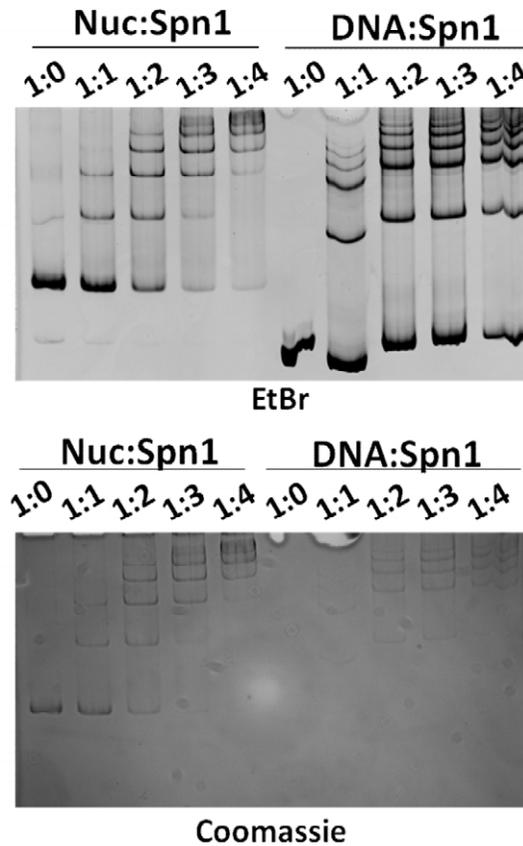
### **3.2d Wild type Spn1 can bind to nucleosomes in vitro**

Due to the Spn1 protein's connections to chromatin through its physical/genetic interactions with Spt6 and the Swi/Snf chromatin remodeling complex<sup>48; 74; 75</sup>, it seems plausible that Spn1 could interact directly with nucleosomes. To test this, both wild type Spn1 and the 141-305 *spn1* mutant were recombinantly-expressed and purified to greater than 95% homogeneity. Each of these proteins were combined with reconstituted mono-nucleosomes prepared using recombinant *Xenopus* histone proteins and the 165-base pair Widom DNA sequence as described

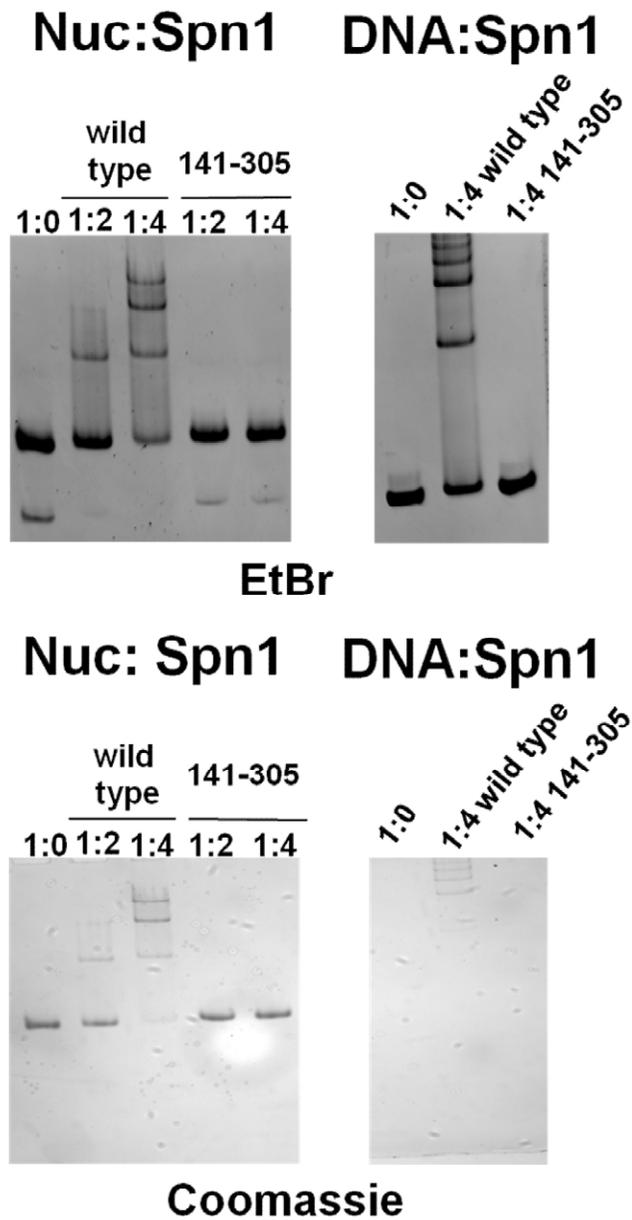
<sup>72; 73</sup>. The binding reactions were run out on a 5% native gel and were subsequently visualized using both ethidium bromide and coomassie stain (Figure 3.6). The results indicate that wild type Spn1 is capable of shifting nucleosomes in a 1:1 ratio of nucleosome to Spn1. This binding is dependent on amino acids 1-140 and 306-410 of the protein because the central domain of Spn1 is not capable of binding to nucleosomes (Figure 3.7). Binding reactions with the central domain in 50-fold molar excess to the nucleosome were attempted to verify this result and no interaction was seen (data not shown). These same binding experiments were performed using free Widom DNA and with histone dimer and tetramer to determine which components of the nucleosome Spn1 is interacting with. Again wild type Spn1 is capable of shifting DNA while the central domain of Spn1 is not. The binding experiments with histone dimer and tetramer were not as successful due to technical issues with the EMSA. Histones have a very basic pI and therefore did not resolve in a native gel very efficiently even when the electrodes were switched. Further solution studies using either AUC or SEC-MALS will be needed to characterize these interactions. One possible conclusion from this data is that the Spn1/nucleosome interaction is an artifact of Spn1 binding to the DNA and is not specific to the structure of the nucleosome. To test this, the same binding reactions described here were performed using a 147-base pair mononucleosome that does not have any free DNA overhangs. Visualized in Figure 3.8, wild type Spn1 is able to bind to this nucleosome as well. Although it is likely that Spn1 does bind to the DNA wrapped around the histones of the nucleosome, the interactions observed with a 147-base pair nucleosome suggest that the interaction is not just an artifact of affinity for DNA.

### **3.2e The N- and C-terminal regions of Spn1 are mostly intrinsically disordered in vitro**

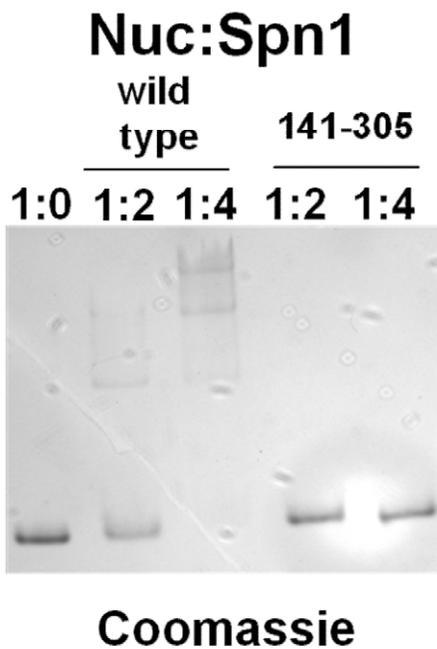
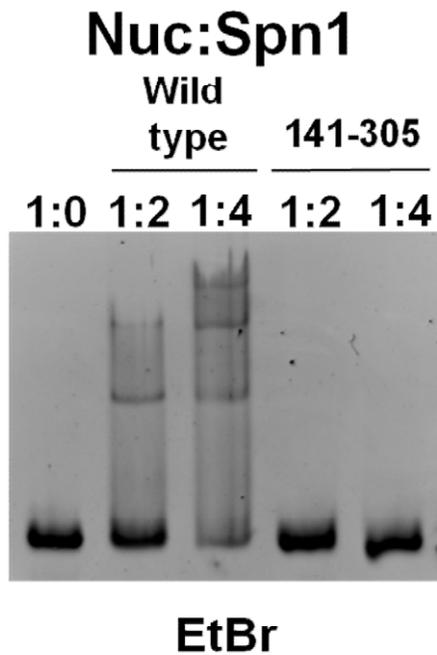
Considering that the affinity of Spn1 for nucleosomes is dependent on the N- and C-terminal regions of the protein, it seems plausible that additional ordered domains are present in these regions that make up part of the interface for this interaction. The structural characterization of these domains could provide further insight into the nature of the Spn1/nucleosome complex



**Fig. 3.6 Spn1 can bind to nucleosomes and free DNA in a 1:1 molar ratio** Wild type Spn1 was combined with reconstituted mono-nucleosomes comprised of *Xenopus* histones and the Widom 165-bp DNA sequence. The nucleosomes were at 5 $\mu$ M with Spn1 in varying ratios of excess as indicated (Nuc:Spn1). Spn1 was also tested for binding with free Widom DNA (DNA:Spn1). All binding reactions were incubated at room temperature for 30 minutes before being loaded onto a 5% native gel. The results were visualized with both ethidium bromide and coomassie stain.



**Fig. 3.7 The central domain of Spn1 cannot bind to nucleosomes or free DNA** Both wild type Spn1 and the 141-305 spn1 mutant were combined with a reconstituted mono-nucleosome comprised of *Xenopus* histones and the Widom 165-bp DNA sequence. The nucleosome was at 5 $\mu$ M with Spn1 in varying ratios of excess as indicated (Nuc:Spn1). Spn1 was also tested for binding to Widom DNA without histones (DNA:Spn1). The binding reactions were incubated at room temperature for 30 minutes before being loaded onto a 5% native gel. The results were visualized with both ethidium bromide and coomassie stain.



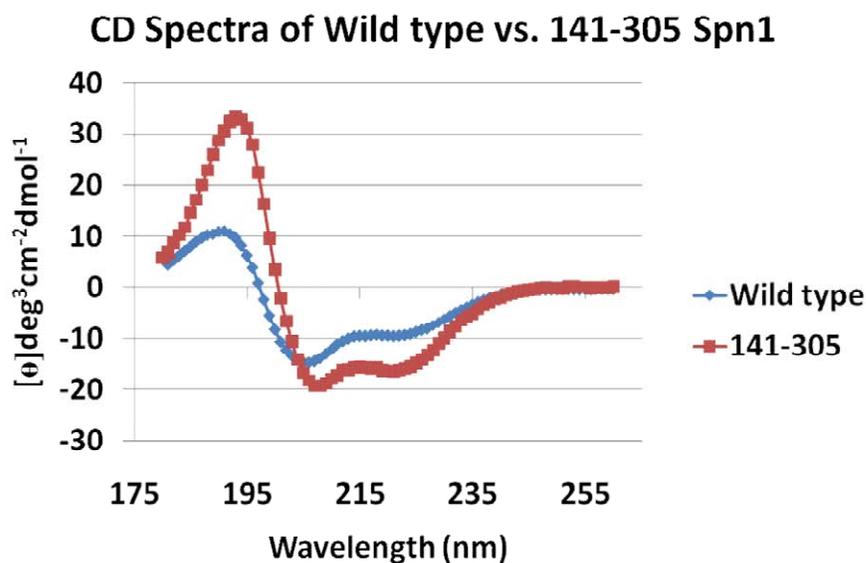
**Fig. 3.8 Spn1 does not require DNA overhangs to bind to nucleosomes** Wild type Spn1 was combined with a reconstituted mono-nucleosome comprised of *Xenopus* histones and the Widom 147-bp DNA sequence. The nucleosome was at 5 $\mu$ M with Spn1 in varying ratios of excess as indicated (Nuc:Spn1). Spn1 was also tested for binding with free Widom DNA (DNA:Spn1). The binding reactions were incubated at room temperature for 30 minutes before being loaded onto a 5% native gel. The results were visualized with both ethidium bromide and coomassie stain.

and the biochemical functions of Spn1 in general. Using recombinantly-expressed and purified wild type Spn1 and the 141-305 spn1 mutant, circular dichroism spectroscopy was performed on each protein to measure the amount of secondary structure present in the outer regions of the protein. Samples were scanned from 260nm to 180nm using a Jasco 720 spectropolarimeter. The resulting CD Spectra were analyzed using reference set 6 (SDP42) and the CDPro secondary structure prediction programs: CDSSTR, CONTINLL, and SELCON 3 (<http://lamar.colostate.edu/~sreeram/CDPro/CDPro.htm#Options%20for%20CDPro>).

From this analysis, wild type Spn1 is 36%  $\alpha$ -helical (~135-157 residues), 12%  $\beta$ -sheet (~47 to 51 residues), 21%  $\beta$ -turn (~88 to 92 residues) and 36% unstructured (117 to 137 residues). The central domain of Spn1 was calculated to be 67%  $\alpha$ -helical (~100 to 120 residues), 7%  $\beta$ -sheet (~8 to 16 residues), 19%  $\beta$ -turn (~28 to 34 residues) and 8% unstructured (~12 to 16 residues). These values are the averages of two independent biological replicates (figure 3.9). The crystal structure of the central domain of Spn1 has 112 residues in an  $\alpha$ -helical orientation and no  $\beta$ -sheet content. The accuracy in calculating  $\beta$ -sheet and  $\beta$ -turn content is not as high as  $\alpha$  helices using CD spectroscopy (more so for  $\beta$ -turn). However, the  $\alpha$ -helical content was close to that of the crystal structure and shows that this protein is a good control for the experiment. The data collected for wild type Spn1 suggests that a modest amount of additional secondary structure could be present in the N-and C-terminal regions, but additional experimentation will be required to verify this. However, it is clear that a fair amount of intrinsic disorder is present in these regions (at least 100 residues and perhaps quite a bit more). Characterization of this additional secondary structure could be difficult due to this intrinsic disorder.

### **3.2f Contributions on the N-and C-terminal regions on the hydrodynamic properties of Spn1**

Although CD spectroscopic data suggests that the N-and C-terminal regions of Spn1 are comprised of a combination of both secondary structure and intrinsic disorder, it is still unclear



CDPro	$\alpha$ helix	$\beta$ sheet	$\beta$ turn	unstructured
Wild type	135 to 157	47 to 51	88 to 92	117 to 137
141-305	100 to 120	8 to 16	28 to 34	12 to 16

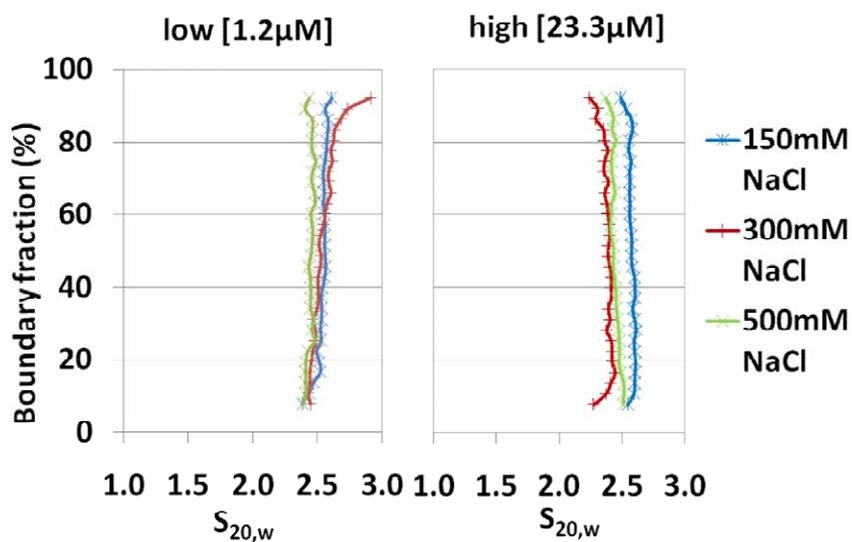
**Fig. 3.9 Analysis of secondary structure in the N-and C-terminal regions of the Spn1 protein.**

The wild type Spn1 and 141-305 spn1 mutant proteins were recombinantly expressed and purified to greater than 95% homogeneity. Proteins were extensively dialyzed against 10mM sod. phos. (pH7.4) + 100mM NaF buffer. Calculations were done using the CDPro software

(<http://lamar.colostate.edu/~sreeram/CDPro/main.html>) and are the ranges of two independent biological replicates (in amino acid residues). Although the wild type protein is predicted to have some additional secondary structure, a significant percentage of disorder present lowers the overall intensity of the spectrum in comparison to the highly globular central domain.

how these regions contribute to the overall behavior of the protein. The hydrodynamic properties and functions of this protein could be dependent on the solution behavior of these regions relative to the central domain. Due to the highly charged nature of the entire protein (N-terminus pI=4.5, C-terminus pI=10.6, central domain pI=8.1 but with several charged patches) it is possible that a range of extended and compact structures could be possible for wild type Spn1 based on intramolecular electrostatic interactions. Performing AUC experiments on Spn1 will also give an idea of how well-behaved the protein is and how amenable it would be to further structural analysis. In order to answer these questions and to further verify the CD spectroscopic data discussed here, sedimentation velocity experiments were done on both wild type Spn1 and the 141-305 spn1 mutant at varying salt and protein concentrations. The 141-305 protein was monitored over a ~30-fold concentration range and wild type Spn1 over a ~20-fold concentration range at 230nm and 280nm. Three different salt concentrations were also used (150, 300, and 500mM NaCl) to test the salt dependence on the hydrodynamic properties of these proteins. Two independent biological samples were tested to ensure that the results were reproducible.

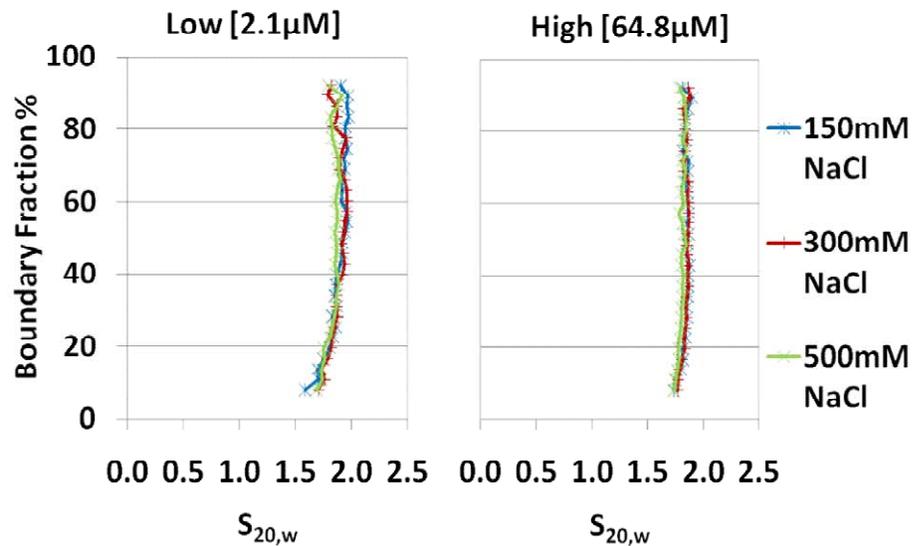
Both proteins sedimented as well-behaved homogeneous species and no dependence on salt or protein concentration were detected. The average sedimentation coefficient for 141-305 was 1.8S and was 2.5S for wild type Spn1 (Figures 3.10 and 3.11). An extensive 2DSA and Genetic Algorithm analysis was done using Ultrascan software for the higher protein concentration data sets and each protein was calculated to be a monomer under the conditions tested with RMSD values no larger than  $6 \times 10^{-3}$ . The frictional ratio is the main parameter calculated from AUC to determine the general shape of a protein in solution. This is calculated by dividing the frictional coefficient of the protein by the frictional coefficient of a sphere with the same mass. A frictional ratio of one indicates that the protein is a perfect sphere while increasingly higher ratios suggest higher levels of asymmetry for the dimensions of the protein. The frictional ratio for the central domain was ~1.4 and was ~2 for full length Spn1. This



Wild type	$S_{20,w}$	MW	f/fo	RMSD
150mM	2.60	54.7	2.05	0.005045
300mM	2.54	50.2	1.98	0.005062
500mM	2.45	50.1	2.05	0.004986

**Fig. 3.10 Wild type Spn1 sediments homogeneously independent of protein or salt concentrations**

Recombinant wild type Spn1 was purified to greater than 95% homogeneity and was dialyzed extensively against 20mM Tris-HCl (pH 7.5) and 150, 300, or 500 mM NaCl prior to the analysis. AUC sedimentation velocity was used to determine the hydrodynamic contributions of the N- and C-terminal regions of this protein at two protein concentrations (1.2µM left, and 23.3µM right). This protein was spun at 45,000 rpm and it sedimented homogeneously regardless of the protein or salt concentration used. Extensive 2DSA and genetic algorithms in Ultrascan were performed (on the 23.3µM data) to calculate the  $S_{20}$  corrected for water (in Svedberg units), the molecular weight (MW in kilodaltons), the frictional ratio ( $f/fo$ ), and the RMSD for each condition.



141-305	$S_{20,w}$	MW	$f/f_0$	RMSD
150mM	1.84	21.2	1.43	0.006612
300mM	1.82	21.1	1.44	0.005762
500mM	1.81	21.6	1.47	0.006671

**Fig. 3.11 141-305 spn1 mutant sediments homogeneously independent of protein or salt concentrations** The central domain of Spn1 was purified to greater than 95% homogeneity and was dialyzed extensively against 20mM Tris-HCl (pH 7.5) and 150, 300, or 500 mM NaCl prior to the analysis. AUC sedimentation velocity was used to determine the hydrodynamic contributions of the N- and C-terminal regions of this protein at two protein concentrations (2.1µM left, 64.8µM right). This protein was spun at 50,000 rpm and it sedimented homogeneously regardless of the protein or salt concentration used. Extensive 2DSA and genetic algorithms in Ultrascan were performed (on the 64.8µM data) to calculate the  $S_{20}$  corrected for water (in Svedberg units), the molecular weight (MW in kilodaltons), the frictional ratio ( $f/f_0$ ), and the RMSD for each condition.

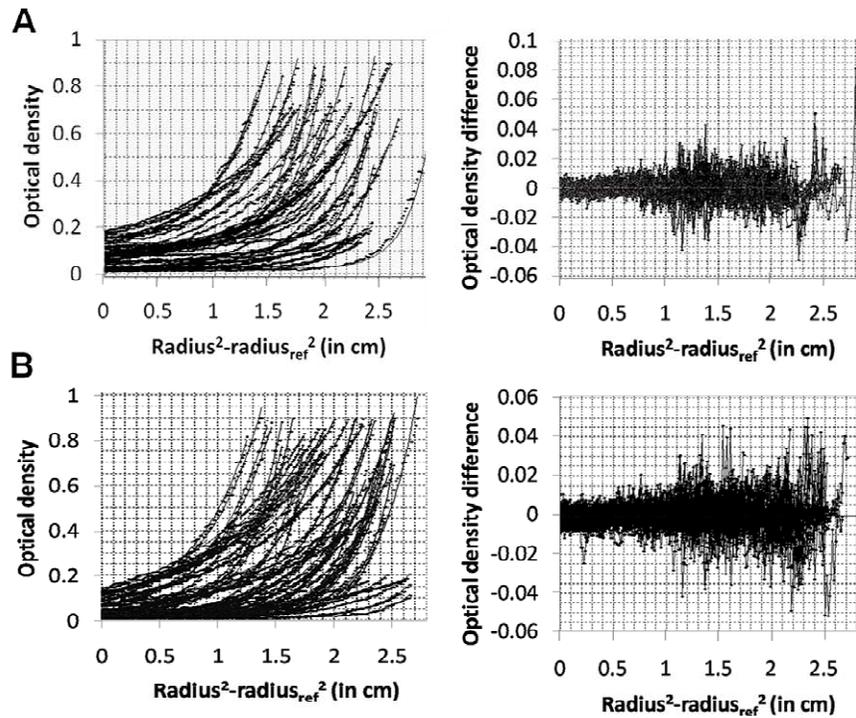
indicates that the Spn1 protein is fairly asymmetric overall, especially in comparison with the central domain. This means that the volume taken up by the additional amino acids comprising the N-and C-terminal regions of the protein are spread out in solution rather than tight and compact creating a larger amount of friction with the surrounding water molecules. This strongly supports the CD data shown here suggesting that many of the amino acids in these regions of the protein are unstructured. Further experimentation will still be required to determine if this disorder is predominantly in one region or the other.

Since the 150mM salt condition resulted in a slightly larger wild type Spn1 protein, sedimentation equilibrium experiments were performed at 150mM and 500mM NaCl to ensure that multimerization is not starting to occur under these conditions. Wild type Spn1 was sedimented over a ~23-fold concentration range at four different speeds and was scanned every four hours to ensure it had reached equilibrium. The data was globally fit to a variety of modeling programs within Ultrascan and each one showed that Spn1 was a monomer under both salt concentrations (Figure 3.12 a and b).

### **3.3 Discussion**

#### **3.3a N-and C-terminal truncated Spn1, the new breed of Spn1 mutation**

The results of the genetic analysis performed here was the first evidence that the N-and C-terminal regions flanking the central domain of Spn1 perform physiologically relevant functions in the cell. Even though the central domain can robustly cover the functions of Spn1 under the growth conditions tested in the SK1 and BY4741 genetic backgrounds, it is likely that the functions performed by the N-and C-terminal regions can be compensated for by other proteins in the cell. The evidence supporting this idea are the synthetic genetic interactions observed between the N-and C-terminal regions of Spn1 and the *DST1* and *RTF1* genes. At varying temperatures and with the addition of caffeine to the medium, simultaneously deleting either of these genes and the outer regions of the Spn1 protein is lethal. Phenotypes of this



**Fig. 3.12 Sedimentation Equilibrium shows that wild type Spn1 behaves as a monomer at 150mM and 500mM NaCl** Purified wild type Spn1 was dialyzed extensively against 20mM Tris-HCl (pH 7.5) and 150 mM NaCl prior to the analysis. Samples spanning an initial 24-fold protein concentration ranging from 1.2 $\mu$ M to 28 $\mu$ M were sedimented to equilibrium at 18700, 26,500, 32,500 and 37,500 rpm. Two equilibrium scans were collected at each speed. A) Forty-six data sets were globally fit to a single-ideal species model in Ultrascan. The fits to the data are shown as *solid lines* on the left. On the right is the goodness of fit residuals for the data. The variance for the fit to a single ideal component model was  $7.21 \times 10^{-5}$ . The analysis returned a mass of 49,910 Da. B) Forty-six data sets were globally fit to a single-ideal species model in Ultrascan. The fits to the data are shown as *solid lines* on the left. On the right is the goodness of fit residuals for the data. The variance for the fit to a single-ideal component model was  $4.56 \times 10^{-5}$ . The analysis returned a mass of 47,760 Da.

severity are not observed for either of the mutations individually suggesting that these proteins are tightly connected regarding certain aspects of transcription.

*DST1* encodes a protein that is also known as TFIIS in higher eukaryotes and has been well-characterized for its ability to bind to Pol II and stimulate elongation by relieving pausing during transcription. Dst1 also plays a role in transcription initiation and stability of the PIC at promoters<sup>76; 77; 78</sup>. It is important to note that amino acids 213-291 of Spn1 and amino acids 1-77 of the mouse TFIIS ortholog share both sequence and structural similarity<sup>58</sup>. With this in mind, the genetic interactions observed here suggest that Spn1 could perform functions involving elongation similar to Dst1, but that these functions require the regions flanking the central domain. Pol II processivity decreases under stress-inducing conditions and the cell cannot compensate for this when missing Dst1 and a fully functional Spn1 protein. What is difficult to understand about this functional relationship is the requirement for either the N-terminus or the C-terminus of Spn1 at elevated temperatures, but the specific requirement for the C-terminus when cells are grown in medium supplemented with caffeine. This suggests that the N- and C-terminal regions of Spn1 are biochemically similar in some way. At first glance this seems unlikely due to the opposite pI's of each region (C, 10.6; N, 4.5), but two small regions share sequence similarity between the two termini (Figure 3.13). If these are crucial interaction sites with other proteins in the cell, this could explain the redundancy observed for each region. In the future, Spn1 should be tested for an ability to stimulate the elongation rate of Pol II in vitro. This would further clarify the phenotypes observed here and indicate that Spn1 and Dst1 perform similar functions.

*RTF1* encodes a member of the Paf complex and has functions related to transcription initiation, elongation, and post-translational modifications of the histone proteins that include methylation and ubiquitination<sup>75; 79; 80; 81</sup>. The phenotypic pattern for these two genes is similar to what is observed with *DST1* except that an almost lethal synthetic genetic interaction is detected

```

44  ETSKGTSDTK
336 PTSRGSSVQTL
    ** * * *

117 SSATPSSRQELEEK
336 PTSRGSSVQTLYEQ
    ** * * *

```

**Fig. 3.13 Partial sequence similarity between the N-and C-terminal regions of Spn1** Sequence alignments were performed using the expasy sequence alignment tool (<http://ca.expasy.org/tools/sim-prot.html>) comparing the sequences of the N and C terminal regions of Spn1. The genetic analysis performed here suggests that the functions of these regions are partially redundant. Although a large amount of similarity was not found, two small clustered regions are very similar and could be a crucial interface for interacting with other proteins.

under standard growth conditions. This suggests that the functional relationship between these two proteins is especially important. An appealing connection for this functional relationship is the histone modifications that the Paf complex is responsible for. Considering the work done here that indicates that Spn1 can interact with nucleosomes, it seems plausible that improper methylation or ubiquitination of histones from deleting *RTF1* could disrupt the Spn1 protein's interaction with nucleosomes. Alternatively these modifications could be required to disrupt this interaction instead.

Another point that should be discussed is that the K192N point mutation genetically interacts extensively with the Swi/Snf chromatin remodeling complex. The K192N mutant is able to suppress several slow-growth phenotypes for *SNF* gene deletions that include growth under alternative carbon sources and –inositol medium. *SNF* gene deletions also suppress the temperature-sensitive phenotype of K192N<sup>48</sup>. The absence of any significant genetic interactions with the *SNF5* gene and the N- and C-terminal regions of *SPN1* suggest that this functional relationship is specific to the central domain. This corresponds with the molecular data indicating that these regions of the protein do not affect the activation of the *CYCI* gene like the K192N mutation. The model for Spn1 function prior to these experiments was that Spn1 constitutively occupies the promoter and inhibits transcription by preventing recruitment of the Swi/Snf chromatin remodeling complex. When switched into ethanol, the Spt6 histone chaperone will interact with Spn1 and alleviate this inhibition allowing for increased transcription. The K192N mutant is deficient for occupancy under all conditions. This results in the constitutive recruitment of Swi/Snf, no Spt6 recruitment in ethanol, and faster activation<sup>48</sup>. CHIP data collected here indicates that the 141-305 and the 1-305 spn1 mutant proteins are deficient for occupancy in glucose, and the K192N, 141-305, and 141-410 mutants are deficient in ethanol. The transcription and occupancy data do not coincide with this model since faster activation was not observed for any of the mutants except for K192N. A plausible explanation for this lack of occupancy is that

the orientation of Spn1 at the promoter occludes the myc tag on the N-terminus of these mutants (adjacent to amino acid 141) preventing the myc antibody from binding during the IP reaction. A conformational change could take place in ethanol that makes the tag a little more accessible. This would explain why a lower level of occupancy is detected for these mutants without a significant change in transcription. If this is the case then there should also not be a significant change in Spt6 or Swi/Snf recruitment to *CYCI*. Spt6 should still be recruited when the cells are grown in ethanol with or without the N-and C-terminal regions of Spn1. This would be followed by recruitment of Swi/Snf and activation of the gene.

The main conclusion from these experiments is that although the N and C terminal regions have physiologically relevant functions in the cell, these functions are different from previous work done on Spn1 using the K192N mutation. Removing these regions of Spn1 is a new type of mutation in the field of Spn1 research and a broader approach is likely needed to better understand all of the functions of the Spn1 protein in the cell.

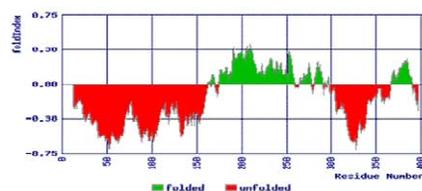
### **3.3b Spn1, a node of RNA Pol II related functions**

The interactions detected between Spn1 and nucleosomes provide a strong basis for new Spn1 research in chromatin dynamics and regulation. This is the first binding partner for Spn1 that is dependent on the N-and C-terminal regions of the protein. The CD and AUC data collected suggesting that these regions of the protein are intrinsically disordered make this interaction with nucleosomes particularly intriguing. Intrinsically disordered proteins are documented for becoming ordered when binding partners physically contact them (reviewed by Uversky and colleagues)<sup>82; 83; 84</sup>. This is a plausible possibility for Spn1 when it binds to nucleosomes. Information on the structure and biochemical properties of the ordered domains within Spn1 that could result from this interaction would provide insight into how and why Spn1 interacts with nucleosomes and the potential relevance of this interaction in vivo. Proteins with substantial disorder are difficult to crystallize, but a co-crystal structure with Spn1 and a nucleosome could

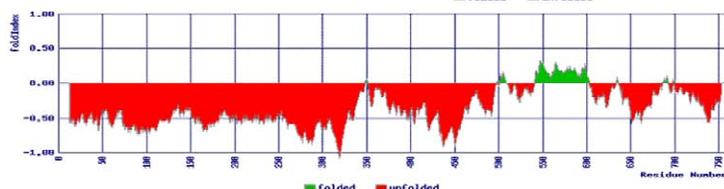
provide more information on the N-and C-terminal regions of the protein. Fold index plot algorithms predict regions of a protein that will fold into ordered domains based on the primary sequence of the protein<sup>85</sup>. For Spn1 orthologs throughout evolution, this analysis results in a central domain that is flanked by intrinsically disordered N-and C-termini (Figure 3.14). This suggests that the overall pattern of order and disorder for the Spn1 protein is conserved and is required for proper functioning in vivo in all eukaryotes. Another possibility that must be considered is that additional secondary structure could be present in one of the outer regions of Spn1 as the CD analysis suggests. Rather than an intrinsically disordered region of Spn1 interacting with nucleosomes, it could be a folded domain that is the interface. The requirement of the N-and C-terminal regions individually should be tested to determine if the interface with the nucleosome is specific to one of them. If this is the case, further biophysical characterization could be done to determine if an additional domain is present in the Spn1 protein. Further experimentation also needs to be done to ensure the interaction between Spn1 and the nucleosome is not an artifact of Spn1 having an affinity for DNA. Testing to see if Spn1 has a higher affinity (Kd) for nucleosomes than for free DNA would answer this question.

The combination of the *CYCI* transcription data and the nucleosome binding data suggest that the central domain of Spn1 could specifically perform functions involving direct regulation of Pol II activity while the N-and C-terminal regions could affect Pol II by regulating chromatin structure. One of the Spn1 protein's binding partners, Spt6, exhibits a similar pattern of activity in the cell. Spt6 has also been shown to affect transcription as an elongation factor and by acting as a histone chaperone to alter chromatin structure<sup>86; 87; 88</sup>. One study showed that a mutation that decreased nucleosome occupancy in the coding region of genes did not alter mRNA levels in the cell, suggesting that they are two separate functions of Spt6 performed by different parts of the protein<sup>88</sup>. The Spn1-Spt6 interaction is especially intriguing since studies have shown that mutations that diminish the affinity of these binding partners for one another result in slower

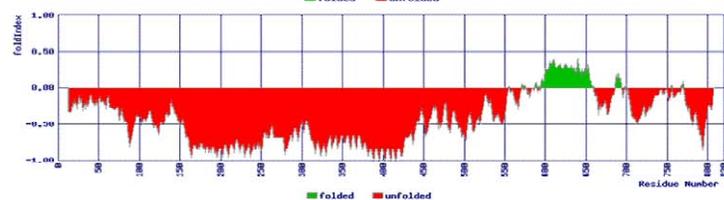
*Saccharomyces cerevisiae*



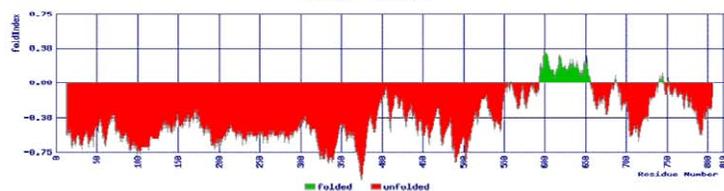
Mouse



*Drosophila*



Human



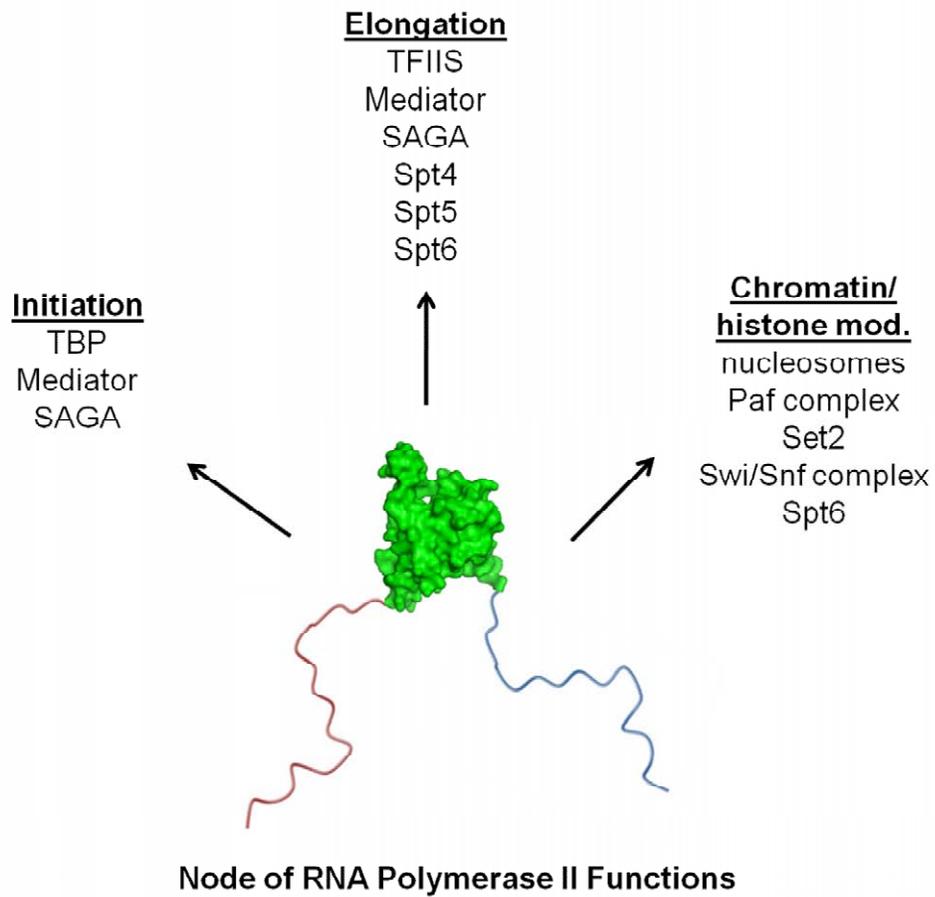
**Fig. 3.14 Fold index plot calculations suggest that the structural characteristics of Spn1 are conserved throughout evolution** Fold index plot algorithms (<http://bip.weizmann.ac.il/fldbin/findex>) that predict the regions of a protein that have secondary structure predicts that Spn1 orthologs throughout evolution are comprised of a highly globular central domain that is flanked by two intrinsically disordered N-and C-termini. This suggests that this is a conserved structural trait that emparts essential functions to the Spn1 protein in vivo. A window size of 25 was used for the plots here.

growth rates for yeast cells. A complete disruption of this interaction makes cells inviable<sup>56; 57</sup>. In this same study, nucleosome gel shifts were performed to better understand the relationship between Spn1-Spt6 and the nucleosome. This group only used an N-terminal truncated version of Spn1 and said that it was not capable of binding to nucleosomes. The work done here would indicate that this is not accurate and that binding would have been detected with use of the full-length Spn1 protein. Further understanding the Spn1-Spt6-nucleosome relationship could provide further insight into this protein's effects on chromatin dynamics.

Although in-depth mechanisms regarding Spn1 function have yet to be determined within the cell, it is evident that the functions of this protein are vast and involve multiple aspects of transcription. Spn1 could play a role in transcription initiation with its connections to TBP, the SAGA complex, the Mediator complex, and the Paf complex, which all aid in the recruitment of Pol II to various promoters<sup>47; 48; 54; 89</sup>. Spn1 also has a well-established list of both physical and genetic interactions with factors involving elongation such as Spt4, Spt5, Spt6, TFIIS, as well as the Mediator complex<sup>48; 54; 56; 57; 58; 74; 75</sup>. Finally Spn1 also has connections to chromatin remodeling and histone modifications via physical and genetic interactions with Swi/Snf, Spt6, the Set2 histone methyltransferase, the Paf complex, and the SAGA complex<sup>48; 54; 56; 57; 58; 60; 74; 75; 90; 91</sup>. A current model of Spn1 function would be the central domain of Spn1 flanked by intrinsically disordered N- and C-termini connected to every aspect of Pol II transcription (Figure 3.15).

### **3.3c Future Directions**

Prior to the research described here, the K192N point mutation was the hallmark mutation for Spn1 research. It is the mutation that suppressed the post-recruitment defective TBP allele and allowed for the identification of Spn1. A significant amount of what is known about Spn1 function has been due to research aimed at characterizing the defects of this mutant. Even the designs of the first two sets of experiments performed here were based on results obtained for



**Fig. 3.15 Spn1, a node of RNA Polymerase II dependent functions** presented here is a model of the different aspects of transcription that Spn1 has been shown to affect and the factors that it interacts with physically and/or genetically.

the K192N mutant. The outcome of the initial genetic analysis and the effects on *CYCI* transcription of the N-and C-terminal regions of Spn1 emphasized the fact that these mutations are not the same as the K192N mutant. These regions of Spn1 do not genetically interact with the Swi/Snf chromatin remodeling complex, and the phenotypes observed in the  $\Delta DST1$  and  $\Delta RTF1$  genetic backgrounds are not the same as for K192N. The N-and C-terminal regions of Spn1 modestly affect the activation of *CYCI*. This is important to keep in mind because future experiments designed to understand the functions of the N-and C-terminal regions should be aimed at different aspects of transcription. The effects of the removal of the N-and C-terminal regions of Spn1 on a constitutively active gene such as a ribosomal protein synthesis gene should be tested. The GAL genes would also be good candidates since they are activation dependent, but still recruitment-regulated.

Another example of broadening the scope of Spn1 experimentation would be the continued characterization of the interactions observed between Spn1 and nucleosomes. Functions involving chromatin would affect transcriptional mechanisms in a much more generalized way as opposed to just postrecruitment-regulation. Titration of recombinant Spn1 into in vitro transcription reactions in the absence and presence of chromatin would be a possible experiment to answer this question. If transcription rates are increased by adding Spn1 to reactions with chromatin templates, this would suggest that Spn1 acts as a histone chaperone to aid in nucleosome remodeling. Another experiment would be to test for cryptic transcription from promoters in vivo using the same strains used here where the N-and C-terminal regions of Spn1 have been deleted. Many histone chaperones are responsible for depositing nucleosomes onto DNA after the transcription machinery has transcribed it. A consequence of disrupting this activity is that the next round of transcription will start at a site downstream of the +1 start site at the 5' end of the ORF of a gene. Spn1 can physically interact with the Spt6 histone chaperone

that has been documented for cryptic start site activity and this is suggestive that Spn1 also plays a role in this process<sup>92; 93</sup>.

Experiments that test the structural requirements of the interaction of the nucleosome also need to be done. One big question that remains is whether Spn1 binding to nucleosomes is specific to the N-terminal region, the C-terminal region, or if both are required. This can be answered by performing the same gel-shift experiments with recombinantly-expressed and purified Spn1 proteins that are lacking each N-and C-terminal region individually. A plausible hypothesis is that both regions are required for the interaction. This is based on the comparison of the opposing pIs of each region. The acidic N terminal region could potentially interact with the histones while the more basic C terminal region could interact with DNA. To answer the question of whether these regions of the Spn1 protein become ordered upon interacting with the nucleosome, CD spectroscopy experiments could be performed on Spn1 with and without nucleosomes. A simpler version of this experiment would be to use DNA without histones. The results from all of these experiments would provide a better mechanistic picture of how Spn1 is functioning in the cell.

## **Chapter 4: Spn1, a highly conserved protein with a connection to aging in *Saccharomyces cerevisiae***

The work described in this chapter probes the Spn1 protein's role in the process of chronological aging in yeast. I wrote this chapter and contributed all of the figures. Aaron Docter helped with performing the assay for several of the biological replicates and for counting the colonies for the majority of the experiments.

### **4.1 Abstract**

Evidence indicates that there are multiple, evolutionarily conserved pathways within the cell that are responsible for determining the rate at which an organism will age. These pathways include: autophagy, ribosome biogenesis, protein translation, mitochondrial activity and function, heterochromatic stability, maintenance of the genome, and apoptosis<sup>3; 4; 5; 6; 7</sup>. A general aging mechanism is emerging that involves the proper maintenance of the molecular metabolism of a cell, or the turnover of its old and damaged proteins, nucleic acids, and lipids over time. In accordance with this idea, the activation of various stress-response pathways that aid the cell in decreasing the total levels of damaged macromolecules over time are considered to be the major determinants of lifespan<sup>3; 4; 5; 6</sup>. A genomic ChIP-chip experiment performed by two independent groups revealed that Spn1 is recruited to a majority of the genes in the yeast genome<sup>1; 2</sup>. The possibility that Spn1 regulates the genes involved in these pathways is highly suggestive that this protein could be an aging factor within the cell. Chronological aging assays revealed that the removal of the disordered N and C terminal regions of the Spn1 protein dramatically increase the lifespan of the BY4741 strain of yeast. These results further verify the physiological importance of this protein and the need for further Spn1 research.

### **4.2 Introduction**

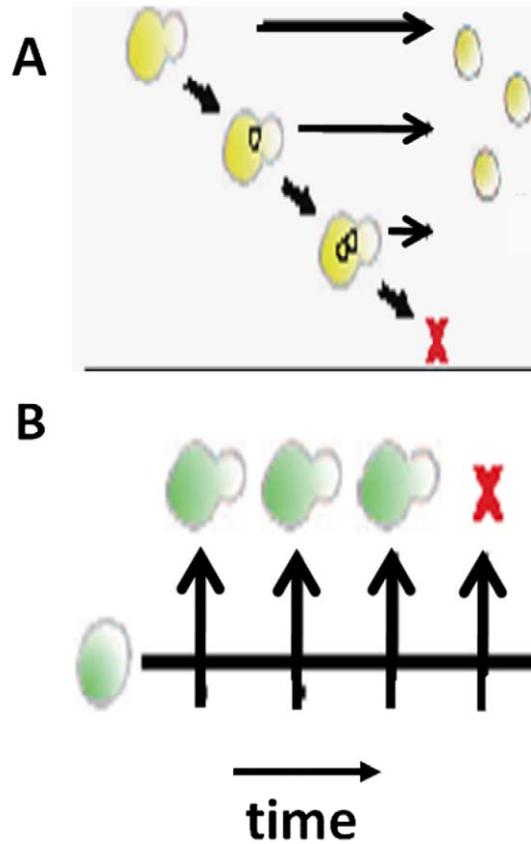
Interest in studying aging at the molecular level came about from the observation that dietary/caloric restriction (CR) increases the lifespan of a variety of model organisms that include

yeast, worms, flies, and mice<sup>94; 95</sup>. Despite these findings, the idea that aging is molecular in nature was approached with skepticism. Instead, it was considered a non-adaptive process since genes were not considered to have “death” or “life” functions<sup>96</sup>. This theory dissipated when several genes were characterized as playing a role in the lifespan of a cell, and since then entire cellular pathways have been implicated in this process<sup>97; 98; 99</sup>. Molecular based mechanisms for aging support the research of this complex process in lower organisms, even in yeast with 2 billion years of divergence. It is the conservation of the structure and function of proteins across evolution that makes this research in lower eukaryotes strongly applicable<sup>96</sup>. Therefore, the primary goal in the field has been the characterization of the molecular mechanism behind CR’s effects on lifespan in hopes of being able to mimic these effects in humans without actually having to eat less. Along with general lifespan increases, these efforts could potentially help to delay the onset of several age-related diseases such as Alzheimer’s disease, macular degeneration, and even cancer<sup>100</sup>.

To date, the perturbation of the following cellular pathways have been implicated in aging: ribosome biogenesis and translation, mitochondrial activity and function, autophagy, heterochromatic stability and maintenance of the genome, and apoptosis<sup>96; 101</sup>. A recent theory suggests that the cellular pathways that affect aging function to defend the cell against internal and external stresses that result from a broad range of environmental conditions<sup>5; 6; 7; 102</sup>. Some of these stresses include: temperature fluctuation, ionizing radiation, oxidative stress, and starvation conditions<sup>7</sup>. Each stress can damage the molecular machinery of the cell and require more continuous recycling and repair of these components over time. Increasing levels of damaged and toxic macromolecules is therefore considered the main molecular cause of aging<sup>5; 6; 7; 103; 104</sup>. Supporters of this model argue that the genetic and environmental perturbations that change the lifespan of various model organisms are the result of alterations in cellular stress responses that regulate molecular homeostasis.

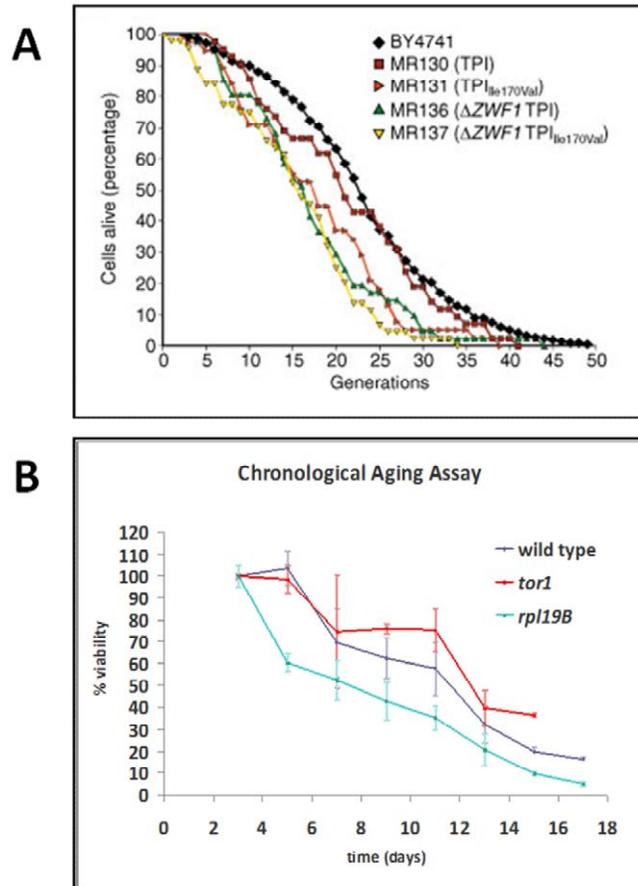
There are two types of aging that can be studied in yeast, replicative aging and chronological aging. RLS (replicative life span) measures the total number of times a mother cell can divide before senescing. The experiment to measure RLS is performed via micro-manipulation of individual cells under a microscope. As each daughter bud forms, it is separated from the mother cell and removed from the plate (Figure 4.1a). In order to obtain statistically relevant data, a hundred or more cells are analyzed at a time<sup>96</sup>. Traditionally, this type of aging is thought to relate to mitotic cells such as stem cells that divide to replenish all of the different types of cells in the tissues and organs of the body. An advantage of an increase in replicative lifespan would be that a stem cell could contribute more divisions for maintaining a particular tissue in a multicellular organism before senescing.

The other type of aging studied in yeast is chronological lifespan (CLS). The experiment to measure CLS asks a different question regarding the lifespan of yeast cells. Rather than determining how many times a cell can divide, it measures how long a population of cells can remain viable in stationary phase and re-enter the cell cycle when introduced to fresh nutrients<sup>105</sup>. CLS is classically done by growing a culture of cells into stationary phase with continued aeration for about 2-4 weeks. Aliquots of the culture are diluted and spread onto solid growth medium plates in order to track the percent viability of a strain over time by counting the number of cells that can form colonies (representative of the percentage of the population of cells that can re-enter the cell cycle, see Figure 4.1b)<sup>105</sup>. CR conditions for both assays involve reducing the glucose concentration from 2% to 0.5% or reducing the amino acid concentration in the media<sup>62</sup>; <sup>105</sup>. CLS is similar to the aging of the non-mitotic or non-dividing cells making up all of the tissues and organs in the human body. Increasing the chronological lifespan of a cell would cause tissues to degenerate more slowly; requiring less total divisions from stem cells to maintain a tissue. This would result in slowing down the onset of age-related diseases (Figure 4.2a and b for a representation of what replicative and chronological data typically look like)



**Fig. 4.1 Representations of the two different types of aging studied in yeast. 1A)** Replicative lifespan: each daughter bud is pulled away until the mother stops dividing. **1B)** Chronological lifespan: culture is grown into stationary phase and is periodically tested for the percentage of cells that still have the ability to start dividing again.

<http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.0030084>



**Fig. 4.2 Representations of yeast lifespan data** **A)** replicative lifespan data (Ralser *et al. Journal of Biology* 2007 6:10 doi:10.1186/jbiol61) and **B)** chronological lifespan data (theoretical).

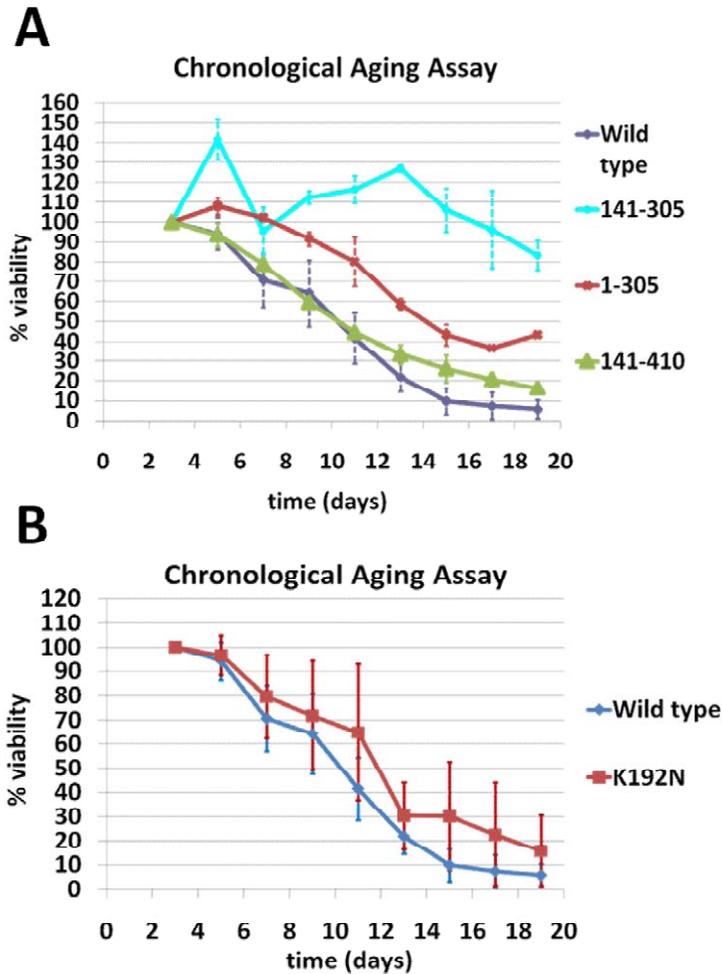
Recently genome-wide occupancy data was obtained in collaboration with Dr. Vishy Ayer to determine whether Spn1 is a specific or general transcription factor and to get an idea of what overall cellular processes Spn1 is likely to affect. The unpublished results of this experiment performed in biological triplicate under standard growth conditions showed that Spn1 significantly occupies the promoter and/or ORF of 67% percent of the ribosomal protein synthesis genes and a variety of genes involved in mitochondrial respiration and energy production, with a p-value below 0.01. Following this experiment, two independent groups published genomic ChIP-chip data indicating that Spn1 is actually recruited to a majority of the genes in the yeast genome<sup>1,2</sup>. In addition to this, fitness data accessible on the Yeast Genome Database indicates that in a diploid strain, deletion of one Spn1 allele followed by treatment with various drugs produces phenotypes similar to strains with various ribosomal protein synthesis genes deleted that are treated with the same drugs<sup>106</sup>. From the results of these experiments it is plausible that Spn1 could regulate the transcriptional output of a large percentage of genes throughout the genome including genes involved in lifespan determination. Many of the protein components of the electron transport chain have peroxidase activity that can help to reduce the total level of reactive oxygen species (ROS) and prevent the cell's molecular machinery from being damaged. The ChIP-chip data and previous data indicating that Spn1 regulates the *CYCI* gene, encoding a protein in the electron transport chain and an ROS scavenger, suggests that Spn1 could also contribute to the overall activity of mitochondria within the cell. In addition, supplementing the growth medium with caffeine extends lifespan of yeast in comparison to medium without caffeine<sup>107</sup>. The double mutant background with the 141-305 *spn1* mutant allele and the deletion of the *DST1* or *RTF1* genes is lethal when the growth medium is supplemented with caffeine (see Ch.3). Collectively these results were the rationale for performing a series of experiments to determine if Spn1 is involved in the process of aging. The most significant result was a dramatically increased lifespan observed for cells transformed with the 141-305 *spn1* allele.

This Spn1 mutant dramatically increases the lifespan of the BY4741 strain of *Saccharomyces cerevisiae*. Although it is unclear what pathways Spn1 is involved in that contribute to lifespan, a connection to mitochondrial function, the dominance/recessive nature of this aging phenotype, a connection to caloric restriction, and the affect of Spn1 on aging in a separate strain of yeast were explored here.

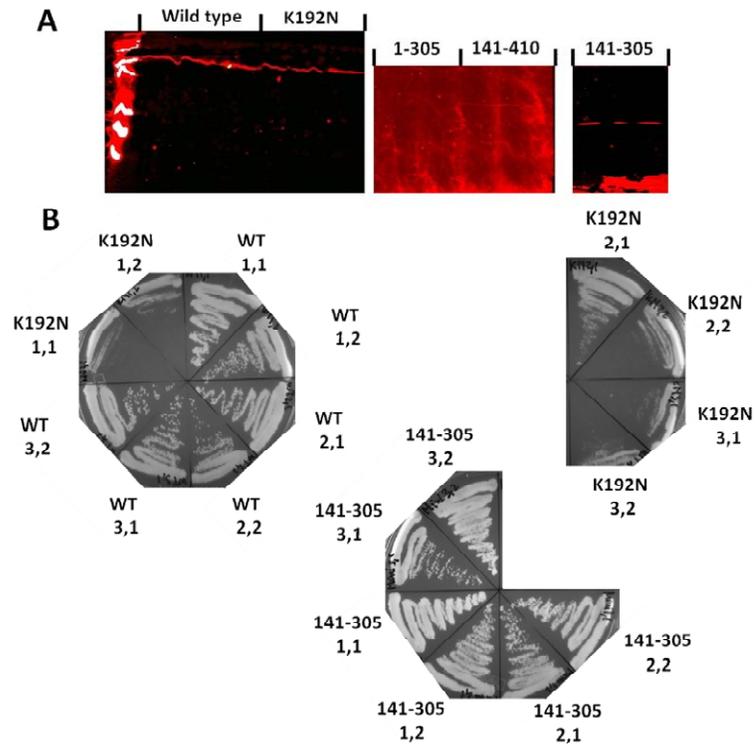
### 4.3 Results

#### 4.3a Removal of the N-and C-terminal regions of Spn1 increase the lifespan of the BY4741 strain of yeast

In the BY4741 genetic background, wild type cells reach 10% viability after 15 days of incubation based on the averages of five biological replicates (Figure 4.3a). The 141-305 *spn1* allele, however, still retains ~100% viability at this time point from the average of four independent biological replicates. Although only two biological replicates have been successfully completed for the 1-305 and 141-410 *spn1* alleles, the results suggest viability between the central domain and wild type Spn1 strains. Removal of the C-terminus resulted in a viability of ~43% and removal of the N-terminus resulted in a viability of ~26% after fifteen days. This indicates that the deletions of both the N-and C-terminal regions synthetically exacerbate the aging phenotype of the deletion of each region individually. This suggests that the increase in lifespan of these Spn1 mutant strains is the result of perturbing the same pathway in each strain. For the K192N mutant *spn1* allele, biological variability has made it difficult to determine whether this mutation has an effect on lifespan. Three biological replicates show viability similar to that of wild type while three other biological replicates show viability between 30% and 70% after fifteen days. The standard deviation for the combined data makes any possible increase in viability not statistically significant (Figure 4.3b). Due to the size difference between the Spn1 mutants, protein extracts were analyzed for many of the biological replicates to ensure they had not been mixed up (Figure 4.4a). To determine if the K192N strains had been mixed up, colonies



**Fig. 4.3 Removal of the N-and C-terminal regions of Spn1 dramatically increases lifespan.** The chronological aging assay was performed on various Spn1 mutants in the BY4741 genetic background as described in the Materials and Methods chapter of this thesis. A) Representation of the average viability of multiple replicates for the wild type (n=5), 141-305 (n=4), 1-305 (n=2), and 141-410 (n=2) Spn1 strains over the course of 19 days. B) Representation of the average viability of multiple replicates for the wild type (n=5) and K192N (n=6) mutant strain over the course of 19 days.



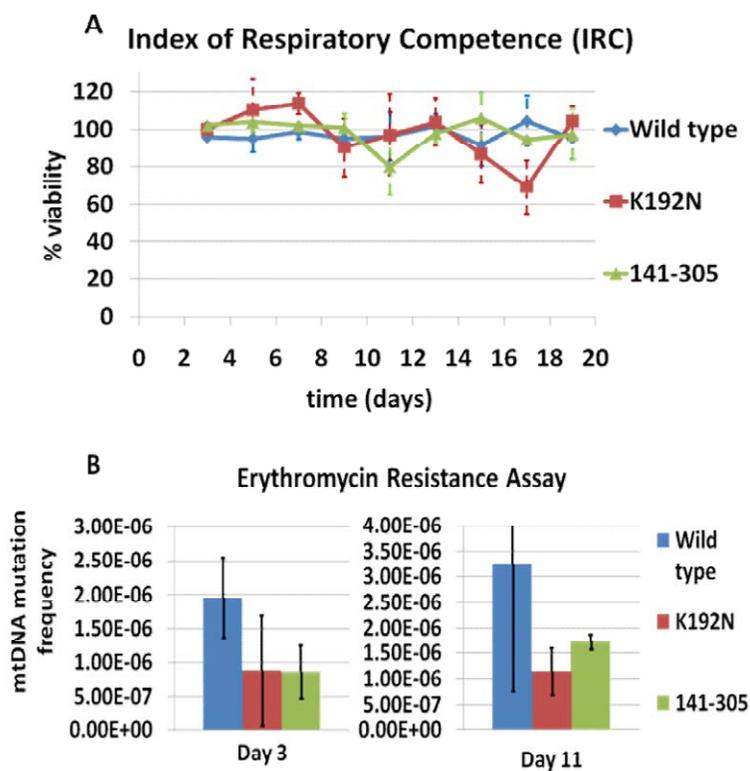
**Fig. 4.4 Aging quality control** For assurance that the aging data obtained was for the correct strains and that none quality control experiments were implemented on all aging data acquired (except for initial biological replicates). **A)** Protein extracts obtained from cells from the quantitative plating assays from each Spn1 mutant strain were analyzed via western blot analysis. **B)** To test whether the wild type and K192N strains had been mixed up, two colonies from the quantitative plating assays were streaked onto YPD medium and grown at 38°C. The K192N strain is temperature sensitive in comparison to wild type cells.

from the quantitative plating assay were streaked onto YPD medium and were incubated at 38°C to test for temperature sensitivity, a published phenotype of this strain<sup>47; 48</sup>(Figure 4.4b). These quality control checks were not implemented immediately and therefore were not used for all of the replicates.

#### **4.3b Spn1 lifespan extension is not due to a change in mitochondrial activity or function**

Two additional assays can be performed with the chronological aging assay called the Index of Respiratory Competence (IRC) and Erythromycin Resistance (ER). These assays test the activity of the mitochondria within the cell throughout the course of the aging assay. The IRC for a culture of cells is the percentage of the viable population capable of growing on non-fermentable carbon sources that require mitochondrial respiration for energy production. To measure IRC, quantitative plating experiments (see Materials and Methods) are performed on YP glycerol plates during a chronological lifespan assay to measure the percent of the population that can grow on this medium in comparison to glucose medium. Yeast can grow on glucose without using mitochondrial respiration and this assay allows for measuring percentage of the population that is deficient for respiration over time. The IRC for the K192N and 141-305 spn1 mutant strains were measured simultaneously with the chronological lifespan. These experiments were performed in biological triplicate and the results suggest that these mutations do not affect the activity of the mitochondria in the cell. All of the strains maintained 100% mitochondrial activity throughout the entire assay (Figure 4.5a). This is highly suggestive that the Spn1 protein's connection to aging does not have to do with respiration.

The ER assay measures the rate at which the mitochondrial genome picks up mutations while cells are aging. Erythromycin is an antibiotic that binds to the 70S subunit of mitochondrial ribosomes<sup>108; 109</sup>. Since a percentage of the components of the electron transport chain are encoded within the mitochondrial genome, this drug will inhibit fungal growth when mitochondrial respiration is required. Specific mutations within the ribosome that prevent this



**Fig. 4.5 IRC and erythromycin resistance assays** Additional assays including the quantitative plating assay were performed to determine if the Spn1 protein's effects on the lifespan of yeast cells is connected to the activity of the mitochondria by measuring the index of respiratory competence (IRC) and erythromycin resistance (ER) of strains encoding the wild type, K192N, or the 141-305 *Spn1* alleles. **A)** IRC results for the wild type, K192N, and 141-305 strains (n=3). **B)** Quantitation of the frequency of mitochondrial DNA mutation after 3 days and 11 days for these same strains (n=3).

interaction allow for growth in the presence of erythromycin<sup>110; 111</sup>. Growth on medium supplemented with a non-fermentable carbon source and erythromycin allows for growth selection of the small percentage of the population that has picked up these mutations. This drug is used as a tool to gauge the overall mutation rate of the mitochondrial genome of a specific strain during chronological aging<sup>105</sup>. To test whether Spn1 affects mitochondrial mutation rate the ER assay was performed in biological triplicate simultaneously with the IRC for the K192N and 141-305 mutant strains. Growth on YPEG medium supplemented with erythromycin was measured on day 3 and day 11 by dividing the # of colonies that grew on this medium by the number of cells that grew on glycerol-based medium on the same days of the IRC assay (see Materials and Methods). A larger amount of variability than was initially expected resulted as is seen in Figure 4.5b. Upon further research on this assay, a high number of biological and technical replicates must be performed to obtain accurate data (other groups did between 20 and 35 biological replicates)<sup>112; 113</sup>. Three biological replicates and three technical replicates were performed here and the data collected is therefore hard to interpret.

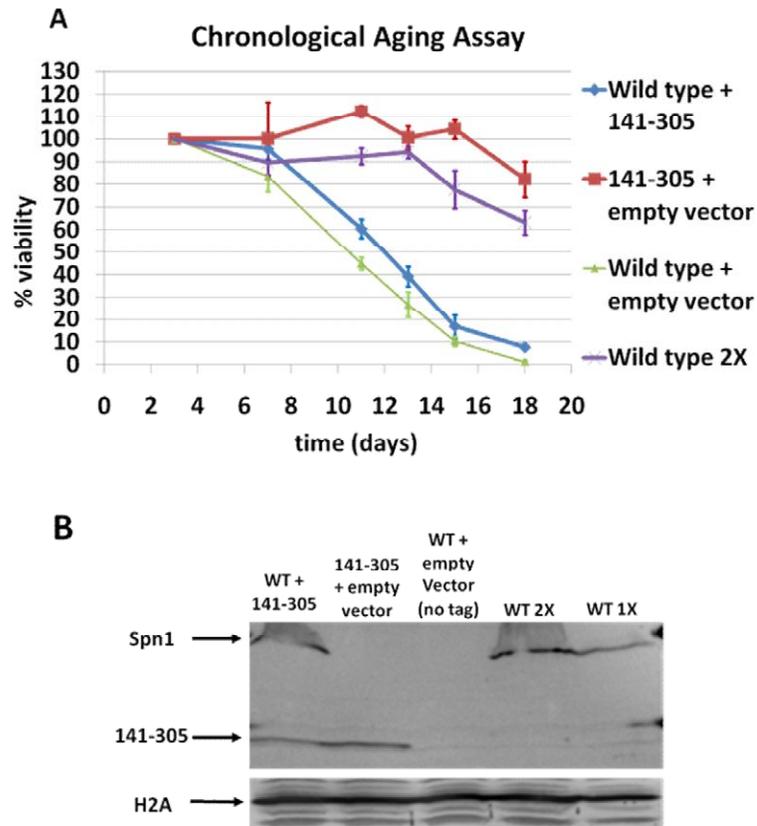
#### **4.3c The extension of lifespan seen for the 141-305 *spn1* allele is likely recessive**

The next question asked regarding the Spn1 protein's role in lifespan determination is whether the phenotype of the central domain is recessive or dominant. If it is dominant then the central domain of Spn1 gains a function/s in the absence of the N-and C-terminal regions and imparts an increased lifespan to cells. If this aging phenotype is recessive, then the N-and C-terminal regions likely have function/s that directly contributes to regulating lifespan. To answer this question experimentally, the same chronological lifespan assay was performed on the wild type and central domain *spn1* allele strains with a few modifications (see Materials and Methods) In brief the central domain (encoded on a *HIS3*-marked plasmid) strain was transformed with the pCR611 plasmid that encodes wild type *SPN1* and a *URA3* selectable marker. Both of these plasmids were maintained by growth in medium lacking histidine and uracil. Three other control

strains were also used: a strain containing two copies of wild type *SPN1* (one in each of the same *HIS3* and *URA3* vectors), the 141-305 strain transformed with an empty *URA3* plasmid, and a wild type strain with *SPN1* encoded on a *URA3* plasmid and transformed with an empty *HIS3* plasmid. Only one biological replicate has been completed, but the results suggest that the central domain aging phenotype is recessive (Figure 4.6a). Unexpectedly, the strain encoding two copies of wild type *SPN1* displays a dramatically increased lifespan. This would suggest that a higher level of *SPN1* expression increases the lifespan of the cell. To measure whether more Spn1 protein is being expressed in this strain, protein extracts were analyzed via western blot analysis. (Figure 4.6b). This western verifies that two copies of wild type *SPN1* result in a higher level of expression. This assay needs to be repeated to verify these results.

#### **4.3d Life-extension phenotype is not from perturbing the same pathways as caloric restriction**

To continue characterizing the Spn1 protein's connection to aging, the same aging experiments described above were performed under caloric restriction conditions to determine if the aging phenotype observed for the 141-305 *spn1* allele or the 1-305 and 141-410 *spn1* alleles are a result of perturbing the same pathways affected by caloric restriction. The performance of these assays proved difficult due to the increased length of time that the strains remain viable. The experiment has been attempted several times and has always resulted in contamination before completion of the assay. A summary of the best data obtained for the wild type, 141-305, 1-305, 141-410 strains performed in biological triplicate is as follows. In this experiment unexplained spikes in viability were observed for at least one of the three replicates for each of the strains tested. It is possible that this is a result of the contamination observed throughout all of cultures, but this has further confounded the use of this data. Data was collected for each strain until contamination was visible within the culture. Wild type cells reach an average viability between 50-60% after 19 days while cells expressing the 141-305 *spn1* mutant still maintain 100% (or



**Fig. 4.6 Lifespan extension observed the 141-305 *spn1* allele is recessive** A variation of the chronological aging assay was performed to determine if the lifespan extension phenotype from the removal of the N-and C-terminal regions of Spn1 is a recessive or dominant. **A)** Representation of preliminary lifespan data (n=1) for strains containing two copies of Spn1 (combinations of wild type and 141-305) or Spn1 and an empty *HIS3* or *URA3* vector. **B)** Results of the western blots performed to ensure that the strains had not been mixed up and to determine whether the strain with two copies of wild type *SPN1* express more of the Spn1 protein than cells with only one copy.

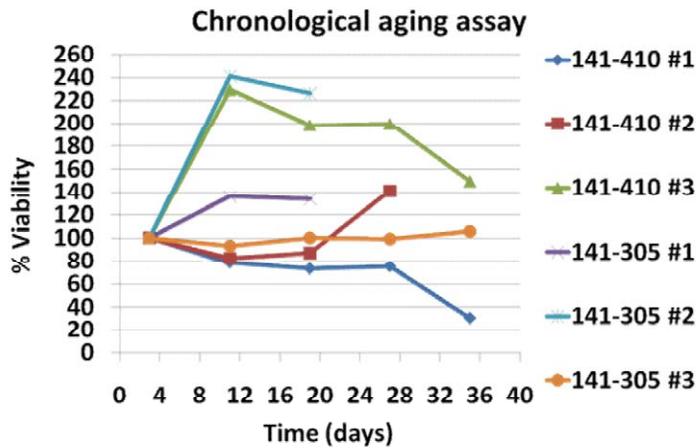
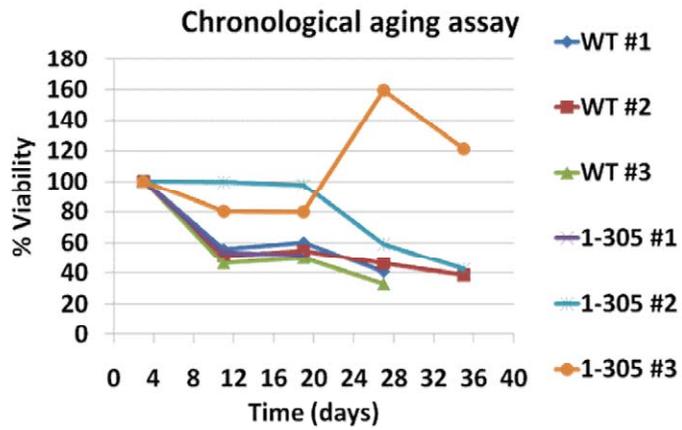
higher) viability at this time point. For the 1-305 *spn1* allele one biological replicate behaved like wild type, one had a viability of 80%, and one replicate maintained 100% viability after 19 days of incubation. For the 141-410 allele, one replicate maintained 100% viability, another replicate spiked above 100%, and the third replicate was down to 75% viability in this amount of time (Figure 4.7). It is currently uncertain what could be done to alleviate the contamination issue.

#### **4.3e Increase in lifespan not seen in the SK1 genetic background**

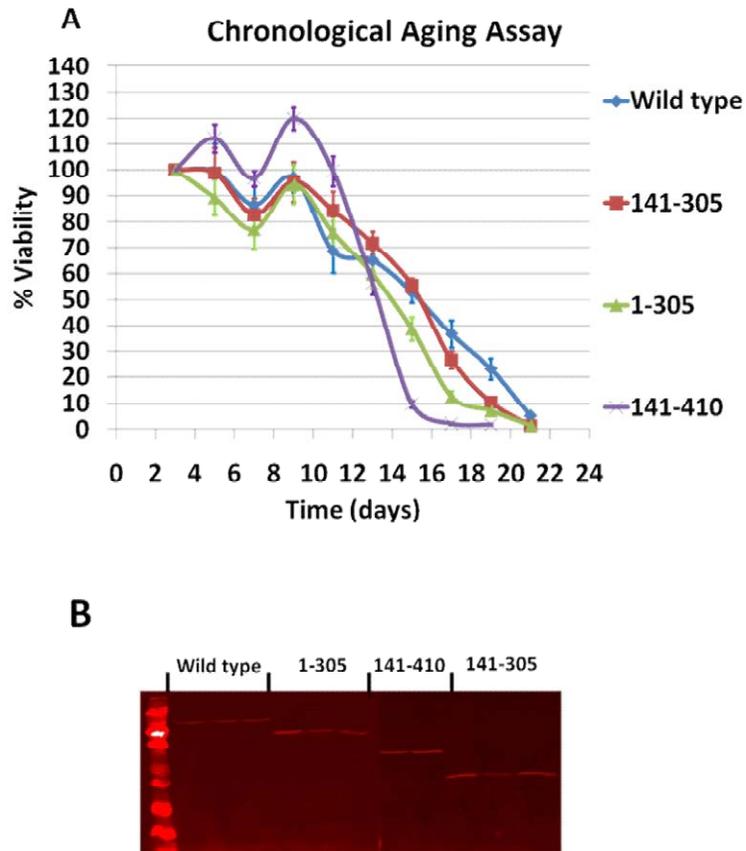
Another set of aging experiments was performed on these same *Spn1* mutants, but in the SK1 genetic background. If the increase in lifespan observed for the 141-305 *spn1* mutant allele is also seen in a different genetic background, then it is more plausible that *Spn1* affects aging in higher eukaryotes. The wild type, 141-305, 1-305, and 141-410 *spn1* SK1 strains were analyzed using the chronological aging assay. The results of these experiments were unexpected and suggest that the increased lifespan is possibly strain-specific for BY4741. All of the mutants analyzed either aged exactly like the wild type strain or had a slight decrease in lifespan (Figure 4.8a). However, these results do not completely rule out that *Spn1* has a conserved role in determining how a cell ages. Other strains should be tested before making this conclusion. Protein extracts were analyzed via western blot to ensure that the replicates had not been mixed up (Figure 4.8b).

#### **4.4 Discussion**

With contributions from ribosome biogenesis and translation, mitochondrial activity and function, autophagy, heterochromatic stability and maintenance of the genome, and even apoptosis, aging affects could arguably be considered the most complicated phenotype to understand in the field of genetics<sup>96; 101</sup>. *Spn1* could be functioning within one or all of these pathways and determining which ones are being directly or indirectly perturbed by the mutations analyzed here is an important future direction of this project. The experiments performed here, however, make a mitochondrial connection unlikely. Based on the results of the IRC experiments,



**Fig. 4.7 Does *Spn1* mutation mimic caloric restriction?** Due to the prevalence of the connection between aging and caloric restriction in the literature, the aging affects of the wild type, 141-305, 1-305, and 141-410 *spn1* alleles were analyzed under caloric restriction. The representation of the best viability data obtained is shown. Due to contamination and unexplained spikes seen in viability, these replicates have not been averaged so that overall trends can be observed.



**Fig. 4.8 The *Spn1* protein's affect on lifespan in the SK1 genetic background** To further validate the evolutionary relevance of the connection between *Spn1* and aging, the same chronological lifespan assay as visualized in Fig. 4.3A was performed using the SK1 genetic background. **A)** Representation of the average viability of multiple replicates for the wild type (n=3), 141-305 (n=3), 1-305 (n=3), and 141-410 (n=2) *spn1* SK1 strains over the course of 19 days. **B)** Verification that the *Spn1* strains had not been mixed up via western blot analysis.

the ability to grow on non-fermentable carbon sources was unchanged for any of the strains tested. Recent high-resolution genomic ChIP-chip experiments suggests that Spn1 is actually present throughout the entire genome of yeast cells<sup>1;2</sup>. With this data, a plausible Spn1 aging model would be that Spn1 contributes to the regulation of a high percentage of genes throughout the genome and therefore controls the total level of protein present within the cell. Over time, the total level of protein requiring recycling would affect the overall molecular homeostasis of the cell and contribute to a change in lifespan. Results reported here suggest that the removal of one or both of the N-and C-terminal regions of Spn1 increases the lifespan of the BY4741 strain of yeast. In this model, these regions of the protein would have a positive effect on overall gene expression. By removing them, the total protein levels would be lower over time and the cell would have less total protein to manage and maintain thus resulting in an increased lifespan. The preliminary data obtained that suggests that the 141-305 *spn1* allele aging phenotype is recessive also supports this model. This experiment suggests that a loss of function of the N-and C-terminal regions of Spn1 increases lifespan by affecting the total level of protein in a cell.

#### **4.4a Future directions**

Although determining whether Spn1 lifespan affects are connected to caloric restriction seems unlikely, a connection to the target of rapamycin (TOR) pathway could be established instead and potentially offer an explanation for how Spn1 is affecting the lifespan of a cell. The TOR pathway is a highly conserved signaling module controlled by the TORC1 kinase protein complex<sup>102</sup>. TORC1 is responsible for altering the expression of the genome in preparation for cellular division in the presence of excess nutrients<sup>114; 115; 116</sup>. Initial interest in the TOR pathway began when results in budding yeast cells with a genomic TOR1 deletion (a subunit of the TORC1 complex) displayed a similar increase in lifespan as is seen for wild type strains under CR conditions. Also, experiments that combined the two did not provide an additive affect regarding lifespan, so it is likely that caloric restriction is working by altering the activity of this

conserved pathway<sup>117</sup>. Characterization of the TOR pathway in lower eukaryotes shows that this complex regulates multiple facets of aging that include ribosome biogenesis and translation, mitochondrial activity, and autophagy. Therefore, many researchers refer to it as a master regulator of aging and the mechanism by which CR increases lifespan in organisms<sup>118</sup>. Future experiments aimed at identifying genetic interactions between Spn1 and TORC1 would further strengthen that this effect is conserved throughout evolution.

Additional strains outside of the BY4741 and SK1 strains also need to be tested. If the pathway/s that Spn1 regulates that determine lifespan are the pathways that are different between the BY4741 and SK1 strains of yeast, then this would explain why a lifespan extension was not observed. Performing a genome-wide microarray could also be done to look at the effects of the central domain of Spn1 on the transcriptome of the cell. This would potentially determine if a large majority of genes are partially down regulated in this strain to support the model that the N- and C-terminal regions of Spn1 positively contribute to the overall protein levels in the cell.

## **Appendix 1: Expression, purification, and characterization of wild type Spn1 and other spn1 mutants**

In this appendix I optimized the expression and purification scheme for the wild type Spn1 protein. I will also describe my efforts to express and purify several Spn1 point mutants including K192N, D172G, K192A, L218P, and the 141-305 + K192N double mutant proteins. I have also included the in vivo characterization of the K192A and 141-305 + K192N mutants. I wrote this appendix, performed all of the experiments and contributed all of the figures.

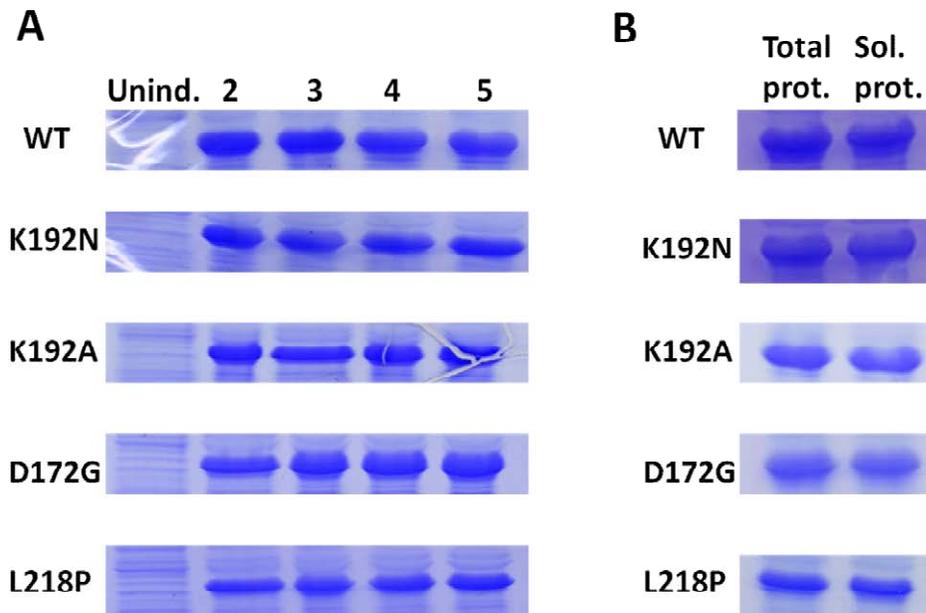
During two of the first three years in the Stargell lab, the main goal was to biophysically characterize the wild type Spn1 protein and various Spn1 point mutant proteins that included: K192N, K192A, D172G, and L218P. With highly purified protein samples, CD spectroscopy, thermomelting, and AUC were to be performed to compare the level of folding and secondary structure present in each protein. Each of these point mutations is located within the central domain of Spn1 and produce genetic and molecular phenotypes in vivo<sup>47; 48; 58</sup>. Measuring how disruptive each of the mutations are to the secondary structure of the protein and learning about how amenable these proteins are in vitro would determine whether future X-ray crystallographic analysis is possible. Visualization of the changes that each mutation imparts on the structure of the central domain will provide a better understanding of how Spn1 functions within the cell. Several hurdles had to be overcome to develop an effective purification scheme for the wild type protein and the purification of the mutant constructs were equally as difficult. Here I will describe the development of the purification scheme for the wild type Spn1 protein in depth and discuss the issues experienced with the mutants. The in vivo characterization of the K192A and 141-305 + K192N spn1 mutant alleles will also be discussed.

### **A1.1 Over-expression and solubility of wild type Spn1 and various point mutants**

The first step in characterizing Spn1 derivatives in vitro is to optimize an over expression scheme to obtain the largest amount of protein possible. The expression of the central domain of

Spn1 has been optimized previously<sup>58</sup>. Large amounts of this protein were obtained by using the Rosetta 2 DE3 pLysS bacterial strain and was a good starting place for the over expression of the Spn1 mutant constructs characterized here (molecular details of this strain can be found at [http://www.emdchemicals.com/life-science-research/rosetta-2de3plyss-competent-cells/EMD\\_BIO-71403/p\\_brGb.s1OagkAAAEjQxI9.zLX](http://www.emdchemicals.com/life-science-research/rosetta-2de3plyss-competent-cells/EMD_BIO-71403/p_brGb.s1OagkAAAEjQxI9.zLX)). Preliminary expression testing on wild type Spn1 suggested that 30°C was a candidate expression condition. The following over-expression time course was performed on the wild type, K192N, K192A, D172G, and L218P Spn1 proteins. Overnight cultures of cells transformed with the appropriate pET-15B + Spn1 ORF (N terminal His tag) plasmid were used to inoculate Luria Broth (LB) medium supplemented with ampicillin (100µg/ml), chloramphenicol (34µg/ml), and glucose (1%). The cultures were grown to an OD600 of 0.4 before being moved to 30°C. The cells were cooled down to this temperature and allowed to grow to an OD600 of 0.6. A 1mL uninduced aliquot was taken, spun down at 4°C and was stored at -80°C before inducing the expression of Spn1 via addition of IPTG to a concentration of 1mM. Additional aliquots were taken at 2, 3, 4, and 5-hour time points after induction. The OD600 at every time point was used to normalize each sample to achieve the same concentration of cells via the addition of an appropriate volume of 1X SDS-PAGE loading dye. The samples were resolved on 10% SDS-PAGE gels (Figure A1a). The results indicate that each Spn1 mutant protein expresses robustly in this cell line under the conditions tested and that a maximum level of expression is achieved after 2 hours.

Follow-up time courses were performed to test whether each mutant is expressed in a soluble form. Ten-mL aliquots were harvested 2 hours after the addition of IPTG. Cell pellets were resuspended in 4mL of ice-cold lysis buffer (20mM Tris pH 7.5, 200mM NaCl, 10% glycerol, 500µM PMSF) and were sonicated for four one minute cycles using 30% duty and an output level of 2. Fifty µl of whole lysate from each sample were combined with 50µl of 2X SDS-PAGE loading dye for a measure of the total protein present. One mL of the samples was



**Fig. A1 A)** Results of an over-expression timecourse of the wild type (WT), K192N, K192A, D172G, and L218P Spn1 proteins at 2,3,4, and 5-hours. One mL of cells were spun down at the indicated time points and were resuspended in an appropriate volume of SDS-PAGE loading dye to normalize the number of cells present between all of the sample. Fifteen  $\mu$ l were resolved on a 10% SDS-PAGE gel and were visualized using coomassie brilliant blue stain. **B)** Visualization of the amount of protein that is soluble after 2 hours of induction for each of the Spn1 proteins tested. Ten mL of cells were spun down, resuspended in ice cold lysis buffer, and lysed using sonication. The insoluble cellular debris was pelleted and 15 $\mu$ l of the samples before (total protein) and after (soluble protein) spinning were resolved on a 10% SDS-PAGE gel to measure the amount of soluble Spn1 protein present.

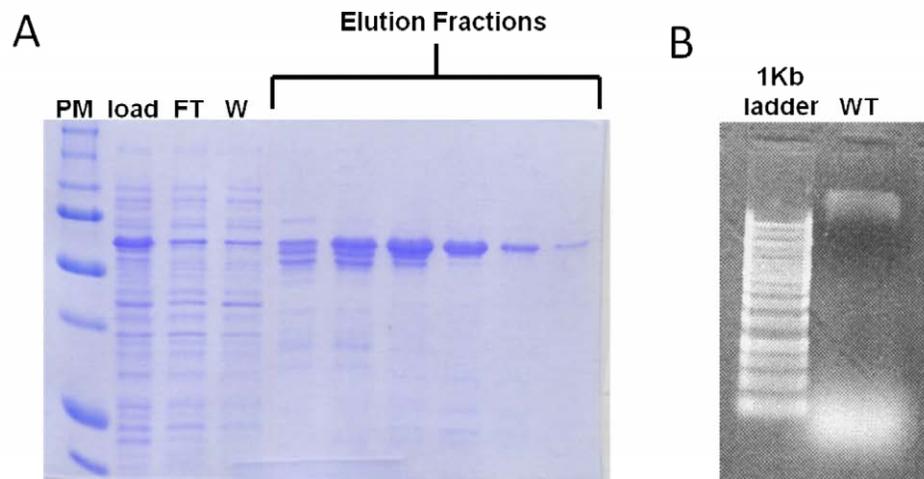
then spun at max speed in a table-top centrifuge at 4°C for 20 minutes to pellet all cellular debris and insoluble/aggregated protein. Fifty µl aliquots were combined with 50µl 2X SDS-PAGE dye. Both the lysate and soluble extract samples were resolved on 10% SDS-PAGE gels (Figure A1b). The results show that the Spn1 protein expressed for each of the mutants is soluble under these buffer conditions.

### **A1.2 Optimization of the purification scheme for wild type Spn1**

The expression of these Spn1 mutants was scaled up to 1L and initial attempts were made to purify them using a GE Hi Trap Ni<sup>2+</sup> chelating affinity column attached to an AKTA FPLC purification system. Elution of the Spn1 proteins using a gradient of imidazole produced highly pure protein via SDS-PAGE analysis, but, the 260:280 UV spectrophotometer ratios ranged from 1.8-2.0. The presence of nucleic acid contamination was verified by visualization on an agarose gel (Figures A2a and A2b). Since sonication was used to lyse the cells, nucleic acid larger than 20,000bp would likely have been sheared. The results of this gel suggest that Spn1 could be forming a complex with a nucleic acid contaminant and that a large amount of free nucleic acid is also being eluted off the column. Visualization of the flow-through fractions suggests that a large amount of Spn1 protein is also not binding to the column. Since X-ray crystallography is one of the downstream experiments for the wild type and mutant proteins, work was done to address these issues and improve the yield of this first Ni<sup>2+</sup>-column capture step.

After ensuring that the column itself or that the Ni<sup>2+</sup> used to charge the column was not causing the low binding efficiency of Spn1, adjustments were made to the buffers being used (20mM Tris-HCl pH7.5, 200mM NaCl, and 10% glycerol). Per GE Healthcare's instructions, the Tris-HCl was replaced with sodium phosphate. According to GE's research, phosphate provides the highest level of binding for proteins and is the preferred buffer of choice for chelating columns

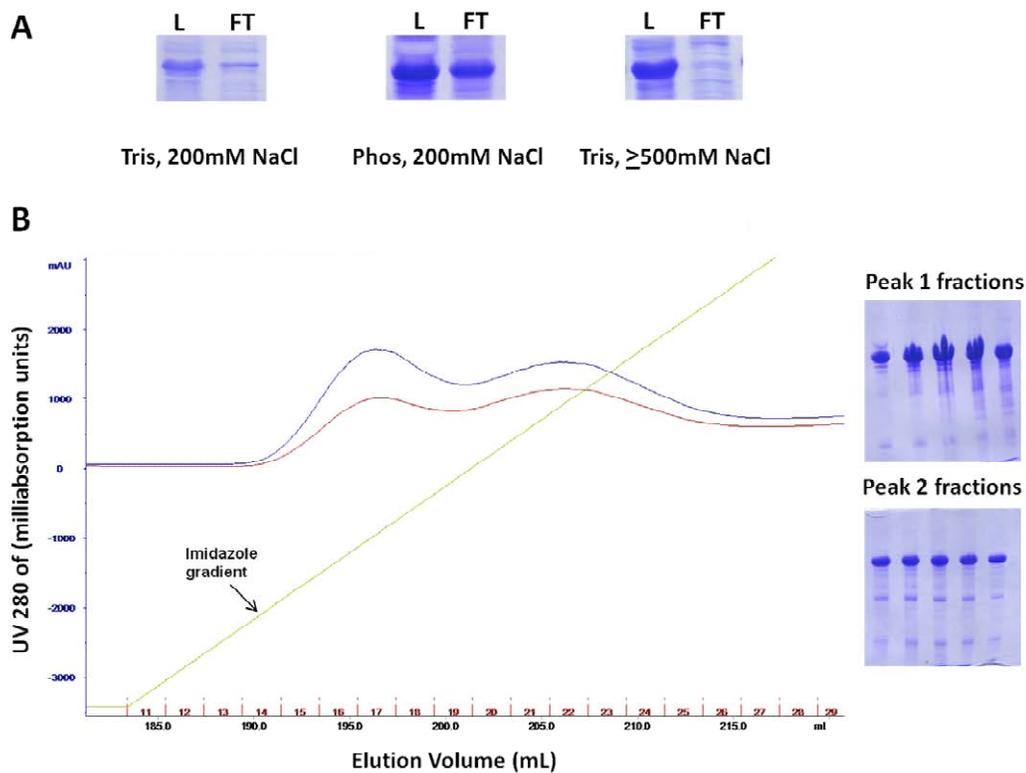
<http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/CF50C877B8D54242C1257628001>



**Fig. A2** Visualization of the initial attempt to purify the wild type (WT) Spn1 protein using a  $\text{Ni}^{2+}$  affinity column. **A)** The load, flow-through (FT), 20mM imidazole wash step (W), and individual elution fractions from a gradient of imidazole up to 500mM were resolved on a 10% SDS-PAGE gel. The coomassie stain of this gel showed that although decently pure protein was obtained, a large amount of Spn1 protein did not bind to the column as is shown in the FT fraction. **B)** Visualization of the contaminating nucleic acid on an agarose gel in the elution fractions from the  $\text{Ni}^{2+}$  column purification of full length wild type Spn1. A high amount of small nucleic acid is present although a band above the ladder suggests a much larger species is also present.

[CBC9D/\\$file/71700500\\_AY\\_WEB.pdf](#)). For Spn1 it must be noted that this is not the case. The use of phosphate buffer causes Spn1 to elute off the column at a lower concentration of imidazole and causes a higher percentage of contaminants to be present in the elution fractions (Figure A3a). Instead, a possible explanation for the low affinity of Spn1 would be that free nucleic acid is competing with Spn1 for the binding sites on the column. Further literature research revealed that nucleic acid can form high-affinity coordination complexes with divalent ions and that this chemistry can actually be utilized to purify both DNA and RNA molecules<sup>119; 120</sup>. After a series of DNase and RNase treatments were unsuccessful at getting rid of the contamination, the ionic conditions of the buffers were adjusted. GE healthcare recommends using anywhere from 0.5M salt to 1M salt to reduce the binding of contaminants with a Ni<sup>2+</sup> affinity column. Higher salt buffers with 500mM, 750mM, and 1M NaCl were subsequently tested and improved the results of the purification. The level of Spn1 binding was dramatically increased based on the SDS-PAGE visualization of the flow through fractions (Figure A3a). Also instead of having one large and broad peak with a high 260:280 ratio, two peaks eluted from the gradient of imidazole that both have a much lower 260:280 ratio (Figure A3b). No differences are evident between the two Spn1 peaks on an SDS-PAGE gel. Due to the slight difference in the 260:280 ratios, it is possible that the first peak is free Spn1 while the second one is a complex of Spn1 and a small nucleic acid contaminant. The last parameter that was optimized was the addition of 50mM imidazole to the buffers. This provided an additional level of purity to the final product.

Size-exclusion chromatography (SEC) was utilized to analyze the differences between the two peaks that elute off the Ni<sup>2+</sup> column. Each peak was pooled separately and concentrated to a ~1mL volume. The two peaks partially overlap each other and it is likely that each pool contains a small percentage of the other peak. Each concentrated pool was loaded onto a 120mL 16/60 Superdex200 SEC column to roughly determine the size differences of the contents. The first Ni<sup>2+</sup> peak with a 260:280 ratio of about 0.5-0.6 eluted from the column in two places, after



**Fig. A3** Visualization of the binding behavior of wild type Spn1 to a  $\text{Ni}^{2+}$ -affinity chelating column under different buffer conditions **A)** The load (L) or total protein put over the column and the flow-through (FT) were resolved via SDS-PAGE and visualized with coomassie brilliant blue dye. The results indicate that Tris buffer with a high amounts of salt allow for the highest percentage of Spn1 to bind to the resin. **B)** Chromatogram of the results of this purification when high salt is used. The y-axis is the UV 280nm absorbance in milliabsorbance units (blue) as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) is also in milliabsorbance units and the green line is the concentration gradient of imidazole used to elute Spn1 from the column. The SDS-PAGE visualization of the two peaks indicates no differences between the protein present in either one.

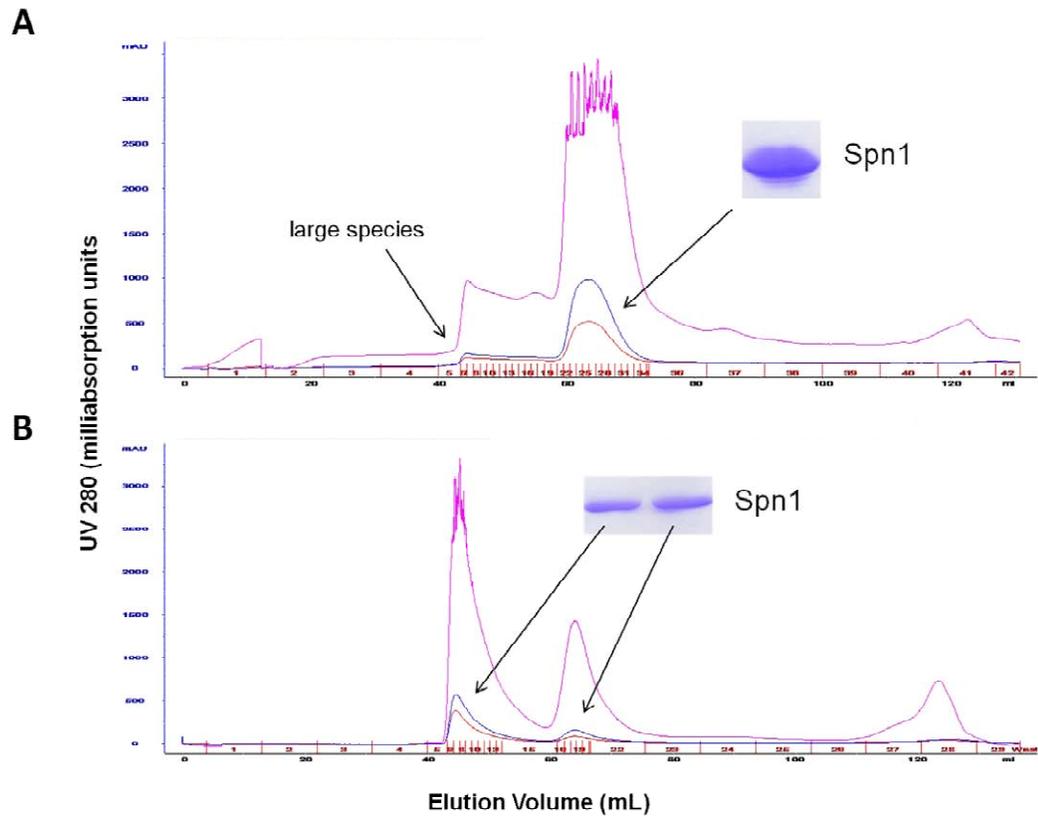
~40 mL (void volume per GE's experience,

[http://www.gelifesciences.com/aptrix/upp00919.nsf/content/E2B05C0463CE9383C1256EB400483E12?OpenDocument&Path=Catalog&Hometitle=Catalog&entry=5&newrel&LinkParent=C1256FC4003AED40-319068987EDF3042C1257019004908BA\\_RelatedLinksNew-C821BEC677D8448BC1256EAE002E3030&newrel&hideseachbox=yes&moduleid=165424](http://www.gelifesciences.com/aptrix/upp00919.nsf/content/E2B05C0463CE9383C1256EB400483E12?OpenDocument&Path=Catalog&Hometitle=Catalog&entry=5&newrel&LinkParent=C1256FC4003AED40-319068987EDF3042C1257019004908BA_RelatedLinksNew-C821BEC677D8448BC1256EAE002E3030&newrel&hideseachbox=yes&moduleid=165424))

and after 60mL (dominant peak) of buffer flowed over the column (Figure A4a). It is possible that the contents of the 40mL peak are the spillover of the 2nd Ni<sup>2+</sup> peak. The analysis of the second Ni<sup>2+</sup> peak on this column supports this hypothesis because it resulted in two peaks at the same volumes, however, this time the 40mL was dominant and the 60mL peak was much smaller (Figure A4b). SDS-PAGE analysis of these two runs verified that both 40 and 60mL peaks contained wild type Spn1 protein. The conclusion is that the first peak from the Ni<sup>2+</sup> purification is what contains functional Spn1 and that the second peak is likely a result of aggregation or a nucleic acid complex. On average, globular proteins about 158 kDa in size elute from this type of SEC column after 60mL. Wild type Spn1 with an N-terminal His tag is ~48kDa. A globular 48kDa protein would elute off of this column after about 80mL

([http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/8F1B26694211F95FC1257628001D2FA4/\\$file/28407384AA.pdf](http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/8F1B26694211F95FC1257628001D2FA4/$file/28407384AA.pdf)). This suggests one of three things about Spn1: either the protein isolated up to this point is not properly folded and is causing it to separate from the resin earlier than usual, a large part of the protein is intrinsically disordered thus increasing the frictional coefficient and apparent size of the protein in solution, or the Spn1 species is multimeric.

Although the biophysical nature of the N-and C-terminal regions flanking the central domain were unknown at this time, fold index plot analysis predicted that these regions are disordered (see Ch. 3 discussion). Further analysis of this protein (as described in Ch. 3) supports this. Since a SEC step separates the possible Spn1/nucleic acid complex and provides an additional level of



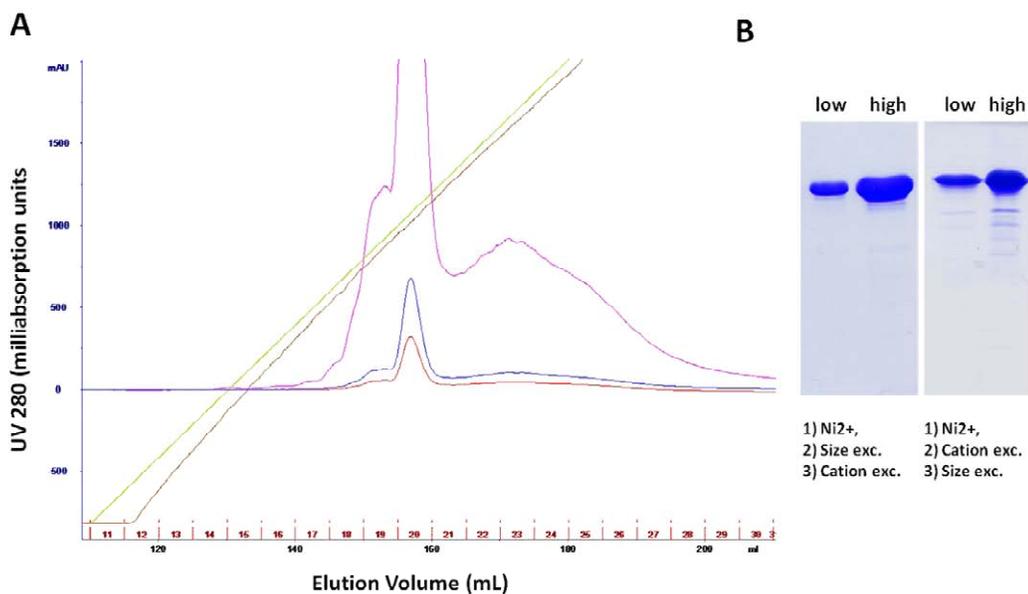
**Fig. A4 A)** Chromatographic results of the purification of the first  $\text{Ni}^{2+}$ -column peak for the wild type protein on a 16/60 Superdex200 column. The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed through the column in mL (x-axis). The UV 260nm (red) and the UV 230nm (pink) are also in milliabsorbance units. **B)** Result of the purification of the second  $\text{Ni}^{2+}$  peak for the full length wild type Spn1 protein. Each peak was visualized using SDS-PAGE. The combined results of these chromatograms indicate that there is a dramatic size difference between the Spn1 species in the two samples.

purity after the initial Ni<sup>2+</sup> column capture, it is a useful second purification step for wild type Spn1.

Wild type Spn1 is prone to small amounts of degradation, evident from the lower bands present on an SDS-PAGE gel even after SEC. Although this might not affect the biochemical activity of the protein, obtaining an accurate calculation of the amount of secondary structure using CD spectroscopy is more difficult with a population of partially degraded molecules in the sample. Ion-exchange chromatography was considered to achieve the final high level of purity required for this experiment. Since the theoretical pI of full length Spn1 (including the His tag) is 7.8 (<http://expasy.org/tools/protparam.html>), either an anion-or cation-exchange column could be used. Using a buffer at pH8.5, Spn1 was loaded onto a Hi Trap Q anion-exchange column at ~85mM NaCl and was eluted in a gradient up to 1M NaCl. Spn1 does not have a high affinity for this column and began eluting between ~100 and 125mM NaCl (data not shown). Using a buffer at pH6.5, Spn1 was also tested on a Hi Trap SP cation-exchange column. The protein was also loaded onto the column at about 85mM NaCl and was eluted with a gradient of NaCl up to 1M. Spn1 has a much higher affinity for this column and does not start to elute until ~300-350mM NaCl (Figure A5a). The SDS-PAGE visualization of the pooled fractions from the peak of this purification showed that highly pure full length Spn1 protein had been obtained. It is important to note that this overall purification scheme was attempted with the size-exclusion and cation-exchange steps reversed and resulted in a lower level of purity (Figure A5b).

### **A1.3 Purification of the K192N spn1 mutant**

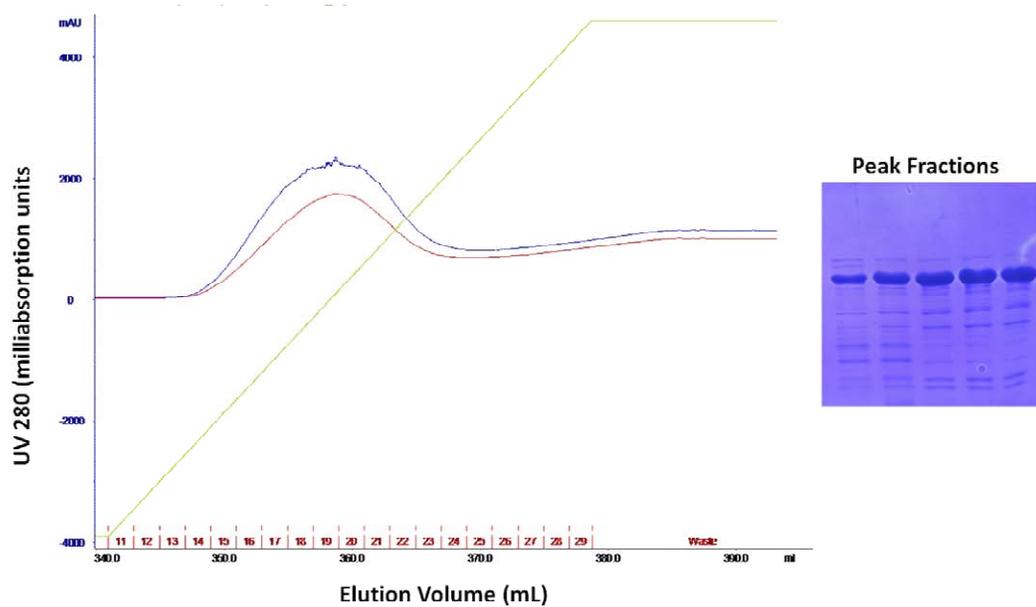
Having determined an effective purification scheme for the wild type Spn1 protein, a purification scheme for the K192N, K192A, D172G, and L218P point mutants was required for the biophysical characterization of these proteins. Due to the difficulties that had to be overcome for wild type Spn1, only the K192N mutant was tested first using the same purification scheme developed above. The results of the first Ni<sup>2+</sup>-affinity capture step were different from that



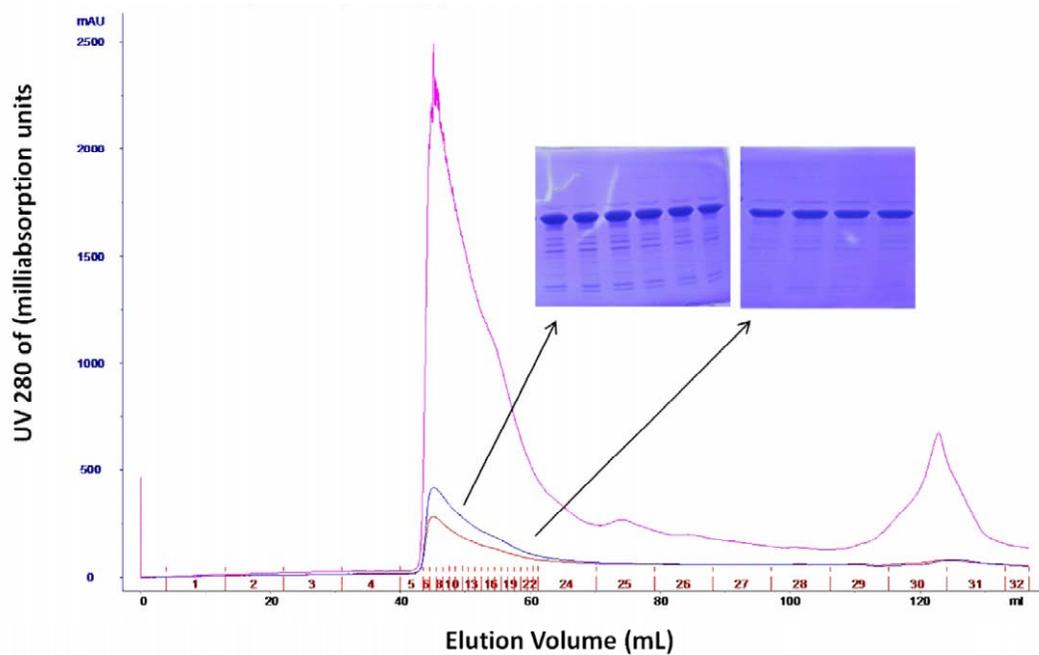
**Fig. A5** Chromatographic results of the final SP cation-exchange step utilized to get rid of the degraded Spn1 present in the sample and obtain a very high level of purity for biophysical analysis. **A)** The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed through the column in mL (x-axis). The UV 260nm (red) and the UV 230nm (pink) are also in milliabsorbance units. The green and brown lines represent the concentration gradient of salt used to elute Spn1 from the column and the conductivity respectively. **B)** SDS-PAGE comparison of the final Spn1 product obtained when the order of the size-exclusion and cation-exchange steps are reversed. Low represents a small amount of protein loaded while high represents a larger amount of sample loaded.

observed for the wild type Spn1 protein. Instead of two peaks eluting with the gradient of imidazole, only one broad peak came off (Figure A6). The fractions from the Ni<sup>2+</sup> peak were pooled and concentrated to ~1mL for a size-exclusion step (120mL 16/60 Superdex200 column using 25mM MES pH6.5, 200mM NaCl, 10% glycerol). Almost all of the K192N protein came out after ~40mL, which again is the void volume for this column and only a very small protein peak came off after ~60mL as is observed for wild type Spn1 (Figure A7). When the peak is visualized using SDS-PAGE, the presence of lower molecular weight contaminants that should separate from Spn1 using this column, suggests that soluble aggregates have formed. This aggregation is likely induced by the concentration step that is performed to prepare the sample for this column. To determine if the K192N protein will enter the column if loaded at a lower concentration, 1mL of the pooled Ni<sup>2+</sup> fractions were loaded onto the column without a concentration step, using a higher ionic strength buffer (20mM Tris pH7.5, 1M NaCl, 10% glycerol). Although the peak sizes were much smaller, the majority of the protein still came out of the column starting at about 40mL.

One possibility is that the presence of the contaminating proteins seen in the SDS-PAGE gel in Figure A7 are inducing soluble multimers to form and that a different purification step that would remove them could allow for the K192N protein to be concentrated for size exclusion. To test this hypothesis, the fractions from the Ni<sup>2+</sup> column were dialyzed into 25mM MES buffer (pH 6.5), 150mM NaCl, 10% glycerol to prepare the sample for a cation-exchange step. The K192N protein proved to be difficult to work with and precipitated out of solution readily during dialysis. Different variations of dialysis and straight dilution were attempted to prepare the K192N protein for cation-exchange, but only one method did not result in precipitated protein. The pooled Ni<sup>2+</sup> fractions were diluted 2 fold before dialysis and the dialysis buffer was adjusted to 400mM NaCl. Prior to loading onto the SP cation-exchange column, the dialyzed sample was further diluted to cut the NaCl concentration down to ~200mM. The sample was subsequently



**Fig. A6** Chromatographic results of the  $\text{Ni}^{2+}$  purification of the K192N spn1 mutant protein. The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) is also in milliabsorbance units and the green line is the concentration gradient of imidazole used to elute Spn1 from the column. The SDS-PAGE visualization of the peak fraction indicate a fair amount of contaminants are still present.

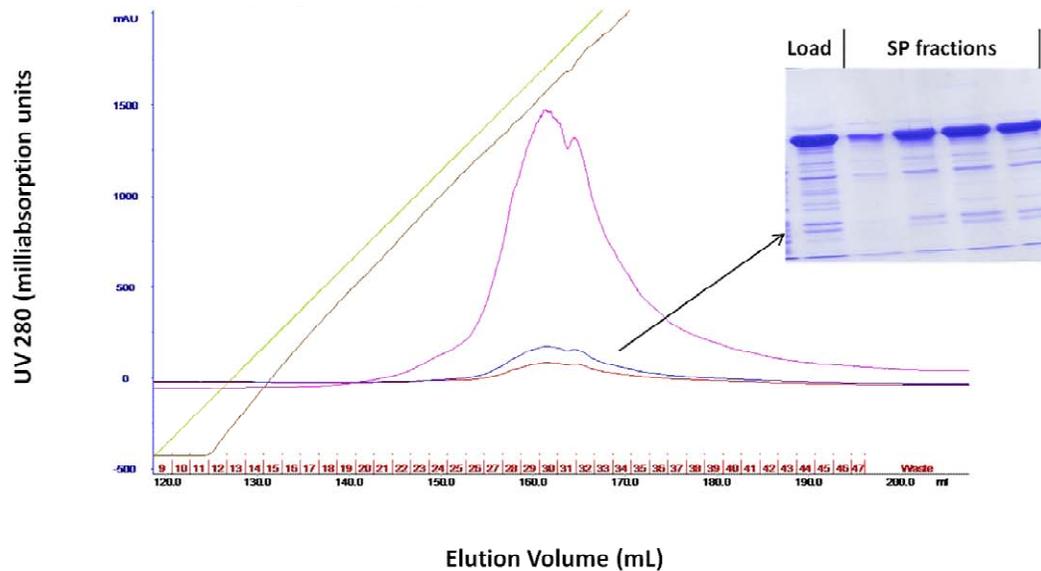


**Fig. A7** Chromatographic results of the S200 SEC purification of the K192N spn1 mutant protein after concentration of the Ni<sup>2+</sup>-column pool to ~1mL. The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) and the UV 230 nm (pink) is also in milliabsorbance units. The peak was resolved on a 10% SDS-PAGE gel and was visualized with coomassie stain.

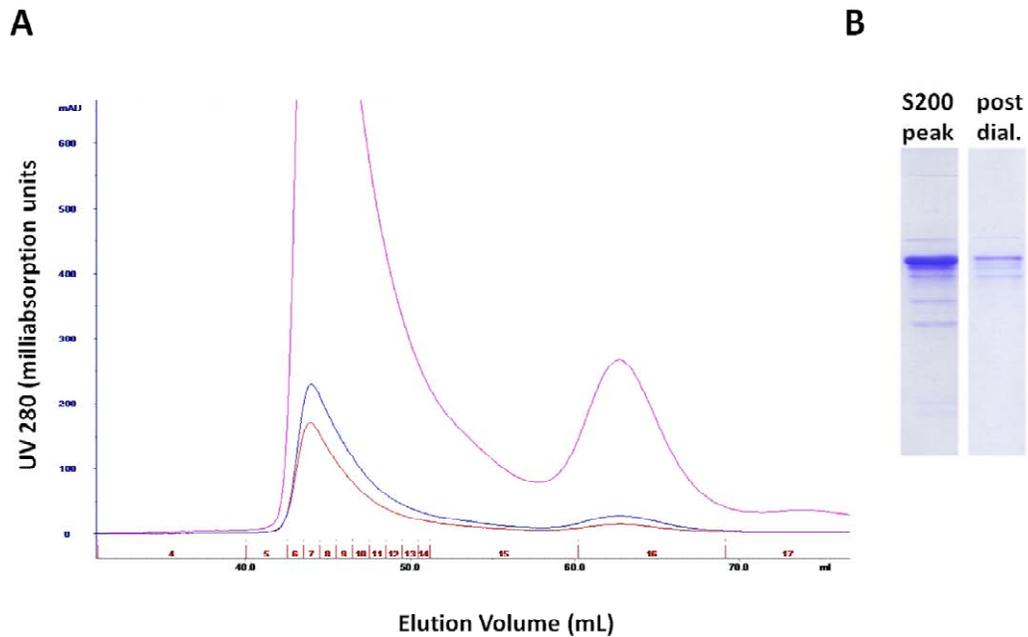
loaded onto a Hi Trap SP cation-exchange column and eluted using a gradient of NaCl (Figure A8). The peak fractions were visualized by SDS-PAGE. The results indicate that the amount of contaminants were decreased. The fractions were pooled and concentrated down about two fold from 12mL to 6mL for SEC. Despite the use of the cation-exchange step, the results were the same as before. The majority of the protein came out in the void volume and only a small percentage came out after 60mL (Figure A9a). Since CD Spectroscopy does not require very much sample, several SEC steps were performed and the 60mL elution peaks were all pooled and dialyzed into CD buffer (see Materials and Methods). However, the K192N protein is not stable in this buffer and crashed out before the experiment could be completed (Figure A9b).

#### **A1.4 Purification and in vivo characterization of the 141-305 + K192N double *spn1* mutant**

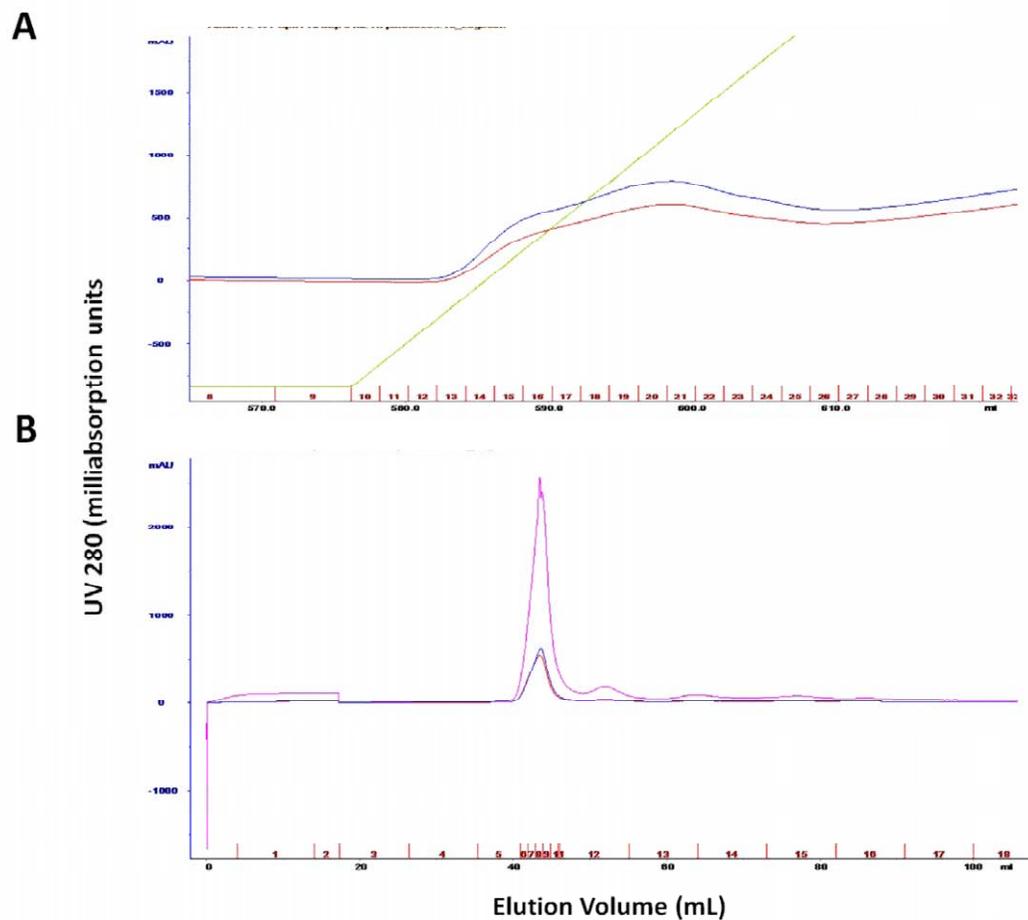
The 141-305 + K192N double *spn1* mutant has not been characterized in vitro or in vivo. Even though the full-length K192N protein is unstable and difficult to work with in vitro (described previously), it is possible that the removal of the N-and C-terminal regions flanking the central domain could help with this. To clone the DNA sequence encoding this double *spn1* mutant into a pET-15B expression vector, Dr. Cathy Radebaugh used point mutagenesis to mutate the K192 residue within the expression vector containing the 141-305 *spn1* allele<sup>58</sup>. One liter of the 141-305 + K192N protein was expressed identically to how the 141-305 protein is expressed (Rosetta 2 DE3 pLysS cell line, induced with 1mM IPTG, expressed at 16°C overnight). The pellet was resuspended in the same buffer and lysed under the same conditions as was described previously for wild type Spn1. After pelleting the cellular debris, the soluble extract was loaded onto a 5mL Hi Trap chelating column and eluted using a linear gradient of imidazole from 50 to 500mM (Figure A10a). The peak fractions were pooled and half of the total volume was concentrated. Almost immediately visible aggregates formed before the sample had been concentrated two-fold. After pelleting the aggregates, 2mL of the supernatant were loaded



**Fig. A8** Chromatographic results of the SP cation-exchange purification of the K192N spn1 mutant protein after dialysis of a  $\text{Ni}^{2+}$  column purification pool. The y-axis is the UV 280nm absorbance in milliabsorbance units as a function of the volume of buffer that has flowed through the column in mL (x-axis). The UV 260nm (red) and the UV 230 nm (pink) is also in milliabsorbance units. The green and brown lines represent the concentration gradient of salt used to elute Spn1 from the column and the conductivity respectively. The long trailing peak suggests a heterogeneous population of protein species. The peak was resolved on a 10% SDS-PAGE gel and was visualized with coomassie stain. The results of the gel would indicate that aggregation between K192N and other contaminants is occurring, since this column provides a higher level of purity to the full length wild type Spn1 protein when used.



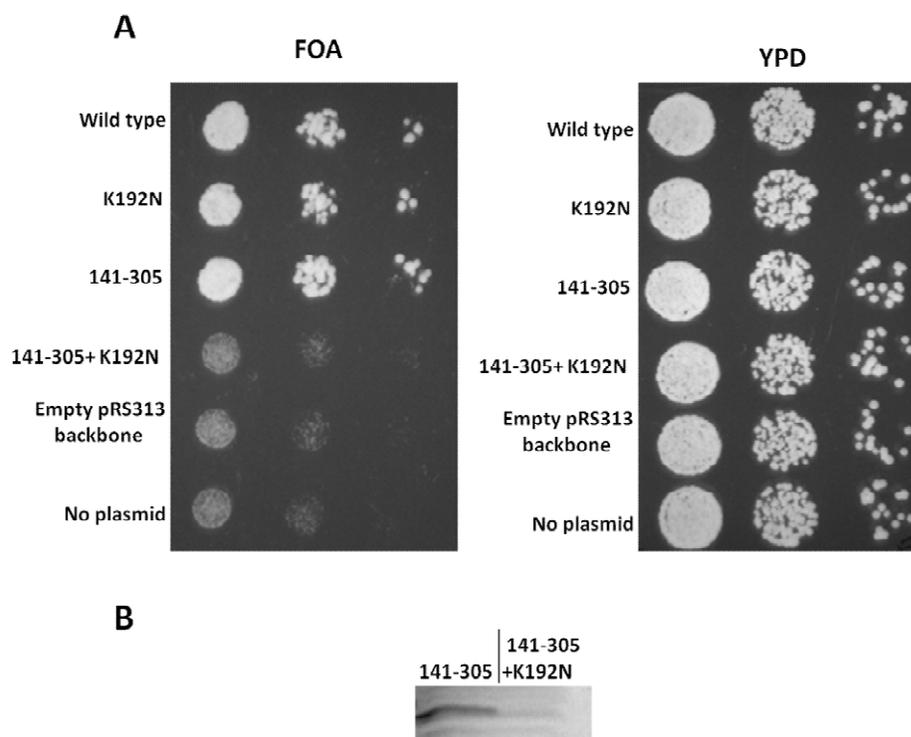
**Fig. A9** Chromatographic results of the S200 SEC purification of the full length K192N spn1 mutant protein after an SP cation exchange step had been performed. **A)** The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) and the UV 230 nm (pink) are also in milliabsorbance units. Despite the higher level of purity that had been obtained by performing a cation-exchange step before the SEC step, the majority of the K192N protein still ended up in the void volume. **B)** SDS-PAGE results of visualizing the K192N protein before and after dialysis into 20mM sodium phosphate (pH7.4) + 200mM NaF buffer for circular dichroism spectroscopy. It is likely that K192N is unstable in this buffer and mostly crashed out during dialysis.



**Fig A10** Chromatographic results of the  $\text{Ni}^{2+}$  purification and the S200 SEC purification of the 141-305 + K192N double spn1 mutant protein. **A)** The y-axis is the UV 280nm absorbance (blue) in milliabsorbance units as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) is also in milliabsorbance units and the green line is the concentration gradient of imidazole used to elute Spn1 from the column. **B)** The y-axis is the UV 280nm absorbance in milliabsorbance units as a function of the volume of buffer that has flowed over the column in mL (x-axis). The UV 260nm (red) and the UV 230 nm (pink) are also in milliabsorbance units.

onto a 120mL 16/60 Superdex200 column (used 20mM Tris pH7.5, 500mM NaCl, 10% glycerol buffer). The only peak that came off the column was after 40mL of buffer had run through it, which is the void volume (Figure A10b). Whatever protein that was left was likely aggregated or unfolded. In order to determine if this was the result of the concentration step performed, the same purification step using 2mL of the Ni<sup>2+</sup> fractions that were not concentrated was performed. The results, however, were the same. No further work has been done with this protein in vitro.

The physiological consequences of this double mutant were also tested in vivo. The ORF encoding this double mutant was cloned into a pRS313 plasmid backbone (*HIS3*, *CEN*) to create the pAA345 plasmid. This was transformed into the BY4741 genetic background (described by Zhang and colleagues)<sup>48</sup>. In brief, this process begins with knocking out the genomic copy of *Spn1* using homologous recombination with a *LEU2* gene fragment flanked by the *SPN1* promoter and terminator. A covering plasmid (*URA3*, *CEN*) with the *SPN1* ORF flanked by the *TOA1* promoter and terminator is required at this step since *SPN1* is essential. From here, any *SPN1* allele can be transformed using 5-fluoroorotic acid (FOA) plasmid shuffling<sup>63</sup>. During this shuffling process, it was determined that the 141-305 + K192N *spn1* double mutant is not capable of covering the genomic knockout of wild type *SPN1*. When transformed with the pAA345 plasmid, the strain was not able to grow on FOA supplemented medium (Figure A11a). To test whether this inability to cover the genomic knockout is due to unstable protein expression in vivo, the pAA345 plasmid was transformed into the BY4741 strain and was cultured in growth medium to select for cells that maintain both the wild type and 141-305 + K192N plasmids (-His, -Ura medium). Protein extracts were analyzed via western blot and *Spn1* was probed using an anti-*Spn1* polyclonal antibody (Figure A11b). The conclusion is that the *Spn1* double mutant is not stably expressed.



**Fig. A11** Results of the spot test to show that the 141-305 + K192N double mutant *spn1* allele cannot cover the genomic knockout of *Spn1*. The indicated *Spn1* ORFs were all transformed into the BY4741  $\Delta$ *SPN1* strain with a *URA3*-marked *Spn1* covering plasmid and were grown on FOA medium in a spot test plating assay. **A)** Only cells that have kicked out the *URA3* covering plasmid can grow on FOA medium. **B)** Results of the western blot to analyze whether the 141-305 + K192N protein can be stably expressed *in vivo*.

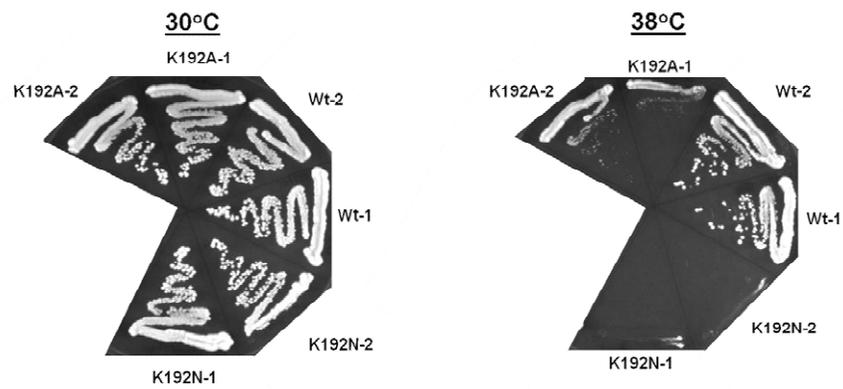
## **A1.5 The K192A *spn1* point mutant is temperature-sensitive in vivo in the BY4741 genetic background**

Previously the K192A *spn1* point mutation had been characterized as not being a temperature-sensitive allele in the SK1 genetic background. This result was difficult to interpret considering that the K192N allele is temperature-sensitive and that the differences between lysine and asparagine are not as dramatic as the differences between lysine and alanine. Therefore this same phenotypic characterization was carried out on the K192A mutation within the BY4741 genetic background to verify this result. After cloning the K192A sequence into the appropriate vector, it was transformed into this strain using FOA shuffling and two colonies were streaked onto YPD medium and were incubated at 38°C. The K192N strain was used as a control for this experiment. The results suggest that although the K192A point mutation is not temperature-sensitive lethal, cells encoding this allele exhibit a slow-growth phenotype under these growth conditions (Figure A12).

### **Materials and Methods**

#### **Plasmids and cloning:**

The *Spn1* ORFs encoding the following *spn1* mutant alleles were all isolated from the pCRK192A, pCRD172G, and pCRL218P plasmids and were subcloned into the pET-15B expression vector backbone: K192A, D172G, and L218P respectively. This was accomplished by performing an *NdeI* and *BamHI* double-restriction enzyme digest. These restriction enzymes were also used to linearize the pET-15B vector. Following agarose gel purification, the *Spn1* gene sequences were subsequently ligated into the plasmid backbone using NEB's T4 DNA ligase. For the in vivo experiment to study full length K192A, the wild type promoter, K192A ORF, and wild type terminator were amplified using PCR on the pJF plasmid using primers with the following sequences:



**Fig. A12** Streaks were performed on cells encoding wild type (WT), K192N, or K192A *SPN1* alleles and were incubated at 38°C to determine if the K192A mutation is temperature-sensitive. Two colonies for each strain were analyzed.

5'TTACATATGAGTACAGCCGATCAAG3' (introduces *NdeI* site to N-terminus of Spn1 ORF in frame) 5'CCTGGATCCTTATTATTTATGCTTCTTGTTTAATC3' (introduces *BamHI* site into C-terminus of Spn1 ORF). Following this PCR reaction, the amplified products were digested with the *PflmI* and *BamHI* restriction enzymes as well as the pCR311 plasmid (*HIS3*, CEN) that contained the wild type *SPN1* gene. Following agarose gel purification, the K192A *spn1* gene sequence was ligated into the plasmid backbone using NEB's T4 DNA ligase. This plasmid was transformed into the SK1 genetic background using an FOA shuffling method<sup>47</sup>.

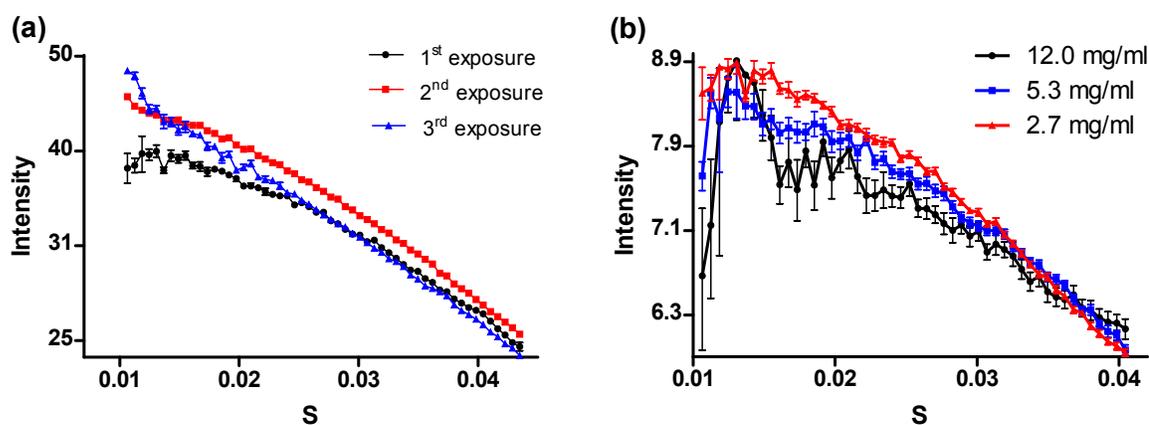
## **Appendix 2: Wild type and 141-305 Spn1 proteins are not good candidates for small angle X-ray scattering**

In this appendix I will describe the SAXS experiments performed on both the wild type and 141-305 Spn1 proteins performed to better understand the structure of the N-and C-terminal regions. I prepared all of the samples for the analysis and wrote this appendix. Dr. Sheena Darcy, Dr. Wayne Lilyestrom, and Nick Clark are responsible for performing the experiments at the synchrotron at UC Berkeley. Dr. Mark van der Woerd is responsible for the calculations, interpretation of the data, and all of the figures.

Small-angle X-ray scattering experiments were performed on the wild type Spn1 protein in an effort to learn more about the structure and confirmation of the N and C terminal regions relative to the conserved central domain. These regions are required for Spn1 to interact with nucleosomes and were determined to be intrinsically disordered although some additional secondary structure could be present (see Ch. 3). Due to the highly charged nature of Spn1 (N-terminus pI=4.5, C-terminus pI=10.6, central domain pI=8.1 with several charged patches) the confirmation of the N and C terminal regions could possibly be relatively fixed or stable. For example the N-and C-terminal regions could interact with one another neutralizing the high level of opposite charge that each region has. Several serine residues within the N-terminus have been observed to be phosphorylated *in vivo*<sup>75, 91</sup>. Phosphorylation could cause the N-and C-regions to repel each other thus activating/inhibiting potential functions of Spn1. Another possibility is that these regions could regulate the functions of the central domain by folding back and interacting with this domain under certain conditions. Understanding more about the orientation of the N-and C-terminal regions could reveal more about Spn1 function. The results of the SAXS experiments performed to answer these questions are reported here. Although the data indicates that Spn1 is difficult to analyze with this technique due to a high susceptibility for radiation damage (along

with other problems), the details of the approach are provided to help with the improvement of future SAXS experiments on this protein.

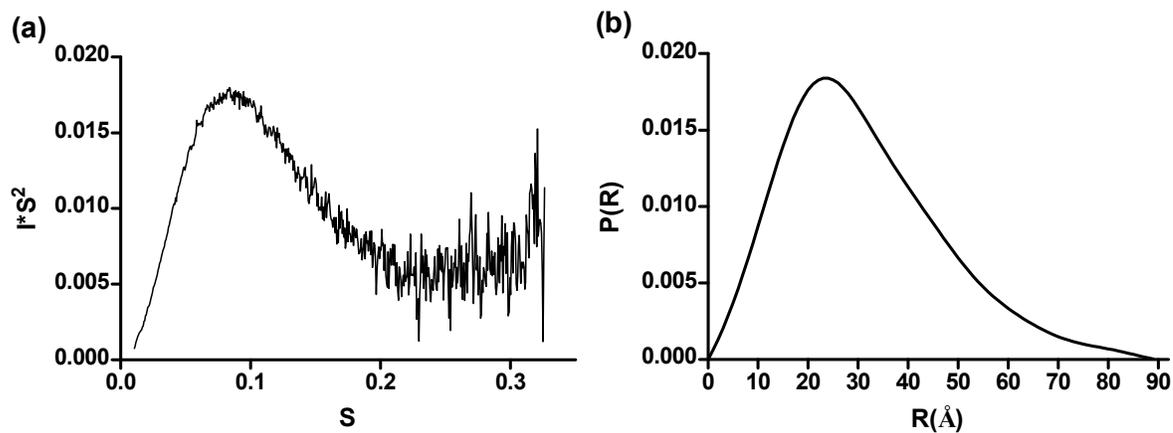
Both the wild type and 141-305 Spn1 proteins were expressed and purified to greater than 95% homogeneity (see Ch. 2). The central domain of Spn1 was used as a control to measure the accuracy of this technique since the structure for this domain has been solved and can be compared to the SAXS results<sup>58</sup>. The experiment was performed on each protein at three different protein concentrations 3, 6, 12 mg/ml (141-305 at 300mM NaCl) and 2.5, 5, 10 mg/ml (wild type at 150mM and 500mM NaCl). Wild type Spn1 was tested under two different salt conditions to determine if the shape of the molecule is salt dependent due to the high amount of charge in the N-and C-terminal regions. Each individual sample was exposed three times consecutively (2 sec, 10 sec, 2 sec). It is customary to use repeat exposures of variable time to check for radiation damage to the sample while variable concentrations should reveal if there is concentration-dependence in the signal<sup>121</sup>. An overview of the pertinent information regarding the results is as follows. Figure A2.1a shows three scattering profiles at the lowest angles for the central domain of Spn1. These angles are chosen because the data in this region is the most sensitive to changes caused by radiation-induced particle aggregation. Since the profiles are clearly not identical, we conclude that the protein solution changes as a function of X-ray exposure. The change in the profiles suggests that radiation-induced aggregation takes place because the subsequent exposures indicate larger particles are present than in the initial exposure. In figure A2.1b, the scaled scattering data from the three different protein concentrations have been overlaid. The shape of the scattering curves, especially at low angles (shown), is not constant and therefore the conclusions for shape and size of the particle in solution will depend on which protein concentration is used.



**Figure A2.1 A)** The central domain of Spn1 is sensitive to radiation damage. Radiation damage is evident from the increase in the X-ray scattering signal with each exposure in every sample tested. In this figure, filled circles represent the first exposure, squares the second exposure and triangles the third exposure, respectively. If the sample were not sensitive to radiation damage, the graphs would superimpose. If any data are used, only the first set should be used, since there is evidence of radiation-induced aggregation in subsequent data sets. The central domain of Spn1 at a concentration of 12 mg/ml in a buffer consisting of 20mM Tris-HCl, 300mM NaCl, and 10% glycerol was exposed in a 15 $\mu$ l cuvette for 2, 10 and 2 seconds, respectively. A reference signal for the identical buffer (without protein) was recorded in the same way. Both images were subtracted after correction for incident X-ray intensity and radially averaged to provide a one-dimensional graph of intensities as a function of scattering angle. **Figure A2.1 B)** Concentration-dependence of the SAXS signal for the central domain of Spn1 shows that this protein is slightly heterogeneous. Especially at low angles (shown), the shape of the curve depends (slightly) on the protein concentration. This suggests that the species in solution (monomer, dimer, etc) are dependent on concentration and the equilibrium may shift with concentration. Therefore different conclusions may be reached about particle size and shape, based on which protein concentration is used.

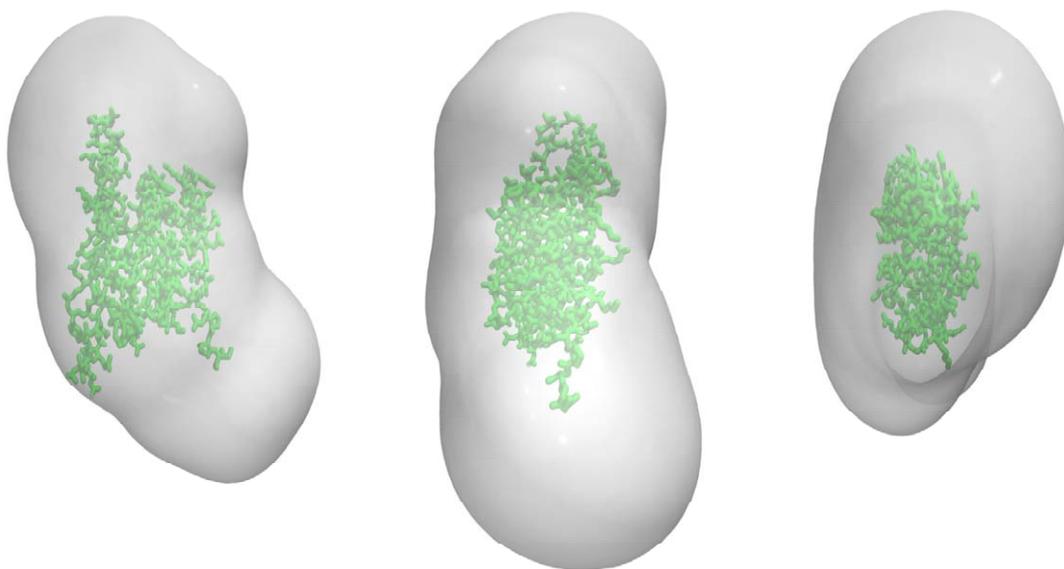
This result makes the outcome of the experiment tenuous. It is difficult to quantify this concentration dependence or the extent of radiation damage from the scattering profiles. Porod's law allows for the estimation of the particle volume and by extension the molecular weight of a protein in solution can be estimated. In the case of the central domain of Spn1 this estimate is  $44,400\text{\AA}^3$  for the lowest protein concentration measured and the molecular weight calculation is 22.2kDa. The calculation of the Porod volume depends on the concentration and on whether noisy data points are included in the calculation. This molecular weight compares favorably with the actual molecular weight (21.03 kDa with His tag) and is consistent with a monomer in solution. The Kratky plot of a completely globular particle in solution is expected to show a parabolic curve, while in the opposite case, a fully unfolded particle (such as a synthetic polymer) is expected to give a linear curve at high angle. The Kratky plot for the central domain of Spn1 shown (Figure A2.2A) is largely parabolic in nature, indicating that this protein is folded, in agreement with the crystal structure. Thus, the analysis of SAXS data acquired in the first exposure is consistent with data from other methods. With this in mind, the data collected were analyzed using the indirect Fourier transform as a crude measure of the shape of the protein (see figure A2.2B). The curve here approaches a bell-shaped curve, which indicates that the particle is spherical in nature. This curve is significantly different for fundamental shapes (sphere, disk, rod etc), and in comparison with fundamental shapes it is difficult to interpret minor changes by visual inspection. Since the crystal structure for the central domain of Spn1 is roughly spherical, this provided confidence in the applicability of the data. The  $P(r)$  provides an upper estimate for  $D_{\text{max}}$  of the central domain of Spn1 (maximum dimension of the particle) at  $90\text{\AA}$  (compare to  $\sim 57\text{\AA}$  for the crystal structure of the core domain, see discussion below).

From the scattering data, shape reconstructions were performed<sup>122</sup> and the low-resolution molecular envelope for the central domain of Spn1 is shown in Figure A2.3. It is evident that the solution-based envelope is larger than the crystallographic model. These reconstructions are an



**Figure A2.2 A)** Kratky plot for central domain of Spn1 at 2.7 mg/ml concentration. The Kratky plot for a fully folded particle should be expected to look like a parabola. The Kratky plot for a completely unfolded particle (such as a synthetic polymer) will be linear at higher angles (and not parabolic). The plot in this figure is a combination of both, with a large fraction of the protein folded and a small fraction unfolded.

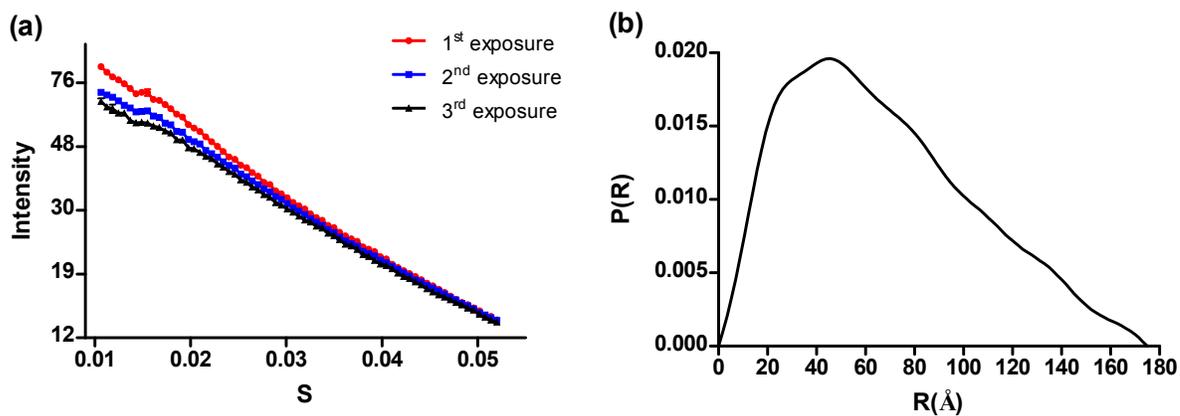
**Figure A2.2 B)** Fourier transformed scattering data indicate that the central domain of Spn1 is largely globular with a maximum dimension of approximately 90Å.



**Figure A2.3:** A model for central domain of Spn1 in solution. The central domain of Spn1 in solution is shown as a transparent green envelope and the crystal structure is superimposed onto the envelope. The model and the crystal structure are dissimilar. This can be explained by the fact that for the crystal structure of the central domain of Spn1 (141-305) only residues 156-300 (145 residues) had significant electron density, thus they are not present in the crystallographic model, and the central domain construct used in this work has a 20 amino acid his tag on the N-terminus. It should therefore be expected that the solution-based envelope is significantly larger than the crystal structure.

indicator that an appropriate control had not been chosen. A comparison between the central domain and full length Spn1 was the goal in an attempt to determine the orientation of the disordered N-and C-terminal regions of the protein. In the crystal structure for the central domain, amino acids 156-300 had sufficiently reliable electron density to be mapped in the structure. This leaves 20 amino acids unaccounted for. Also in the SAXS experiments performed here, a 20 amino acid His tag was left on the proteins. This means that the model shown here is from fitting a 145-amino acid crystal structure inside an envelope developed from a 185-amino acid protein. In the future a better control would be to PCR a central domain construct that only had the amino acids from the crystal structure and to remove the His tag during the purification scheme.

Wild type Spn1 was tested under two different salt conditions (150mM and 500mM NaCl). Under high salt conditions the this protein also appears to be sensitive to radiation because subsequent exposures change the measured scattering signal (Figure A2.4A). Unexpectedly wild type Spn1 appears to get smaller as a consequence of exposure to radiation. This is unusual and triggers the question of whether the initial graph and exposure is indicative for Spn1 monomers or multimers (or both). Different concentrations of Spn1 in solution give the same scattering curves (data not shown). The curves superimpose well and subsequent exposures trigger behavior similar to that shown in Figure A2.4A. However, since the signal does not change for each protein concentration it is suggestive that an equilibrium between monomers and multimers is not present. Despite this the Porod volumes and molecular weights do depend on the protein concentration (Table A2.1). Although the AUC experiments in Ch. 3 suggest that wild type Spn1 is a monomer under the buffer conditions tested here, a different technique such as SEC-MALS is needed to analyze this protein at higher protein concentrations to ensure that the protein is mono-dispersed under the same concentrations used for SAXS experiments. Even though shape reconstructions cannot be done due to the radiation damage and heterogeneity of the wild type Spn1 sample, the interpretation of the Fourier-inverted data for this protein does



**Figure A2.4 A)** Three subsequent exposures of full length Spn1 at low angle, showing that the signal changes with exposure. Contrary to the behavior expected from a protein when radiation damage occurs, these data suggest that there is *anti*-aggregation induced by radiation. That is, the subsequent graphs indicate smaller particles. This is a concern because we now must ask if the first exposure is indicative of a Spn1 monomer, or a multimer. **Figure A2.4 B)** The P(r) function for full length Spn1 at 500mM salt suggests that the particle is elongated with flexible tails. The estimated maximum particle dimension is approximately 175Å, while the shape of the curve (trailing end) suggests that there are flexible, extended tails. Compare with figure A2.2b.

<b>Full length Spn1 Concentration (mg/ml)</b>	<b>Molecular weight (kDa)</b>	<b>Porod Volume (<math>10^5 \text{ \AA}^3</math>)</b>
<b>10</b>	<b>105</b>	<b>2.1</b>
<b>5.1</b>	<b>115</b>	<b>2.3</b>
<b>2.8</b>	<b>85</b>	<b>1.7</b>

**Table A2.1:** Porod volumes (in  $10^5 \text{ \AA}^3$ ) as a function of protein concentration for wild type Spn1 at 500mM NaCl. The concentrations were accomplished by dilution of the most concentrated sample. The actual concentration estimates are based on scattering intensity. The estimated volume for a theoretical full length Spn1 monomer is  $0.96 * 10^5 \text{ \AA}^3$ . Therefore the Porod volumes suggest that Spn1 may be a mixture of monomers and multimers in solution.

<b>Full length Spn1 Concentration (mg/ml)</b>	<b>Molecular weight (kDa)</b>	<b>Porod Volume (<math>10^5 \text{ \AA}^3</math>)</b>
<b>10</b>	<b>150</b>	<b>3.0</b>
<b>4.7</b>	<b>100</b>	<b>2.0</b>
<b>1.9</b>	<b>85</b>	<b>1.7</b>

**Table A2.2:** Porod volumes (in  $10^5 \text{ \AA}^3$ ) as a function of protein concentration for wild type Spn1 at 150mM NaCl. The concentrations were accomplished by dilution of the most concentrated sample. The actual concentration estimates are based on scattering intensity. The estimated volume for a theoretical full length Spn1 monomer is  $0.96 * 10^5 \text{ \AA}^3$ . Therefore the Porod volumes suggest that Spn1 may be a mixture of monomers and multimers in solution.

support previous data that the N-and C-terminal regions are intrinsically disordered in solution. The graph and the behavior of the Dmax (maximum particle dimension) estimate suggests that wild type Spn1 is not a globular molecule, but rather an extended molecule with an estimated extended size of 175Å (Figure A2.4b). The P(r) plot also suggests that this protein has an elongated, rod-like shape.

The experiments described above were repeated under lower salt conditions (150mM NaCl), with very similar results. The Porod volume under these conditions is estimated between 1.7 and  $3.0 \times 10^5 \text{ \AA}^3$  (see Table A2.2). The changes in the size of full length Spn1 are still evident, however, and are very hard to understand.

### **Materials and Methods: Small Angle X-ray scattering**

SAXS data for all samples were collected at the SIBYLS beam line (12.3.1) at the ALS (Berkeley) with an X-ray energy of 10keV ( $\lambda=1.2398\text{\AA}$ ). A Mar CCD detector was used to record the scattering data. A 15 $\mu$ l sample was placed in a 1mm thick chamber with two windows of 25 $\mu$ m mica. The distance between the sample and the detector was 1.5m. Stock solutions of the wild type and the 141-305 Spn1 proteins at 10mg/ml and 12mg/ml respectively were prepared. Additional samples were prepared by diluting the stock solutions to one half and one quarter concentration with reference buffer (20mM Tris, pH 7.5, 150, 300, or 500mM NaCl, 10% glycerol). The intensity curves (intensity as a function of momentum of transfer  $s=4\pi \sin(\theta)/\lambda$  with  $2\theta$  the total scattering angle) were measured at all concentrations, and corrected for buffer scattering. Repeat exposures were taken to check for radiation damage, while two different exposures, typically of 2 and 10s in duration, were taken to optimize the signal-to-noise ratio and avoid detector saturation. Initial data processing was performed with the program PRIMUS<sup>123</sup>. The maximum particle dimension was estimated by regularized indirect Fourier transform with the program GNOM, which provides the P(r) function<sup>124</sup>. The P(r) function is a histogram of inter-atomic distances and goes to zero at the maximum intra-molecular distance Dmax. Low-

resolution molecular envelopes were calculated from the scattering data for the 141-305 Spn1 protein using the program DAMMIN<sup>122</sup> and superimposed using SUPCOMB. Finally, the pre-aligned models were averaged with DAMAVER, giving an effective occupancy of each voxel<sup>125</sup>. A convex shell of all models can be generated by keeping all occupied voxels, while filtering at half-maximal occupancy provides ‘filtered’ models. To visualize the results, the reconstructed models were converted to volumetric maps using real-space convolution with a Gaussian kernel with the program Situs<sup>126; 127</sup>. A kernel width of 6Å and voxel spacing of 1Å were used. Molecular representations were made with VMD<sup>128</sup>.

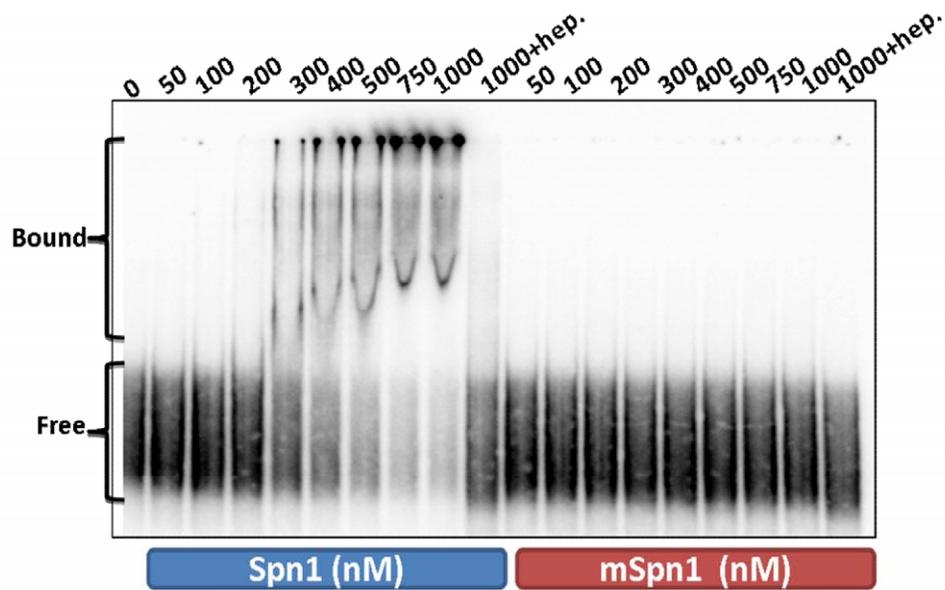
### **Appendix 3: Spn1 can bind to single-stranded RNA in vitro and has a preference for mRNA**

Experiments to determine whether Spn1 is able to bind to RNA in vitro will be described here. I prepared all of the protein samples and wrote this appendix. Jerome Lee prepared the RNA that was used, performed the binding experiments, and did the calculations. He also contributed figure A3.1.

Along with the Spn1 protein's functions involving transcription initiation, elongation, and chromatin remodeling, Spn1 also plays a role in mRNA processing and export in the nucleus. In mammalian cells, siRNA knockdown of Spn1 results in splicing defects and bulk retention of poly A+ mRNAs in the nucleus<sup>59</sup>. Further research has indicated that Spn1 is required for the recruitment of RNA export factor REF1/Aly and that it also interacts with the nuclear exosome subunit Rrp6<sup>59</sup>. Although less evidence is available characterizing these functions in *Saccharomyces cerevisiae*, yeast Spn1 has been shown to physically interact with NAB2 (nuclear polyadenylated RNA binding factor)<sup>129</sup> which is required for proper mRNA export<sup>130</sup> and control of the length of the poly A tail on mRNA<sup>131</sup>. Yeast Spn1 also physically interacts with Npl3<sup>132</sup>, an RNA-binding protein that carries poly(A) mRNA from the nucleus to the cytoplasm<sup>133; 134</sup> and is required for pre-mRNA splicing<sup>135</sup>. This data is suggestive that Spn1 functions to couple transcription with pre-mRNA processing and export. It is therefore plausible that Spn1 helps to coordinate these processes in part by binding directly to RNA within the cell. To test this possibility, in vitro RNA-binding experiments were performed on Spn1 in collaboration with Ph.D. graduate student Jerome Lee.

Both wild type Spn1 and the 141-305 Spn1 protein were purified to greater than 95% homogeneity and subsequently combined with radioactively labeled in vitro-transcribed mRNA-like molecules in an electrophoretic mobility shift assay (EMSA) to detect binding. Variations of capped and uncapped mRNA along with mRNA with and without a poly A tail were also used in this assay. Wild type Spn1 exhibited a preference for poly A-containing mRNA while no

preference was detected for capped or uncapped mRNA. The  $K_d$  for wild type Spn1 and a capped poly A mRNA was 464.1nM (figure A3.1). The central domain of Spn1, however, does not exhibit the ability to bind to any of these same mRNA molecules. The interactions detected in these gel shift assays are all abolished in the presence of heparin which indicates that they are not sequence specific and are likely due an interaction with the sugar-phosphate backbone. The lack of sequence-specificity in these reactions is expected since recent microarray data suggests that Spn1 is recruited throughout the entire genome<sup>1</sup>. If Spn1 plays a role in pre-mRNA splicing or export, equal affinity for all mRNA molecules is logical. Further experimentation is required to verify that Spn1 truly has a preference for mRNA with a poly A tail. In the experiments shown here, the poly A-containing molecules were longer than the other mRNAs used. Therefore the increased affinity could be from a higher number of binding sites rather than from specificity for adenine. If the specificity for the poly A tail holds true, then this would be suggestive that Spn1 transitions to binding to mRNA in the latter steps of transcription where subsequent processing or export functions could be performed. This transition would likely be dependent on the presence of one or both of the N-and C-terminal regions of the Spn1 protein. Due to the basic pI of the C-terminal region, it is likely that this region is responsible for these considering the negative charge of the sugar phosphate backbone. Constructs that have each of these, regions individually could be expressed and purified to test this hypothesis. Another experiment that could be performed to show in vivo relevance would be an RNA IP. This could be used to see if Spn1 immunoprecipitates with any mRNA molecules that could be detected using linear or real-time PCR.



Spn1 Affinity [Capped poly(A)+ RNA]	
Mean $K_D$ (nM)	Std. Dev.
464.1	26.4

Substrate	Fraction Bound at 500nM Spn1
Capped poly(A)	0.57
Uncapped poly(A)	0.53
Capped $A_n$	0.29
Uncapped $A_n$	0.31

**Fig. A3.1** EMSA performed with capped poly(A)+ RNA (Gem A60) and recombinant wild type Spn1 or the central domain of Spn1. RNAs were in vitro-transcribed and labeled with UTP [ $P^{32}$ ] using SP6 RNA polymerase. A constant amount of RNA was used in each reaction and Spn1 was titrated in nanomolar amounts as indicated. The binding can be disrupted by the addition of Heparin ( $2\mu\text{g}/\mu\text{L}$ ) indicating that it is likely a non-sequence-specific interaction with the sugar phosphate backbone. The central domain of Spn1 protein does not bind to RNA in this experiment. Affinities were calculated using a phosphorimager and graphpad prism. The values are from the averages of three experiments. Furthermore, it appears the wild type protein may have slightly lower affinity for non-polyadenylated RNAs (though this merely may be dependent on length).

## Materials and Methods:

### Electrophoretic Mobility Shift Assays using in vitro transcribed RNA

RNAs were in vitro transcribed and labeled with UTP [ $P^{32}$ ] using SP6 RNA polymerase from pGem TNF<sup>136</sup>, pGem ARE<sup>137</sup>, pGem TNF<sup>136</sup>, and pGem4 plasmids. pGem TNF contains 250 nts of 3'-UTR sequence flanking the ARE of TNF<sup>136</sup>. pGem ARE contains just the 34-nt ARE from TNF mRNA. pGem TNF is derived from pGem TNF by deletion of 53 nts containing the ARE, and pGem4 is a cloning vector from Promega and was used to generate the control Gem RNA. All plasmids were linearized with *HindIII* prior to transcription. Increasing amounts of recombinant wild type Spn1 and the 141-305 Spn1 proteins (see Ch. 2) were incubated with 3 fmol of the indicated RNA in the presence of 20 units of RNase inhibitor, 0.15mM spermidine, 20 mM HEPES (pH 7.9), 8% glycerol, 100 mM KCl, and 2 mM MgCl<sub>2</sub> for 5 min at 30 °C in a total volume of 10  $\mu$ l. Low-molecular weight heparin (Sigma) was added to a final concentration of 2  $\mu$ g/ $\mu$ l when indicated. Samples were chilled on ice for five minutes and 2  $\mu$ l of loading dye (0.5% bromphenol blue, 0.5% xylene cyanol, 30% glycerol) were added. This was followed by electrophoresis at room temperature on 5% native polyacrylamide gels in 1X TBE buffer at 10 V/cm. Gels were dried, exposed to a phosphor screen, and visualized by Phosphor-Imaging using a Typhoon Trio Imager (GE Healthcare) or FX Personal Imager (Bio-Rad) and the accompanying software. The fraction of RNA bound was calculated by quantifying the amount of RNA associated with protein and dividing it by the total amount of RNA in each lane. Dissociation constants ( $K_d$ ) were defined as the protein concentration required to achieve half-maximal binding at equilibrium.

## References:

1. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J. & Cramer, P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17, 1272-8.
2. Venters, B. J., Wachi, S., Mavrich, T. N., Andersen, B. E., Jena, P., Sinnamon, A. J., Jain, P., Roller, N. S., Jiang, C., Hemeryck-Walsh, C. & Pugh, B. F. A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. *Mol Cell* 41, 480-92.
3. Osiewacz, H. D. & Scheckhuber, C. Q. (2006). Impact of ROS on ageing of two fungal model systems: *Saccharomyces cerevisiae* and *Podospora anserina*. *Free Radic Res* 40, 1350-8.
4. Hipkiss, A. R. (2009). Error-protein metabolism and ageing. *Biogerontology* 10, 523-9.
5. Kennedy, B. K. & Kaeberlein, M. (2009). Hot topics in aging research: protein translation, 2009. *Aging Cell* 8, 617-23.
6. Kaeberlein, M. & Kennedy, B. K. (2007). Protein translation, 2007. *Aging Cell* 6, 731-4.
7. Vellai, T., Takacs-Vellai, K., Sass, M. & Klionsky, D. J. (2009). The regulation of aging: does autophagy underlie longevity? *Trends Cell Biol* 19, 487-94.
8. Levine, M. & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* 424, 147-51.
9. Sauer, F. & Tjian, R. (1997). Mechanisms of transcriptional activation: differences and similarities between yeast, *Drosophila*, and man. *Curr Opin Genet Dev* 7, 176-81.
10. Maniatis, T., Goodbourn, S. & Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* 236, 1237-45.
11. Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* 11, 347-51.
12. Orphanides, G. & Reinberg, D. (2002). A unified theory of gene expression. *Cell* 108, 439-51.
13. Conaway, J. W. & Conaway, R. C. (1999). Transcription elongation and human disease. *Annu Rev Biochem* 68, 301-19.
14. Villard, J. (2004). Transcription regulation and human diseases. *Swiss Med Wkly* 134, 571-9.
15. Vogelstein, B. & Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nat Med* 10, 789-99.
16. Kim, T. M. & Park, P. J. (2011). Advances in analysis of transcriptional regulatory networks. *Wiley Interdisciplinary Reviews-Systems Biology and Medicine* 3, 21-35.
17. Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Biol* 11, 394-403.
18. Ostergaard, S., Olsson, L. & Nielsen, J. (2000). Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 64, 34-50.
19. Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and Molecular Biology Reviews* 62, 465-503.
20. Margaritis, T. & Holstege, F. C. (2008). Poised RNA polymerase II gives pause for thought. *Cell* 133, 581-4.
21. Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., Murray, B. G., Kapraun, D. F., Greilhuber, J. & Bennett, M. D. (2007). Eukaryotic genome size databases. *Nucleic Acids Res* 35, D332-8.

22. Matsunaga, M. & Jaehning, J. A. (2004). Intrinsic promoter recognition by a "core" RNA polymerase. *J Biol Chem* 279, 44239-42.
23. Warne, S. E. & deHaseth, P. L. (1993). Promoter recognition by *Escherichia coli* RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the -10 and -35 regions are dependent on spacer DNA sequence. *Biochemistry* 32, 6134-40.
24. Lemon, B. & Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* 14, 2551-69.
25. Kuras, L. & Struhl, K. (1999). Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature* 399, 609-613.
26. Kornberg, R. D. (2001). The eukaryotic gene transcription machinery. *Biol Chem* 382, 1103-7.
27. Goodrich, J. A. & Tjian, R. (1994). Transcription Factors IIE and IIIH and ATP Hydrolysis Direct Promoter Clearance by RNA Polymerase II. *Cell* 77, 145-156.
28. Kim, T. K., Ebright, R. H. & Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science* 288, 1418-22.
29. Saunders, A., Core, L. J. & Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7, 557-67.
30. Sobennikova, M. V., Shematorova, E. K. & Shpakovskii, G. V. (2007). [C-terminal domain (CTD) of the subunit Rpb1 of nuclear RNA polymerase II and its role in the transcription cycle]. *Mol Biol (Mosk)* 41, 433-49.
31. Mapendano, C. K., Lykke-Andersen, S., Kjems, J., Bertrand, E. & Jensen, T. H. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol Cell* 40, 410-22.
32. Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868-71.
33. Svejstrup, J. D. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta* 1677, 64-73.
34. Kornberg, R. D. & Lorch, Y. (1992). Chromatin structure and transcription. *Annu Rev Cell Biol* 8, 563-87.
35. Kornberg, R. D. & Lorch, Y. (2002). Chromatin and transcription: where do we go from here. *Curr Opin Genet Dev* 12, 249-51.
36. Kingston, R. E. & Narlikar, G. J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* 13, 2339-52.
37. Clapier, C. R. & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78, 273-304.
38. Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. *Nature* 386, 569-577.
39. Price, D. H. (2008). Poised polymerases: on your mark...get set...go! *Mol Cell* 30, 7-10.
40. Radonjic, M., Andrau, J., Lijnzaad, P., Kemmeren, P., Kockelkorn, T., van Leenen, D., van Berkum, N. & Holstege, F. (2005). Genome-Wide Analyses Reveal RNA Polymerase II Located Upstream of Genes Poised for Rapid Response upon *S. cerevisiae* Stationary Phase Exit. *Molecular Cell* 18, 171-183.
41. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77-88.
42. Rougvie, A. E. & Lis, J. T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of *D. melanogaster* is transcriptionally engaged. *Cell* 54, 795-804.

43. Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J. & Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nat Genet* 39, 1507-11.
44. Zeitlinger, J., Stark, A., Kellis, M., Hong, J. W., Nechaev, S., Adelman, K., Levine, M. & Young, R. A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39, 1512-6.
45. Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., Larsson, J., Lambertsson, A., Handa, H. & Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev* 17, 1402-14.
46. Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. & Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. *Mol Cell Biol* 22, 2918-27.
47. Fischbeck, J. A., Kraemer, S. M. & Stargell, L. A. (2002). SPN1, a conserved yeast gene identified by suppression of a post-recruitment defective yeast TATA-binding protein mutant. *Genetics* 162, 1605-1616.
48. Zhang, L., Fletcher, A., Cheung, V., Winston, F. & Stargell, L. A. (2008). Spn1 regulates the recruitment of Spt6 and the Swi/Snf complex during transcriptional activation by RNA polymerase II. *Mol Cell Biol* 28, 1393-403.
49. Sherman, F., Stewart, J. W., Margoliash, E., Parker, J. & Campbell, W. (1966). The structural gene for yeast cytochrome C. *Proc Natl Acad Sci U S A* 55, 1498-504.
50. Hortner, H., Ammerer, G., Hartter, E., Hamilton, B., Rytka, J., Bilinski, T. & Ruis, H. (1982). Regulation of synthesis of catalases and iso-1-cytochrome c in *Saccharomyces cerevisiae* by glucose, oxygen and heme. *Eur J Biochem* 128, 179-84.
51. Guarente, L. & Mason, T. (1983). Heme regulates transcription of the CYC1 gene of *S.cerevisiae* via an upstream activation site. *Cell* 32, 1279-1286.
52. Guarente, L., Lalonde, B., Gifford, P. & Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S.cerevisiae*. *Cell* 36, 503-511.
53. Boss, J. M., Darrow, M. D. & Zitomer, R. S. (1980). Characterization of yeast iso-1-cytochrome c mRNA. *J Biol Chem* 255, 8623-8.
54. Lee, S. K., Fletcher, A.G.L., Zhang, L., Chen, X., Fischbeck, J.A., Stargell, L.A. (2010). Activation of a poised RNAPII-dependent promoter requires both SAGA and Mediator. *GENETICS* 184, 659-672.
55. Martens, C., Krett, B. & Laybourn, P. (2001). RNA polymerase II and TBP occupy the repressed CYC1 promoter. *Molecular Microbiology* 40, 1009-1019.
56. McDonald, S. M., Close, D., Xin, H., Formosa, T. & Hill, C. P. Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. *Mol Cell* 40, 725-35.
57. Diebold, M. L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J. & Romier, C. The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIIS, Elongin A and Med26. *Embo J* 29, 3979-91.
58. Pujari, V., Radebaugh, C. A., Chodaparambil, J. V., Muthurajan, U. M., Almeida, A. R., Fischbeck, J. A., Luger, K. & Stargell, L. A. (2010). The transcription factor Spn1 regulates gene expression via a highly conserved novel structural motif. manuscript in preparation.
59. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* 21, 160-74.

60. Yoh, S. M., Lucas, J. S. & Jones, K. A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* 22, 3422-34.
61. Hampsey, M. (1997). A review of phenotypes in *Saccharomyces cerevisiae*. *Yeast* 13, 1099-1133.
62. Parrella, E. & Longo, V. D. (2008). The chronological life span of *Saccharomyces cerevisiae* to study mitochondrial dysfunction and disease. *Methods* 46, 256-62.
63. Boeke, J. D., Trueheart, J., Natsoulis, G. & Fink, G. R. (1987). 5-fluoroorotic acid as a selective agent in yeast molecular genetics. *Meth. Enzymol.* 154, 164-175.
64. Hirsch, J. P. & Henry, S. A. (1986). Expression of the *Saccharomyces cerevisiae* inositol-1-phosphate synthase (INO1) gene is regulated by factors that affect phospholipid synthesis. *Mol Cell Biol* 6, 3320-8.
65. Sreerama, N. & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287, 252-60.
66. Sreerama, N. & Woody, R. W. (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol* 383, 318-51.
67. Carruthers, L. M., Schirf, V. R., Demeler, B. & Hansen, J. C. (2000). Sedimentation velocity analysis of macromolecular assemblies. *Methods Enzymol* 321, 66-80.
68. Demeler, B., Behlke, J. & Ristau, O. (2000). Molecular parameters from sedimentation velocity experiments: whole boundary fitting using approximate and numerical solutions of Lamm equation. *Methods Enzymol* 321, 38-66.
69. Demeler, B. & van Holde, K. E. (2004). Sedimentation velocity analysis of highly heterogeneous systems. *Anal Biochem* 335, 279-88.
70. Iyer, V. & Struhl, K. (1996). Absolute mRNA levels and transcriptional initiation rates in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5208-5212.
71. Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. (2001). Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev* 15, 2069-82.
72. Luger, K., Rechsteiner, T. J. & Richmond, T. J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol* 304, 3-19.
73. Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M. & Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* 375, 23-44.
74. Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R. r. & Hartzog, G. A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol.*, 1368-78.
75. Krogan, N. J. m., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S. & Greenblatt, J. F. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol* 22, 6979-92.
76. Shimoaraiso, M., Nakanishi, T., Kubo, T. & Natori, S. (1997). Identification of the region in yeast S-II that defines species specificity in its interaction with RNA polymerase II. *J Biol Chem* 272, 26550-4.
77. Christie, K. R., Awrey, D. E., Edwards, A. M. & Kane, C. M. (1994). Purified yeast RNA polymerase II reads through intrinsic blocks to elongation in response to the yeast TFIIS analogue, P37. *J Biol Chem* 269, 936-43.

78. Guglielmi, B., Soutourina, J., Esnault, C. & Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci U S A* 104, 16062-7.
79. Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F. & Shilatifard, A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* 11, 721-9.
80. Mueller, C. L. & Jaehning, J. A. (2002). Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol Cell Biol* 22, 1971-80.
81. Warner, M. H., Roinick, K. L. & Arndt, K. M. (2007). Rtf1 is a multifunctional component of the Paf1 complex that regulates gene expression by directing cotranscriptional histone modification. *Mol Cell Biol* 27, 6103-15.
82. Uversky, V. N. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev* 40, 1623-34.
83. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim Biophys Acta* 1804, 1231-64.
84. Reingewertz, T. H., Shalev, D. E. & Friedler, A. Structural disorder in the HIV-1 Vif protein and interaction-dependent gain of structure. *Protein Pept Lett* 17, 988-98.
85. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21, 3435-8.
86. Bortvin, A. & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* 272, 1473-6.
87. Adkins, M. W. & Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol Cell* 21, 405-16.
88. Ivanovska, I., Jacques, P. E., Rando, O. J., Robert, F. & Winston, F. Control of chromatin structure by spt6: different consequences in coding and regulatory regions. *Mol Cell Biol* 31, 531-41.
89. Collins, S. R., Miller, K. M., Maas, N. L., Roguev, A., Fillingham, J., Chu, C. S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., Ding, H., Xu, H., Han, J., Ingvarsdottir, K., Cheng, B., Andrews, B., Boone, C., Berger, S. L., Hieter, P., Zhang, Z., Brown, G. W., Ingles, C. J., Emili, A., Allis, C. D., Toczyski, D. P., Weissman, J. S., Greenblatt, J. F. & Krogan, N. J. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806-10.
90. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S. & Krogan, N. J. (2007). Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol Cell Proteomics* 6, 439-50.
91. Chen, S. H., Albuquerque, C. P., Liang, J., Suhandynata, R. T. & Zhou, H. A proteome-wide analysis of kinase-substrate network in the DNA damage response. *J Biol Chem* 285, 12803-12.
92. Kaplan, C. D., Laprade, L. & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096-9.
93. Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R. & Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol* 6, e277.
94. Roth, G. S., Ingram, D. K. & Lane, M. A. (1999). Calorie restriction in primates: will it work and how will we know? *J Am Geriatr Soc* 47, 896-903.

95. Sohal, R. S. & Weindruch, R. (1996). Oxidative stress, caloric restriction, and aging. *Science* 273, 59-63.
96. Bitterman, K. J., Medvedik, O. & Sinclair, D. A. (2003). Longevity regulation in *Saccharomyces cerevisiae*: linking metabolism, genome stability, and heterochromatin. *Microbiol Mol Biol Rev* 67, 376-99, table of contents.
97. Imai, S., Armstrong, C. M., Kaerberlein, M. & Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* 403, 795-800.
98. Kennedy, B. K., Austriaco, N. R., Jr., Zhang, J. & Guarente, L. (1995). Mutation in the silencing gene SIR4 can delay aging in *S. cerevisiae*. *Cell* 80, 485-96.
99. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature* 366, 461-4.
100. Kenyon, C. (2005). The plasticity of aging: insights from long-lived mutants. *Cell* 120, 449-60.
101. Rockenfeller, P. & Madeo, F. (2008). Apoptotic death of ageing yeast. *Exp Gerontol* 43, 876-81.
102. Stanfel, M. N., Shamieh, L. S., Kaerberlein, M. & Kennedy, B. K. (2009). The TOR pathway comes of age. *Biochim Biophys Acta* 1790, 1067-74.
103. Scherz-Shouval, R. & Elazar, Z. (2007). ROS, mitochondria and the regulation of autophagy. *Trends Cell Biol* 17, 422-7.
104. Donati, A., Recchia, G., Cavallini, G. & Bergamini, E. (2008). Effect of aging and anti-aging caloric restriction on the endocrine regulation of rat liver autophagy. *J Gerontol A Biol Sci Med Sci* 63, 550-5.
105. Burtner, C. R., Murakami, C. J., Kennedy, B. K. & Kaerberlein, M. (2009). A molecular mechanism of chronological aging in yeast. *Cell Cycle* 8, 1256-70.
106. Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M., Koller, D., Altman, R. B., Davis, R. W., Nislow, C. & Giaever, G. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320, 362-5.
107. Wanke, V., Cameroni, E., Uotila, A., Piccolis, M., Urban, J., Loewith, R. & De Virgilio, C. (2008). Caffeine extends yeast lifespan by targeting TORC1. *Mol Microbiol* 69, 277-85.
108. Tzagoloff, A. & Myers, A. M. (1986). Genetics of mitochondrial biogenesis. *Annu Rev Biochem* 55, 249-85.
109. Clark-Walker, G. D. & Linnane, A. W. (1967). The biogenesis of mitochondria in *Saccharomyces cerevisiae*. A comparison between cytoplasmic respiratory-deficient mutant yeast and chlormaphenicol-inhibited wild type cells. *J Cell Biol* 34, 1-14.
110. Sor, F. & Fukuhara, H. (1983). Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria. *Nucleic Acids Res* 11, 339-48.
111. Cui, Z. & Mason, T. L. (1989). A single nucleotide substitution at the rib2 locus of the yeast mitochondrial gene for 21S rRNA confers resistance to erythromycin and cold-sensitive ribosome assembly. *Curr Genet* 16, 273-9.
112. Doudican, N. A., Song, B., Shadel, G. S. & Doetsch, P. W. (2005). Oxidative DNA damage causes mitochondrial genomic instability in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25, 5196-204.
113. Chen, X. J., Wang, X. & Butow, R. A. (2007). Yeast aconitase binds and provides metabolically coupled protection to mitochondrial DNA. *Proc Natl Acad Sci U S A* 104, 13738-43.

114. Cardenas, M. E., Cutler, N. S., Lorenz, M. C., Di Como, C. J. & Heitman, J. (1999). The TOR signaling cascade regulates gene expression in response to nutrients. *Genes Dev* 13, 3271-9.
115. Loewith, R., Jacinto, E., Wullschleger, S., Lorberg, A., Crespo, J. L., Bonenfant, D., Oppliger, W., Jenoe, P. & Hall, M. N. (2002). Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control. *Mol Cell* 10, 457-68.
116. Weisman, R. & Choder, M. (2001). The fission yeast TOR homolog, tor1+, is required for the response to starvation and other stresses via a conserved serine. *J Biol Chem* 276, 7027-32.
117. Kaeberlein, M., Powers, R. W., 3rd, Steffen, K. K., Westman, E. A., Hu, D., Dang, N., Kerr, E. O., Kirkland, K. T., Fields, S. & Kennedy, B. K. (2005). Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science* 310, 1193-6.
118. Schieke, S. M. & Finkel, T. (2007). TOR and aging: less is more. *Cell Metab* 5, 233-5.
119. Kanakaraj, I., Jewell, D. L., Murphy, J. C., Fox, G. E. & Willson, R. C. Removal of PCR error products and unincorporated primers by metal-chelate affinity chromatography. *PLoS One* 6, e14512.
120. Nastasijevic, B., Becker, N. A., Wurster, S. E. & Maher, L. J., 3rd. (2008). Sequence-specific binding of DNA and RNA to immobilized Nickel ions. *Biochem Biophys Res Commun* 366, 420-5.
121. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40, 191-285.
122. Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76, 2879-86.
123. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36, 1277-1282.
124. Svergun, D. I. (1992). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *Journal of Applied Crystallography* 25, 495-503.
125. Volkov, V. V. & Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 36, 860-864.
126. Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 125, 185-95.
127. Wriggers, W. & Chacon, P. (2001). Using Situs for the registration of protein structures with low-resolution bead models from X-ray solution scattering. *Journal of Applied Crystallography* 34, 773-776.
128. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph* 14, 33-8, 27-8.
129. Batisse, J., Batisse, C., Budd, A., Bottcher, B. & Hurt, E. (2009). Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. *J Biol Chem* 284, 34911-7.
130. Green, D. M., Marfatia, K. A., Crafton, E. B., Zhang, X., Cheng, X. & Corbett, A. H. (2002). Nab2p is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. *J Biol Chem* 277, 7752-60.
131. Hector, R. E., Nykamp, K. R., Dheur, S., Anderson, J. T., Non, P. J., Urbinati, C. R., Wilson, S. M., Minvielle-Sebastia, L. & Swanson, M. S. (2002). Dual requirement for

- yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *Embo J* 21, 1800-10.
132. Tardiff, D. F., Abruzzi, K. C. & Rosbash, M. (2007). Protein characterization of *Saccharomyces cerevisiae* RNA polymerase II after in vivo cross-linking. *Proc Natl Acad Sci U S A* 104, 19948-53.
  133. Lee, M. S., Henry, M. & Silver, P. A. (1996). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. *Genes Dev* 10, 1233-46.
  134. Gilbert, W., Siebel, C. W. & Guthrie, C. (2001). Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. *Rna* 7, 302-13.
  135. Kress, T. L., Krogan, N. J. & Guthrie, C. (2008). A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol Cell* 32, 727-34.
  136. Moraes, K. C. M., Wilusz, C. J. & Wilusz, J. (2006). CUG-BP binds to RNA substrates and recruits PARN deadenylase. *Rna-a Publication of the Rna Society* 12, 1084-1091.
  137. Gao, M., Wilusz, C. J., Peltz, S. W. & Wilusz, J. (2001). A novel mRNA-decapping activity in HeLa cytoplasmic extracts is regulated by AU-rich elements. *Embo J* 20, 1134-43.