Characterization of SMN and Gemin2: Insights into Spinal Muscular Atrophy

by

Tracy Niday

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2012 by the
Graduate Supervisory Committee:

James Allen, Chair
Giovanna Ghirlanda
Rebekka Wachter

ARIZONA STATE UNIVERSITY

December 2012

ABSTRACT

Spinal muscular atrophy (SMA) is a neurodegenerative disease that results in the loss of lower body muscle function. SMA is the second leading genetic cause of death in infants and arises from the loss of the Survival of Motor Neuron (SMN) protein. SMN is produced by two genes, *smn1* and *smn2*, that are identical with the exception of a C to T conversion in exon 7 of the *smn2* gene. SMA patients lacking the *smn1* gene, rely on *smn2* for production of SMN. Due to an alternative splicing event, *smn2* primarily encodes a non-functional SMN lacking exon 7 (SMN D7) as well as a low amount of functional full-length SMN (SMN WT). SMN WT is ubiquitously expressed in all cell types, and it remains unclear how low levels of SMN WT in motor neurons lead to motor neuron degradation and SMA. SMN and its associated proteins, Gemin2-8 and Unrip, make up a large dynamic complex that functions to assemble ribonucleoproteins.

The aim of this project was to characterize the interactions of the core SMN-Gemin2 complex, and to identify differences between SMN WT and SMN D7. SMN and Gemin2 proteins were expressed, purified and characterized via size exclusion chromatography. A stable N-terminal deleted Gemin2 protein (N45-G2) was characterized. The SMN WT expression system was optimized resulting in a 10-fold increase of protein expression. Lastly, the oligomeric states of SMN and SMN bound to Gemin2 were determined. SMN WT formed a mixture of oligomeric states, while SMN D7 did not. Both SMN WT and D7 bound to Gemin2 with a one-to-one ratio forming a heterodimer and several higher-order oligomeric states. The SMN WT-Gemin2 complex favored high

molecular weight oligomers whereas the SMN D7-Gemin2 complex formed low molecular weight oligomers. These results indicate that the SMA mutant protein, SMN D7, was still able to associate with Gemin2, but was not able to form higher-order oligomeric complexes. The observed multiple oligomerization states of SMN and SMN bound to Gemin2 may play a crucial role in regulating one or several functions of the SMN protein. The inability of SMN D7 to form higher-order oligomers may inhibit or alter those functions leading to the SMA disease phenotype.

DEDICATION

I would like to dedicate this to Dale Niday, who instilled in me the values of education and learning.

ACKNOWLEDGMENTS

I would like to thank my family and friends for their constant support during my graduate school career.

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

1 SMN PROTEIN AND SPINAL MUSCULAR ATROPHY 1

1.1 Brief history of SMN and spinal muscular atrophy 1

1.2 Clinical Features of spinal muscular atrophy 1

1.3 Genetics of SMA 4

1.4 SMN Protein 6

1.5 The SMN complex 8

1.6 The SMN complex and splicing 10

1.7 Molecular mechanisms of SMA 12

1.8 Current therapies in SMA 13

1.9 The core components 14

1.10 Project overview 15

2 EXPRESSION AND PURIFICATION OF GEMIN2 17

2.1 Introduction 17

2.2 Methods 19

2.3 Results and Discussion 22

3 EXPRESSION AND PURIFICATION OPTIMIZATION OF SMN 46

3.1 Introduction 46

CHAPTER Page

3.2 Methods 49

3.3 Results and Discussion 51

4 CHARACTARIZATION OF GEMIN2 AND SMN PROTEINS 68

4.1 Introduction 68

4.2 Methods 70

4.3 Results and Discussion 72

5 CONCLUSION 97

5.1 Conclusion 97

5.2 Future Work 97

REFERENCES 98

LIST OF TABLES

Table		Page
1.1.	SMA disease types	3
2.1.	Solubility limit of His-Gemin2	30
2.2.	Crystallization conditions for N45-G2	44
3.1.	Comparison of refolding techniques	55

LIST OF FIGURES

Figure		Page
1.1.	Motor neurons and SMA	2
1.2.	Genetics of SMA	5
1.3.	SMN protein domains	8
1.4.	The SMN complex	9
1.5.	The role of SMN in assembly splicing machinery	11
1.6.	The assocaiton of Gemin2 and SMN	15
2.1.	Cloning of gemin2 constructs	24
2.2.	Vector maps of gemin2 constructs	25
2.3.	Expression and purification of His-Gemin2	28
2.4.	Stability of His-Gemin2	32
2.5.	Predicted cleavage sites and gemin2 deletion constructs	35
2.6.	Expression and stability of N45-Gemin2	37
2.7.	Stability of N45-Gemin2	40
2.8.	Crystallization Results of N45-Gemin2	44
3.1.	Smn constructs	51
3.2.	Optimization of SMN expression	54
3.3.	Optimization of nickel affinity SMN purification	57
3.4.	Stability of SMN	62
3.5.	Stability of SMN WT compared to SMN TRX	66
4.1.	SMN WT oligomerization states	75
4.2.	SMN D7 oligomerization states	78

Figure Page

4.3. Model of the self-association of SMN in SMA 81

4.4. Gemin2 oligomers 83

4.5. The oligomerization state of SMN WT-Gemin2 86

4.6. SMN-Gemin2 complex and disulfide bonds 89

4.7. Gemin2-SMN interaction differences between SMN WT and D7 92

Chapter 1

SMN PROTEIN AND SPINAL MUSCULAR ATROPHY

1.1 Brief history of SMN and spinal muscular atrophy

Both Micheal Lerner and Latchman (McAllister 1988, Sharpe 1989) discovered the Survival of Motor Neuron (SMN) Protein in the early 1980's. Both scientists reported SMN to be a gene product that was involved in RNA splicing. SMN was named Sm polypeptide N because of sequence similarities to the Sm class of proteins that function to assemble the splicing machinery (Schmaus 1989). Six years later deletions and mutations in the *smn* gene were identified in patients with a disease called spinal muscular atrophy (SMA), a debilitating and fatal genetic disease that begins with lower motor neuron degeneration in the spinal cord followed by loss of muscle function in the trunk and lower body and eventually death (Levebre 1995). Although, much has been uncovered in the last thirteen years about the detailed molecular function of SMN in splicing and other RNA processing events, the specific role of SMN in the pathology of SMA is still unknown and there are no treatments for the disease.

1.2 Clinical Features of Spinal Muscular Atrophy

Spinal muscular atrophy (SMA) is an autosomal recessive disease that affects approximately 1 in 6000 infants and is the second leading genetic cause of death in children after cystic fibrosis. A physician named Werndig (Levebre 1995, Xiao 2011) reported neurological symptoms of SMA such as partial or total muscular atrophy of lower limbs and chest as early as 1891. The muscle atrophy is preceded by the loss of function of the lower motor neurons called alpha motor

neurons, which are located in the anterior horn of the spinal cord. Alpha motor neurons are composed of three main parts; the cell body located in the spinal cord, the axon that extends from the spinal cord to the muscle and the muscle fibers (Figure 1.1). The function of alpha motor neurons is to send an electrical signal through the axons to the muscle fibers, which causes the muscles to contract. The neuromuscular junction (NMJ) is located at the muscle-axon interface and is responsible for transducing the electrical signal across the synapse to the muscle fibers. In SMA, motor neurons are shown to be non-functional, which was indicated by abnormal axon and NMJ composition relative to healthy motor neurons (Figure 1.1) (Baumer 2010). The molecular mechanism that leads to the loss of function of alpha motor neurons in SMA is unknown.

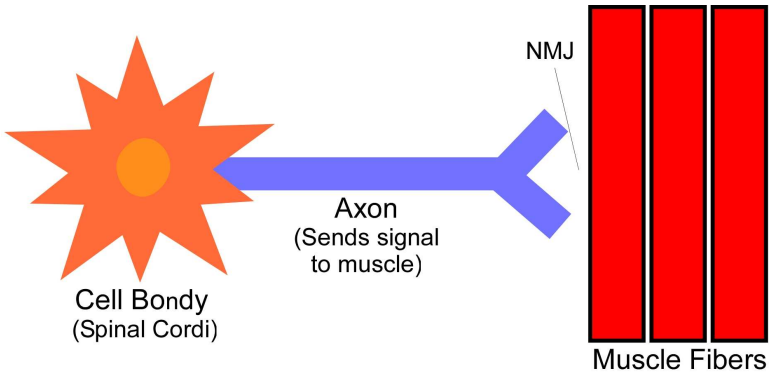


Figure 1.1- Motor Neurons and SMA
Diagram of an alpha motor neuron showing the cell body (orange) located in the spinal cord connected to the axon (blue), which connects to the muscle fibers (red). The neuromuscular junction (NMJ) is located between the axon and the muscle.

SMA is categorized into type I-IV with type I being the most severe and type IV being the least severe (Table 1.1). Type I SMA patients have onset of disease symptoms in infancy, prior to 6 months of age. Type I patients are never able to walk or sit up without assistance. As the muscle failure progresses into the

diaphragm and trunk muscles. The disease progresses rapidly as type I patients usually do not live past their second birthday. Type II patients have a slightly later disease onset of 6-18 months and have less severe symptoms allowing them to sit up and sometimes walk. The life expectancy of type II patients ranges from 5-11yrs of age. Type III SMA patients have a normal life expectancy and a rare type IV is characterized by an adolescent or adult disease onset (Xiao 2011,). The symptoms of the less severe types III and IV are impairment of muscle function usually leading to use of a wheelchair (Xiao 2011). Currently, physicians can only monitor the progression of the disease and help patients live a comfortable life with breathing, feeding and mobility assistance.

Disease Type	Age of Onset	Symptoms
I	0-6 months	Cannot sit up without support
II	7-18 months	May be able to sit up and possibly walk
III	18 months to adolescence	Walk with prevalent weakness, eventually needing a wheelchair
IV	After 35 yrs	Progressive muscle weakness as they age

Table 1.1- SMA disease types
SMA has four disease types (I-IV) based on age of onset and severity of symptoms.

Understanding how mutations in the SMA patient genotype lead to the SMA disease phenotype at a molecular level is crucial in the drug discovery process. Identifying the molecular link of a disease requires two processes, determining the mutated gene product or protein produced from transcription and translation within the cell, and then determining how the mutated protein functions to cause the disease. The atomic structure of proteins provides insight

into the molecular function. Determining the differences in molecular function between native and mutant disease proteins provides potential drug targets that then allow for high-through put molecule screens or rational drug design to identify a successful drug. A drug is defined as successful by its ability to specifically alter the molecular function of the disease protein, relieve the symptoms or underlying pathology of the disease and its level of safety for human consumption. Consequently identifying a drug requires several costly rounds of trial and error to meet the above criteria. Because the drug discovery process is risk intensive and expensive, more knowledge about the structure-function relationship and the molecular pathway of a mutant protein increases the probability and efficiency of creating a drug.

1.3 Genetics of SMA

SMA research has progressed rapidly in the last decade because of the dynamic increase and efficiency of gene sequencing and cellular imaging techniques. These technologies have resulted in a broad picture of the genotype-phenotype link in SMA. SMA patients have gene mutations that lead to reduced levels of full-length SMN protein, which cause cell death in the anterior motor neurons. In humans, the two *smn* genes are located on a 500kb segment of the 5q13 chromosome as two inverted copies, which resulted from a gene duplication event during evolution (Burglan 1995, Capon 1995, Lefebre 1995). The *smn1* gene is located on the centromeric side of the chromosome, whereas *smn2* is located on the telomeric side. The two genes are identical in sequence with the exception a C to T conversion in exon 7 of *smn2*. The T in exon 7 of *smn2* causes

an alternative splicing event after transcription that excludes exon 7. Due to alternative splicing, *smn2* is translated into 90 % of SMN lacking exon 7 (SMND7) and 10% of full-length SMN (SMN WT) (Lorson 2000, Lorson 1999, Crawford 1996) (Figure1.2).

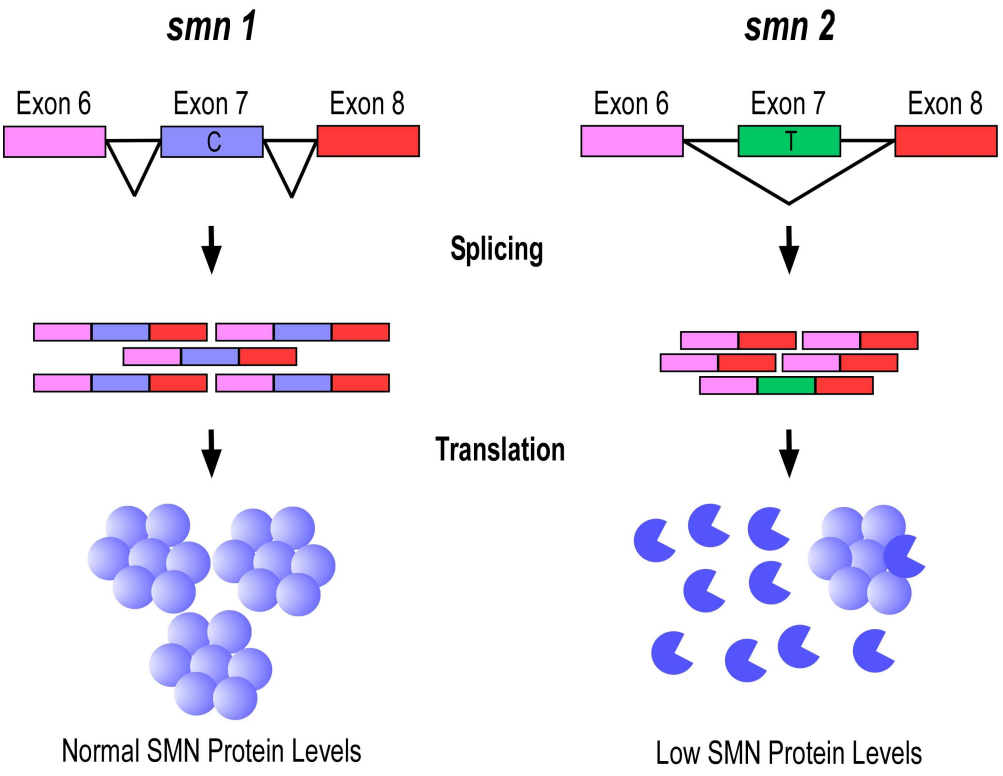


Figure 1.2- Genetics of SMA
SMN is produced from two genes, *smn1* and *smn2*. The transcription and translation *smn1* leads to full-length SMN (SMN WT shown in light blue circles). Due to a C to T conversion in exon seven (green) of *smn2*, an exon skipping event occurs producing primarily SMN lacking the region encoded by exon seven (SMN D7 shown in dark blue).

Ninety-five percent of SMA patients have deletions in exon 6 of the *smn1* gene, which result in a C-terminal truncated protein that is ironically similar to the SMN D7 protein. Another 5% of patients have missense mutations in the *smn1* gene, which also result in non-functional SMN protein (Levebre 1995, Thompson 1995). Because SMA patients do not produce full-length SMN protein or SMN

WT from the *smn1* gene, they rely on the 10% of full-length SMN produced from the *smn2* gene. Although a small percentage of healthy people lack *smn2*, all SMA patients have a copy of the *smn2* suggesting that the lack of any SMN WT production is embryonically lethal (Workman 2009). The number of *smn2* genes and the amount of full-length SMN WT protein produced in SMA patients is directly correlated to the severity of the SMA phenotype with type III patients having 3-4copies and thus more SMN WT, type II having 3 copies and type I having 2 copies (Mailman 2002,Wirth 2006). SMN is expressed in all cell types; however, the mechanism by which low levels of SMN cause a phenotype only in motor neurons of SMA patients is unknown.

1.4 SMN protein

The *smn* gene contains nine exons, 1, 2a, 2b, and 3-8 that translate into a 292 amino acid protein with an approximate molecular weight of 32kDa (Lui 1997). SMN has four domains, the K-rich domain, the P-rich domain, the Tudor domain and the YG domain (Burghes 2009) (Figure 1.3). The K-rich domain is encoded by exons 2a and 2b. Exon 2a is important for binding to Gemin2 and exon 2b contains a self-association region (Young 2000). The P-rich region includes exon 6 and part of exons 4 and 5 and is thought be important for profilin3 binding (Giesemann 1999). SMN does not contain any apparent sequence homology to other proteins with the exception of the Tudor domain, amino acids 90-160, a conserved domain of unknown function originally discovered in drosophila (Buhlor 1999). The Tudor domain located in exon 3 has a Sm protein like fold, which consists of a highly bent anti-parallel beta strand.

The Tudor domain interacts with splicosomal Sm proteins B, D1 and D3 and other RG rich proteins (Selenko 2001, Paushkin 2002). The RG protein-Tudor domain interaction is enhanced by the di-methylation of the arganine residues and is most likely an important regulation step for the several identified RG proteins that bind the Tudor domain. The C-terminal YG box domain contains a self-association region that extends partially into exons 6 and 7 with a stronger affinity relative to the N-terminal self-association region encoded by exon 2b (Talbot 2007, Young 2000). It is postulated that the N and C terminal self-association domains interact to form a dimer or the relatively weaker exon 2 site could interact with other SMN proteins to promote oligomerization (Young 2000). Supporting the importance of these self-association interaction regions, SMND7 lacking part of the self-association region has a decreased ability to self-associate and form oligomers, which disrupts the splicosomal assembly function of SMN. SMND7 and SMA missense mutants that are unable to form oligomers are unstable and rapidly degraded *in vivo* (Burnett 2008). Only 5% of SMA patients have missense mutations, and the majority of those are located in the self-association region and the Tudor domain, with two occurring in the Gemin2 binding region. The reduction in the self-association of SMN missense mutants correlates to SMA disease severity (Lorson 1998, Pellerzoni1999). In SMA mice that have been engineered to express low levels of SMN WT, expression of SMN D7 was able to oligomerize with SMN WT and partially rescue neurodegeneration defects (Le 2005). SMA mutations in both the Tudor domain and C-terminus have been shown to inhibit Sm protein binding and other SMN

substrates such as nuclear proteins RNA pol I and GARI (Jones 2001, Pellizoni 2001, Buhler 1999, Paushkin 2002). Although much is known about the function of each domain of the SMN protein, little is known about the concerted molecular function of the SMN domains.

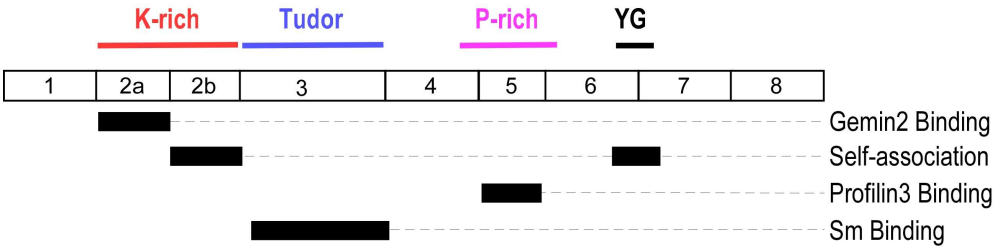


Figure 1.3- SMN protein domains
The SMN protein is encoded by 9 exons that contain the K-rich (red), Tudor (blue), P-rich (pink), and YG (black) domains. Several regions of the protein (black boxes) have been associated with functions such as protein binding and self-association.

1.5 The SMN complex

SMN exists in a large a dynamic macromolecular complex containing Gemins2-8 and unrip proteins. The SMN complex is a scaffold that assists in the assembly of ribonucleoproteins (RNPs) and localizes both to the cytoplasm and also to small foci in the nucleus called gems and coiled bodies (CBs) (Battle 2006). The function of CBs is unknown, but CBs contain the marker protein coilin, splicosomal RNP's and other RNPs and are thought to function in RNA processing and RNP assembly. Gems are often neighbors to CBs, but they do not contain coilin and are highly enriched with both the SMN complex and SnRNPs and may be a storage or recycling site for these complexes. SMN along with Gemins 2-8 and Unrip are termed the SMN complex because of their stable association (Otter 2007) (Figure 1.4). The SMN complex purified from Hela nuclear extracts is similar in size to the 20S ribosomal RNA complex, while

complexes purified from whole Hela cells range in size from 20S-80S (Meister 2000, Carassimi 2006).

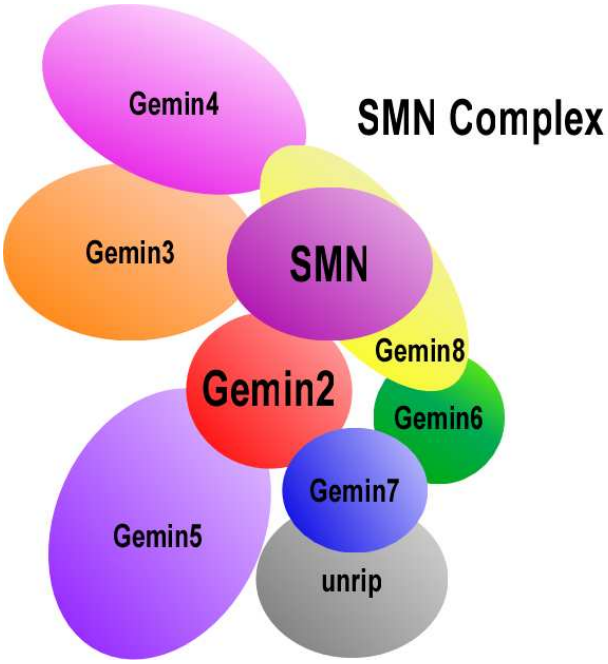


Figure 1.4- The SMN complex
The SMN complex is made up of SMN, Gemins 2-8 and Unrip. SMN interacts with Gemins2, 4 and 8. Gemin2 interacts with SMN, Gemins 5 and 7. Gemin8 interacts with SMN, Gemin6 and 4. Gemin3 is associated to the complex via Gemin4. Unrip is associated with the complex via Gemin7.

Gemin2, first characterized, as SMN interacting protein 1 (SIP1), does not have a clear role in the function of the SMN complex, but according to recent structural results, functions as a bridge between SMN and Sm proteins D1, D2, E, F, and G and may even regulate RNA-Sm protein interactions (Sarachan 2012, Zhang 2011). Gemin3, is a DEAD box protein, a class of RNA helicase proteins, which may be important in the RNA-protein assembly process (Lee 2005). Gemin 4 associates with Gemin3 and has an unknown function, but includes a nuclear localization signal that may transport the SMN complex into the nucleus (Charroux 2000, Lorson 2008). Gemin 5 is an RNA binding protein that

associates with the SMN complex through its interaction with Gemin2 and associates with snRNAs to recruit them to the SMN complex (Battle 2006). Gemins 6 and 7 form a heterodimer that surprisingly has a Sm-fold. The heterodimer is proposed to bind to the Sm heptamer in place of B and D3 in the snRNP assembly process (Leung 2005, Ma 2005). Gemin8 and Unrip both have unknown functions (Grimmler 2005, Carissimi 2006). As more of the components of the SMN complex have been discovered, the interactions between the proteins appear to be complex (Figure 1.3). The evidence points to a myriad of dynamic and complex interactions that have not yet been correlated to the function of the SMN complex. Adding to the complexity is the fact that several of the Gemin proteins and Unrip have SMN-independent functions such as DNA repair, virus replication, and even SMN is proposed to have functions independent of the SMN complex such as transcription regulation (Pellizzoni 2001, Hamamoto 2006, Grimmler 2005, Moureletatos 2002).

1.6 The SMN complex and splicing

RNP's are composed of non-coding RNAs and their associated proteins and they perform a variety of functions in the cell including processing and modification of ribosomal RNA (snoRNP's), splicing (snRNP's), mRNA processing, Histone mRNA processing (U7snRNP) and telomere maintenance (Jones 2001, Pillai 2003, Bachand 2002, Levebre 2002, Pellizoni 2002). The SMN complex is associated with each of the above RNPs; however, its role in assembling splicosomal snRNPs is the best understood. The RNA component of splicosomal RNAs, uridine-rich small nuclear RNAs (snRNA) are transcribed in

the nucleus and transported to the cytoplasm where the SMN complex recognizes a specific code, the sm site (AUUUA), a sequence specific region 5' and a non-specific 7-12bp RNA stemloop usually at 3' side of the sm site. Gemin5 contains an RNA binding WD repeat that specifically recognizes and binds snRNAs. The SMN complex recruits the seven heptameric Sm proteins (B, D1, D3, E, F and G) both through interactions of the SMN tudor domain (B, D1, and D3) and the Gemin proteins and deposits the ring in an ATP-dependent manner onto the snRNA (Meister 2001, Battle 2006) (Figure 1.5). Once the Sm proteins are assembled, the exosome is able to trim the 3' end of the RNA. The RNA is then capped with a trimethylgaunasine, and re-imported back into the nucleus. Inside the nucleus, snRNPs are directed to the Cajal bodies (CBs) to meet up with the last of the proteins required for snRNP function as well as receiving some final RNA modifications before being sent to the sites of transcription for pre-mRNA splicing (Eggert 2006, Workman 2012).

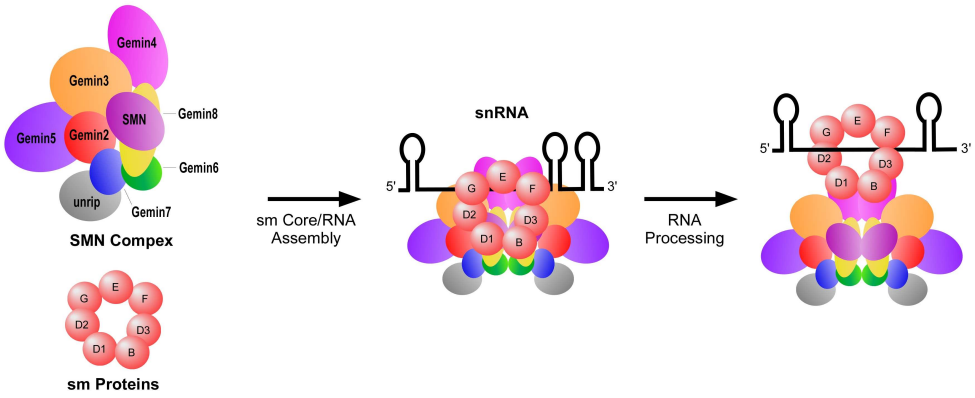


Figure 1.5- The role of SMN in assembly splicing machinery
The SMN complex chaperones the assembly of snRNPs. SMN first associates with the heptamer of Sm proteins and then recruits the snRNA. SMN assembles the Sm core onto the snRNA in an ATP-dependent manner. The snRNA undergoes RNA processing including trimming of the 3' end.

1.7 Molecular mechanisms of SMA

Because of SMN's well-known role in splicing and the fact that low levels SMN do cause a decrease in Sm protein binding and splicing in the cell, the first proposed molecular mechanism that causes SMA is a defect in splicing. In this case, the motor neurons in SMA patients due to deficient amounts of full-length SMN WT are theorized to be more sensitive to global splicing defects then other types of cells or a specific mRNA transcript in the motor neuron is affected by a splicing defect (Burghes 2009). Because a specific or global splicing defect has yet to be revealed, other molecular mechanisms proposing a splicing-independent function have been investigated. The observation that SMN and some Gemins, but not Sm proteins localize with β-actin mRNA at the growth cones of motor neurons and the SMN interaction with Hn-RNP-R that binds to b-actin mRNA led to the splicing-independent theory that SMN may be responsible for axonal transport of β-actin mRNA (Rossoll 2002, Rossoll 2009, Todd 2010a, Todd 2010b). In the β-actin mechanism, SMA motor neurons have defects in the axon growth and axon growth cones of motor neurons because the β -actin mRNA cannot be transported to the growth cones. More research into both the splicing and axonal transport mechanisms will likely determine the role if each or even possibly both of these molecular events is involved in the pathogenesis of SMA (Fallini 2010).

1.8 Current therapies in SMA

Without a full understanding of the complex molecular mechanism for SMA, current treatment avenues are focused on increasing the amount of SMN WT protein in SMA patients. The first treatment approach is to use small molecules and antisense oligonucleotides (ASO's) that inhibit the alternative splicing and exon seven exclusion of the *smn2* pre-mRNA transcript and therefore allow for the translation of all eight exons (Humphrey 2012). Treatment of ASOs in the SMA mouse model has shown 100% incorporation of exon 7 and a significant increase in SMN expression, but the method of local delivery into the brain and spinal cord needs optimization (Tsia 2012). Furthermore, some scientists argue that local delivery of drugs may not be able to rescue the SMA phenotype (Humphrey 2012, Hoa 2011). The second treatment approach is to use small molecules or drugs that stabilize SMN, such as HDAC, proteasome and phosphatase inhibitors. Although some of these small molecules increased SMN expression in animal models, clinical trials in patients were largely unsuccessful (Xiao 2011). Another promising treatment approach based on animal model results is to use gene therapy with adenoviruses to deliver SMN protein into the brain. A virus-mediated drug AA8 was able to increase lifespan by 880% in SMA mice. One disadvantage of gene therapy is the small chance that the adenovirus DNA vector could be integrated into the patients genomic DNA. Another lab is currently investigating an adenovirus that is mutated so it is unable to integrate into human genomic DNA (Humphrey 2012). Since, the current SMN-dependent drug therapies have not yielded any beneficial results in SMA patients, and SMN

has recently been proposed to affect NMJ formation and axonal RNA transport, recent investigations in SMN-independent drug targets have also begun (Tsai 2012).

1.9 Gemin2 and SMN as the core components of the SMN complex

There is much to be discovered about the specific functions of each of the components of the SMN complex and how each works together to assemble and transport RNPs. The SMN complex is thought to assemble in a systematic process with Gemin2 and SMN being the core components (Battle 2007, Liu 1997). Both *in vivo* and *in vitro* studies suggest a strong and consistent interaction between SMN and Gemin2 relative to the other SMN complex proteins. *In vitro*, Gemin2 has been shown to promote the N terminal self-association and oligomerization of SMN, while SMN has also been reported to stabilize the Gemin2 protein (Owaga 2007, Sarachan 2012) (Figure 1.6). Furthermore, the *in vitro* association of SMN and Gemin2 is extremely stable as it is resistant to 1 M NaCl (Lui 1997). *In vivo* sedimentation experiments of SMN purified from Hela extracts showed that Gemin2 and SMN are the major components present in stable complexes larger then 300kDa (Helmken 2000, Liu 1997,). In addition, *in vivo*, SMN localizes with Gemin2 in both the nucleus and cytoplasm and the stoichiometry of the two proteins has been shown to be important for the formation of neurite RNP granules in motor neurons (Zhang 2006). Because the SMN-Gemin2 interaction has been demonstrated to be crucial for the function of SMN in motor neurons, understanding the SMN-Gemin2 interaction is important for uncovering the molecular mechanism of SMN.

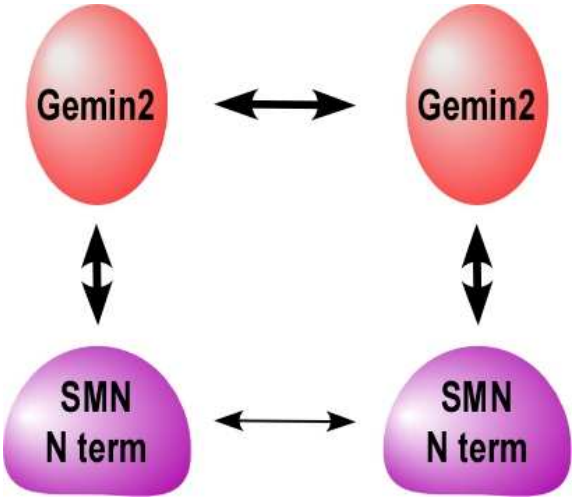


Figure 1.6- The association of SMN and Gemin2
Gemin2 self-association and interaction with the N-terminus of SMN (bold arrows) increases the stability of the SMN N-terminal self-association (light arrow). Figure adapted from Owaga et al 2007.

1.10 Project overview

The aim of this project is to characterize the interaction of the SMN Gemin2 complex and determine differences between SMN WT and SMN D7 with the hope of gaining insights into the molecular mechanism of SMN in the pathology of SMA. Chapter 2 describes the expression and purification of a stable N-terminal deletion Gemin2 protein (N45-G2). N45-G2 was purified via nickel affinity chromatography and identified to be more stable and soluble then the full-length His-Gemin2 protein. Initial crystallization resulted in 2µm crystals. Chapter 3 describes the optimization of the expression and purification of the SMN WT and SMN D7 proteins. It was determined that the SMN WT expression system was leaky and the toxicity of a small amount of soluble SMN protein resulted in a 10-fold decrease of protein expression. The purification and refolding of the SMN protein was optimized, increasing the percentage of

refolded protein, protein purity and protein stability. In chapter 3, size exclusion chromatography, SDS and Native PAGE identified a one-to-one ratio of SMN-G2 binding as well as the oligomeric states of SMN, SMN D7 and N45-G2. Surprisingly, the heterodimer formation of SMN WT-G2 *in vitro* was similar to that of SMN D7-Gemin2. However, SMN WT-G2 *in vitro* favored the formation of high molecular weight oligomers, which was less favorable for the SMN D7-Gemin2 complex.

Chapter 2

EXPRESSION AND PURIFICATION OF GEMIN2

2.1 INTRODUCTION

Recombinant expression of proteins in *E. coli* is the most common technique used to obtain the high concentration of purified protein needed for protein crystallography and other *in vitro* protein characterization techniques. The success of expressing and purifying recombinant proteins is largely dependent on the expression level of the target protein. The expression level is intrinsic to both the vector DNA sequences and the translated amino acid sequence of the target protein. Because of the complexity of the biological dogma that takes DNA to RNA followed by RNA to protein, the correlation between the DNA sequence of the plasmid transformed into *E. coli* and the associated protein expression level is not well understood. Several factors are known to have a major influence on protein expression such as the toxicity level of the protein, RNA hairpin sequences at the 5' and 3' UTR of the protein coding sequence, the stability and folding of the target protein, and rare codons in the protein mRNA sequence (Jana 2005). In order to overcome some of these issues, induction temperature and time are often varied as a way to decrease the rate of transcription and translation as well as employing the use of a variety of *E. coli* cell lines that have been genetically altered to address the above factors that negatively influence protein expression levels. These bacterial cell lines aid in the decrease of protein toxicity, protease activity and leaky expression and increase

the amount of rare tRNAs as well as the folding and stability of proteins (Samuelson 2011).

The aim of this project is to characterize and crystallize the Gemin2 protein with SMN WT and SMN D7 to understand the role of the SMN complex in SMA. Gemin2 was first discovered as an SMN interacting protein 1 (SIP1), with an unknown function and limited homology to the yeast protein Brr1 involved in snRNP biogenesis (Lui 1997). Gemin2 has been produced recombinantly in *E. coli* and purified in small-scale microgram amounts most often with a GST tag (Ogawa 2007, Friesen 2001). To characterize Gemin2 in complex with SMN, we set out to purify large-scale milligram amounts of Gemin2. While several studies of Gemin2 and SMN have been conducted in mammalian cells, very little information is known about the detailed Gemin2 and SMN interaction and/or the role that Gemin2 plays in the SMN complex (Fischer 1997, Jablonka 2001). We expressed and purified several Gemin2 proteins using nickel affinity chromatography. We identified a more stable N-terminal deletion Gemin2 protein that resulted in microcrystals.

2.2 METHODS

2.2.1 Cloning

Full-length *gemin2* was PCR amplified from Hela S3 cDNA with forward and reverse primers containing a 5' NdeI and a 3' XhoI site. PCR products were digested and ligated into pET 28a and pET 30a vectors to create an N-terminal histidine tagged gemin2 and native gemin2 construct. Both native and N-terminal histidine tagged Gemin2 clones were verified by sequencing at the ASU DNA lab and referred to as His-Gemin2 and Native Gemin2. Gemin2 constructs encoding N-terminal deletions of 14 and 45 amino acids were PCR amplified and ligated into the pET 28a vector as above using the same reverse primer and sequence specific forward primers.

2.2.2 Expression

Gemin2 constructs were transformed into Rosetta PlysS *E. coli* and single colonies were inoculated into 15mL LB cultures with 0.5ug/ul of kanamycin and 0.35ug/ul of chloramphenicol for 12-16hrs at 30°C overnight. Overnight 15ml cultures were inoculated into 1.5L LB large cultures containing the same concentration of antibiotics and grown at 30°C until the OD_{600} of the bacterial culture reached approximately 0.8. Cultures were brought to 18° or 25°C by incubation on ice and induced for 12hrs with the addition of 1mM IPTG. The 1.5L cultures were spun down at 6000rpm for 10 minutes and cell pellets were frozen at -20°C for future purification.

2.2.3 Purification

Cells were re-suspended in binding buffer (50mM Sodium Phosphate pH 7.4, 500mM NaCl, 10% Glycerol, 5mM BME) with the addition of DNAse and $MgCl_2$ at a final concentration of 20mM and sonicated at amplitude of 6 for 1 min repeated five times to break open the cells. Sonicated *E. coli* lysate was clarified via centrifugation at 13,000rpm for 30mins at 4°C and the supernatant was incubated with equilibrated nickel beads (GE fast flow 6) at 4°C on an orbital shaker. The bead-lysate mixture was then allowed to settle in a column by gravity (biorad) and the lysate was allowed to flow through. The column was then washed with wash buffer containing either 40 or 80mM imidazole and eluted with binding buffer containing 250mM imidazole. Fractions were monitored by UV absorbance at 280nm and pooled. Purified protein was either dialyzed into a HEPES low salt buffer (20mM HEPES pH 7.5, 200mM NaCl, 5% Glycerol, 5mM BME), Tris Buffer (20mM Tris HCl pH 8.0, 250mM NaCl, 5-20mM BME) or left in elution buffer and stored at 4°C with or with out addition of PMSF or a protease inhibitor cocktail (Sigma) , or flash frozen and stored at -80°C.

2.2.4 SDS PAGE and western blot analysis

Purified proteins were analyzed by running samples heated at 90°C or incubated at 4°C in laemmli buffer with or without BME on a 12% SDS PAGE gel and subsequent staining with coomassie dye. Protein samples were either heated at 90°C for 10 minutes or incubated at 4°C in laemmli buffer with or without BME.. For Western Blot analysis, samples were run 12% SDS PAGE gel and then transferred on ice to a PVDF membrane for 35 minutes at 90V in transfer buffer on ice (Tris Glycine Running Buffer containing 0.01% SDS and 20% Methanol).

Blots were blocked in PBST with 5% BSA or nonfat milk at 4°C for overnight, washed with PBST and then probed with primary (1:3000 anti-His, GE;1:250 anti-His, Santa Cruz sc-8036 or 1:10,000 anti-SMN –MAMSMA 1, Lorson 1998) and secondary (1:10,000 goat anti-mouse HRP, Peirce) antibodies for one hour at room temperature in PBST with 5% nonfat milk. Proteins were detected using ECL (Super signal West Pico Chemiluminescent Substrate KIT, Peirce), visualized via a CCD imager for 30 seconds to 2 minutes and quantitated using Image J software (NIH).

2.2.5 Crystallization

Purified His-Gemin2 and His-N45-Gemin2 were either left in imidazole elution buffer or buffer exchanged into 20mM Tris HCl pH 8.0, 250mM NaCl with or with out 10% glycerol and 5mM BME and concentrated to 1-4mg/ml in a 10kDa or 30kDa MWCO Amicon centrifugal units at 4°C. Protein was added to hanging drop diffusion crystal trays using the Hampton screen HR2-110 and 112 kits at 16°C and room temperature.

2.3 RESULTS AND DISCUSSION

2.3.1 Cloning of Gemin2 constructs

Four gemin2 constructs were created by cloning native gemin2 or N-terminal deletions of gemin2 into pET30 and pET28 vectors (Novagen). The constructs were designed to express and purify recombinant Gemin2 protein in order to characterize Gemin2 and the SMN-Gemin2 complex. A native gemin2 construct and a histidine tagged gemin2 construct were created along with two additional N-terminal deletion gemin2 constructs, which will be discussed in more detail in Section 2.3.6. Gemin2 cDNA was PCR amplified using the appropriate forward (For) and reverse primers (Rev) (Figure 2.1a-c). The G2 For and Rev primers annealed to the 5' and 3' ends of the gemin2 cDNA respectively to produce full-length gemin2 PCR products. For the gemin2 deletion PCR products, the same G2 Rev primer was used in combination with sequence specific forward primers. The G2-N14 For primer annealed to the gemin2 cDNA sequence starting at the DNA sequence that encodes the fifteenth amino acid valine and the G2-N45 For primer annealed to the gemin2 cDNA sequence that encodes the fourty-sixth proline (Figure 2.1b). PCR products were digested with NdeI and XhoI restriction enzymes and ligated into either the pET28a or pET30a vectors (Figure 2.2a-d). The pET28a vector has an N-terminal histidine tag whereas the pET30a vector does not. Colonies grown in cultures and purified plasmids from each colony were screened by restriction digest to determine if the gemin2 insert was ligated into the pET vectors. Positive clones were verified by sequencing at the ASU DNA lab. We successfully produced four gemin2

constructs, two full-length and two with N-terminal deletions that were then used to express recombinant Gemin2 protein (Figure 2.2a-d). Of note, is that both the gemin2 deletion primers had mismatched nucleotides between the primer sequence and the gemin2 cDNA sequence as shown by the nucleotides that are not underlined (Figure 1.2a). The final sequences did not show any mutations from these mismatches, however, these primers would not be ideal for future cloning of gemin2.

Primer Name	Primer Sequence
G2 For	5’-gggaattcCATA**TG**cgccgagcggaactggctggtt-3’
G2 Rev	5’-gatCTCGAG**tcaagatggctcatcagctaaatcacg**-3’
N14-G2 For	5’-ggaattccCATATG**gtaccagcggagtccgcagtgga**gt-3’
N45-G2 For	5’-gggaattcCATATG**ccgaggacgcctcaggaataccctga**-3’

a

Gemin2 Sequence

```
M  R  R  A   E  L  A   G  L  K   T  M  A  W   V  P  A   E  S  A   V  E  E  L
ATGCGCCGAG CGGAACTGGC TGGTTTGAAA ACCATGGCGT GGGTACCAGC GGAGTCCGCA GTGGAAGAGT
TACGCGGCTC GCCTTGACCG ACCAAACTTT TGGTACCGCA CCCATGGTCG CCTCAGGCGT CACCTTCTCA
  M  P  R   L  L  P   V  E  P  C   D  L  T   E  G  F   D  P  S  V   P  P  R
TGATGCCTCG GCTATTGCCG GTAGAGCCTT GCGACTTGAC GGAAGGTTTC GATCCCTCGG TACCCCCGAG
ACTACGGAGC CGATAACGGC CATCTCGGAA CGCTGAACTG CCTTCCAAAG CTAGGGAGCC ATGGGGGCTC
 T  P  Q   E  Y  L  R   R  V  Q   I  E  A   A  Q  C  P   D  V  V   V  A  Q
GACGCCTCAG GAATACCTGA GGCGGGTCCA GATCGAAGCA GCTCAATGTC CAGATGTTGT GGTAGCTCAA
CTGCGGAGTC CTTATGGACT CCGCCCAGGT CTAGCTTCGT CGAGTTACAG GTCTACAACA CCATCGAGTT
I  D  P  K   K  L  K   R  K  Q   S  V  N  I   S  L  S   G  C  Q   P  A  P  E
ATTGACCCAA AGAAGTTGAA AAGGAAGCAA AGTGTGAATA TTTCTCTTTC AGGATGCCAA CCCGCCCCTG
TAACTGGGTT TCTTCAACTT TTCCTTCGTT TCACACTTAT AAAGAGAAAG TCCTACGGTT GGGCGGGGAC
  G  Y  S   P  T  L   Q  W  Q  Q   Q  Q  V   A  Q  F   S  T  V  R   Q  N  V
AAGGTTATTC CCCAACACTT CAATGGCAAC AGCAACAAGT GGCACAGTTT TCAACTGTTC GACAGAATGT
TTCCAATAAG GGGTTGTGAA GTTACCGTTG TCGTTGTTCA CCGTGTCAAA AGTTGACAAG CTGTCTTACA
 N  K  H   R  S  H  W   K  S  Q   Q  L  D   S  N  V  T   M  P  K   S  E  D
GAACAAACAT AGAAGTCACT GGAAATCACA ACAGTTGGAT AGTAATGTGA CAATGCCAAA ATCTGAAGAT
CTTGTTTGTA TCTTCAGTGA CCTTTAGTGT TGTCAACCTA TCATTACACT GTTACGGTTT TAGACTTCTA
E  E  G  W   K  K  F   C  L  G   E  K  L  C   A  D  G   A  V  G   P  A  T  N
GAAGAAGGCT GGAAGAAATT TTGTCTGGGT GAAAAGTTAT GTGCTGACGG GGCTGTTGGA CCAGCCACAA
CTTCTTCCGA CCTTCTTTAA AACAGACCCA CTTTTCAATA CACGACTGCC CCGACAACCT GGTCGGTGTT
  E  S  P   G  I  D   Y  V  Q  I   G  F  P   P  L  L   S  I  V  S   R  M  N
ATGAAAGTCC TGGAATAGAT TATGTACAAA TTGGTTTTCC TCCCTTGCTT AGTATTGTTA GCAGAATGAA
TACTTTCAGG ACCTTATCTA ATACATGTTT AACCAAAAGG AGGGAACGAA TCATAACAAT CGTCTTACTT
 Q  A  T   V  T  S  V   L  E  Y   L  S  N   W  F  G  E   R  D  F   T  P  E
TCAGGCAACA GTAACTAGTG TCTTGGAATA TCTGAGTAAT TGGTTTGGAG AAAGAGACTT TACTCCAGAA
AGTCCGTTGT CATTGATCAC AGAACCTTAT AGACTCATTA ACCAAACCTC TTTCTCTGAA ATGAGGTCTT
L  G  R  W   L  Y  A   L  L  A   C  L  E  K   P  L  L   P  E  A   H  S  L  I
TTGGGAAGAT GGCTTTATGC TTTATTGGCT TGTCTTGAAA AGCCTTTGTT ACCTGAGGCT CATTCACTGA
AACCCTTCTA CCGAAATACG AAATAACCGA ACAGAACTTT TCGGAAACAA TGGACTCCGA GTAAGTGACT
  R  Q  L   A  R  R   C  S  E  V   R  L  L   V  D  S   K  D  D  E   R  V  P
TTCGGCAGCT TGCAAGAAGG TGCTCTGAAG TGAGGCTCTT AGTGGATAGC AAAGATGATG AGAGGGTTCC
AAGCCGTCGA ACGTTCTTCC ACGAGACTTC ACTCCGAGAA TCACCTATCG TTTCTACTAC TCTCCCAAGG
 A  L  N   L  L  I  C   L  V  S   R  Y  F   D  Q  R  D   L  A  D   E  P  S
TGCTTTGAAT TTATTAATCT GCTTGGTTAG CAGGTATTTT GACCAACGTG ATTTAGCTGA TGAGCCATCT
ACGAAACTTA AATAATTAGA CGAACCAATC GTCCATAAAA CTGGTTGCAC TAAATCGACT ACTCGGTAGA
*
TGA
ACT
```

b

PCR Primers	Vector	Final Construct
G2 For, G2 Rev	pET 28a	His-Gemin2
G2 For, G2 Rev	pET 30a	Native Gemin2
N14-G2 For, G2 Rev	pET 28a	His-N14-Gemin2
N45-G2 For, G2 Rev	pET 28a	His-N45-Gemin2

c

Figure 2.1- Cloning of gemin2 constructs

Gemin2 cDNA was PCR amplified with the appropriate primers containing a 5’ NdeI and 3’ XhoI site shown in Italics. PCR products were then digested and cloned into the designated digested pET vectors to create the four Gemin2 contructs. (a) Four primer sequences with NdeI and XhoI restriction sites are indicated by CAPS and the underlined bold sequence indicates the portion of the

primer that annealed to the gemin2 cDNA for PCR amplification. (b) The gemin2 cDNA sequence with the encoded amino acid sequence above. Primer annealing sites are indicated with bold underlined text. (c) The forward and reverse primers used for each PCR reaction are listed with the corresponding vector each PCR product was cloned into and the resulting final DNA construct.

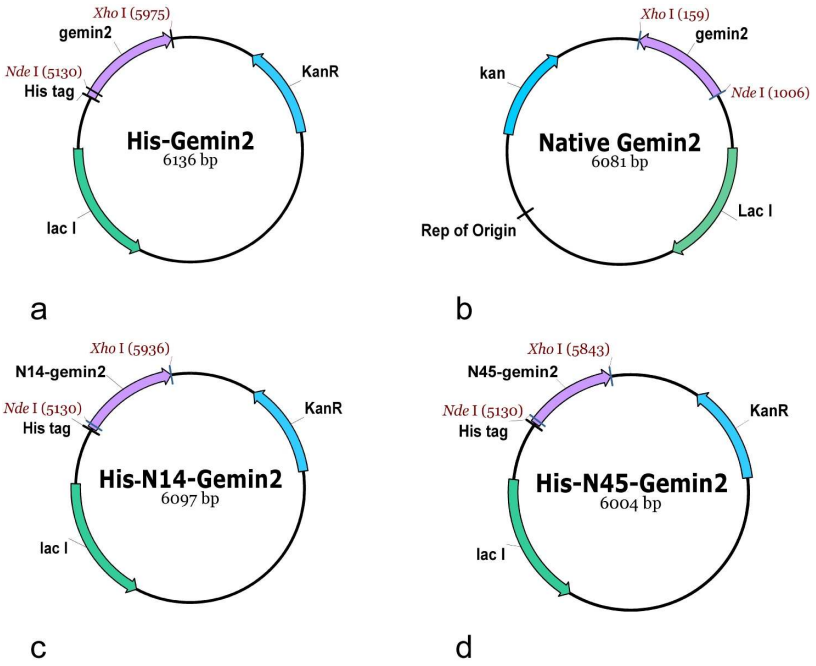


Figure 2.2- Vector maps of gemin2 constructs
The *gemin2* gene was cloned into the various pET vectors with NdeI and XhoI restriction sites (italics) to create the DNA plasmid constructs. The pET vectors contained a kanamaycin resistance gene (blue), the *lac I* gene (green), and a Histidine tag (a,c,d). (a) Histidine tagged gemin2 (b) Native gemin2 (c) Histidine tagged gemin2 encoding a N terminal 14 amino acid deletion and (d) Histadine tagged gemin2 encoding a N-terminal 45 amino acid deletion.

2.3.2 Expression optimization of His-Gemin2

The verified His-Gemin2 clone (Figure 2.2a) was transformed into Rosetta PlysS cells because the amino acid sequence contains both a high percentage of rare codons as well as several successive groups of rare codons that are known to inhibit the soluble expression of recombinant proteins in *E. coli* (Kane, 1995). His-Gemin2 in Rosetta cells was grown at 37°C and then induced with IPTG for

various lengths of time and at various temperatures to determine the optimal expression conditions. After several unsuccessful expression conditions yielding less then five milligrams of soluble protein, optimal expression of His-Gemin2 was achieved by reducing the induction temperature to 18°C for 12 hrs (Figure 2.3a). These conditions were determined by a recently published paper (Takizawa 2010) and produced a 10 -fold increase of expression compared to other similar low temperature expression conditions (Figure 2.3a). The major difference between the previously attempted low temperature induction at 16°C for 16hrs and the successful high expression conditions is that the bacteria culture was grown at 30°C instead of 37°C during all growth conditions prior to induction (Figure 2.3b). The difference in growth temperature is likely attributed to the additional rare codon tRNAs in Rosetta PlysS *E. coli* strain. This is supported by the fact that several publications using Rosetta PlysS use similar growth temperatures prior to induction and is indicative that lower temperature growth prior to induction promotes optimal conditions for the expression, stability and/or function of the rare tRNAs encoded by the PlysS plasmid.

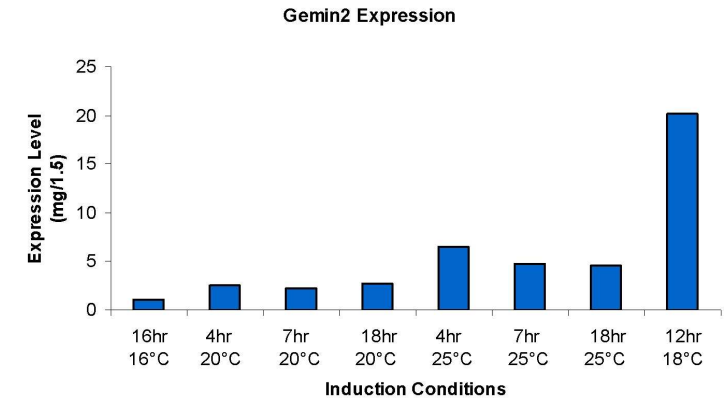
2.3.3 Purification of His-Gemin2 using nickel affinity chromatography

Over-expressed His-Gemin2 induced at 18°C was run on a nickel affinity column to purify it from the remaining *E. coli* lysate proteins. Samples were collected during each purification step and analyzed via SDS PAGE to determine if the purification was successful. The large 34kDa protein band in the elution sample but not in the flow through and wash samples showed that His-Gemin2 was efficiently purified during chromatography (Figure 2.3b). The other protein

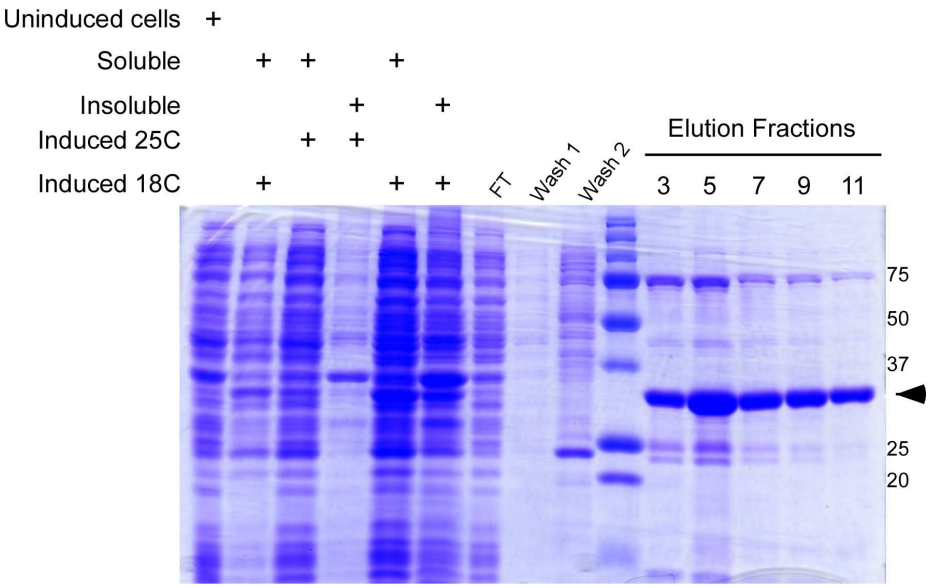
band seen at 75kDa was likely an impurity that bound to the column. This was further confirmed by the observation that increasing the imidazole concentration in the wash buffer from 50mM to 75mM led to a significant decrease of the 75kDa band in the eluted protein fractions (data not shown).

To verify that the 34kDa protein observed in SDS PAGE was in fact His-Gemin2, samples were blotted and detected via an anti-histidine antibody with His-GFP as a positive control. The blot indicates that the major purification product at 34kDa is in fact His-Gemin2 (Figure 2.3c). In addition, the blot confirms the optimal induction temperature for His-Gemin2 at 18°C.

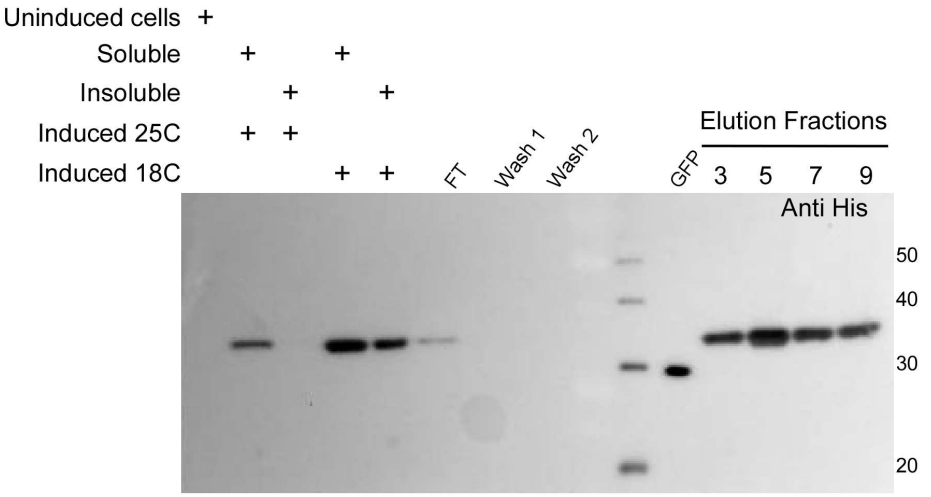
His-Gemin2 was found to be optimally expressed by growing Rosetta bacteria cells at 30°C and inducing for 12hrs at 18°C. In addition, the 34kDa His-Gemin2 was successfully purified to greater then 80% purity via nickel affinity chromatography.


Gemin2 Expression
Expression Level (mg/1.5)
25
20
15
10
5
0
16hr 16°C
4hr 20°C
7hr 20°C
18hr 20°C
4hr 25°C
7hr 25°C
18hr 25°C
12hr 18°C
Induction Conditions

a


Uninduced cells +
Soluble + + +
Insoluble + +
Induced 25C + +
Induced 18C + + +
FT
Wash 1
Wash 2
Elution Fractions
3 5 7 9 11
75
50
37
25
20

b


Uninduced cells +
Soluble + +
Insoluble + +
Induced 25C + +
Induced 18C + +
FT
Wash 1
Wash 2
GFP
Elution Fractions
3 5 7 9
Anti His
50
40
30
20

c

Figure 2.3- Expression and purification of His-Gemin2
(a) His-Gemin2 was induced from 16-25°C for 4-18hrs, *E. coli* samples were lysed and the Gemin2 expression level was determined by analyzing the soluble *E. coli* lysate on an SDS PAGE followed by a western blot using an anti-his tag antibody. Expression levels of soluble His-Gemin2 in (mg/1.5L culture) are shown relative to a control His-GFP protein in each blot. (b) Large-scale expression of *E. coli* containing the His-Gemin2 construct expressed either at 18°C for 12hrs or 25°C for 4hs (left portion of the gel). The plusses indicate whole cell induced and uninduced samples as well as soluble and insoluble lysate samples. The 18°C soluble *E. coli* lysate was purified via nickel affinity chromatography (right portion of gel). The nickel purification samples include flowthrough, wash 1 of binding buffer, wash 2 of 50mM imidazole and elution fractions 3-9. The arrow shows the band corresponding to His-Gemin2. (c) The expression and purification samples in (b) were analyzed by western blot anlaysis using an anti-his tag antibody. His-GFP was used as a positive control.

2.3.4 Solubility of His-Gemin2 protein

His-Gemin2 in elution buffer precipitated out of solution at 4°C. In order to increase the solubility of purified His-Gemin2 for future characterization, His-Gemin2 was immediately dialyzed after purification into several buffers containing a range of concentrations of salt and beta-mercapetoethanol (BME). Upon dialysis into HEPES buffer including a high concentration of BME, a solubility limit of 1.0 OD_{280} was obtained (Table 2.1). Previous studies have shown that the majority of recombinant Gemin2 is in the dimer form with a small portion in the trimer and monomer forms (Ogawa 2007). The higher BME concentration may have increased the solubility of the protein by reducing disulfide bonds and decreasing aggregation.

Buffer	**OD_{280}**
Elution buffer (500mM NaCl, 250mM imidizole)	NA (all crashes out)
20mM Tris pH 8.0	0.50
20mM HEPES pH 7.5	0.40
20mM HEPES (no salt, no glycerol)	NA (all crashes out)
20mM HEPES pH 7.5 (20mM BME)	1.00

Table 2.1- Solubility limit of His-Gemin2
His-Gemin2 protein was dialyzed into buffers containing 10% Glycerol, 200mM NaCl and 5mM BME or left in elution buffer at 4°C overnight. Changes to buffer conditions are indicated in parentheses. After dialysis, the protein was centrifuged and the remaining amount of soluble protein was determined by measuring the absorbance at 280nm.

2.3.5 The degradation of His-Gemin2

Although the HEPES buffer increased the solubility of His-Gemin2, after several days at 4°C His-Gemin2 continued to crash out of solution and appeared to undergo degradation. In an effort to inhibit the degradation of His-Gemin2, the protease inhibitor PMSF was added to His-Gemin2 and samples of His-Gemin2 with and without protease inhibitor were monitored for degradation over several days at 4°C. Even in the presence of PMSF, His-Gemin2 was degraded into several smaller molecular weight bands, which started at day three of incubation. In addition, His-Gemin2 continued to precipitate out of solution over time (Figure 2.4a).

The same stability test was repeated with a commercial protease inhibitor cocktail. The degradation of His-gemin2 occurred in the presence of PMSF as well as a protease inhibitor cocktail (2.4b). The two main degradation fragments at approximately 32kDa and 30kDa were calculated to have a loss of 1.50kDa and

5.01kDa respectively excluding the histidine tag and linker sequence. A western blot of degraded His-Gemin2 was run to give insight into the location of the protein cleavage. The anti-histidine antibody only detected the full-length His-Gemin2 protein suggesting that the degraded His-Gemin2 fragments contained at least an N-terminal deletion (Figures 2.4 b-c). It may be possible that the His-Gemin2 fragments were not detected due to both a lower concentration and/or because of inefficient transfer due to their lower molecular weights, however, the detection of the 29kDa His-GFP protein makes the later unlikely.

His-Gemin2 was purified to twenty milligram per liter amounts, but precipitated out of solution and degraded after several days even in the presence of protease inhibitors. Due to the lack of solubility and stability, we concluded that His-Gemin2 was not an ideal protein for crystallography studies and further biochemical analysis.

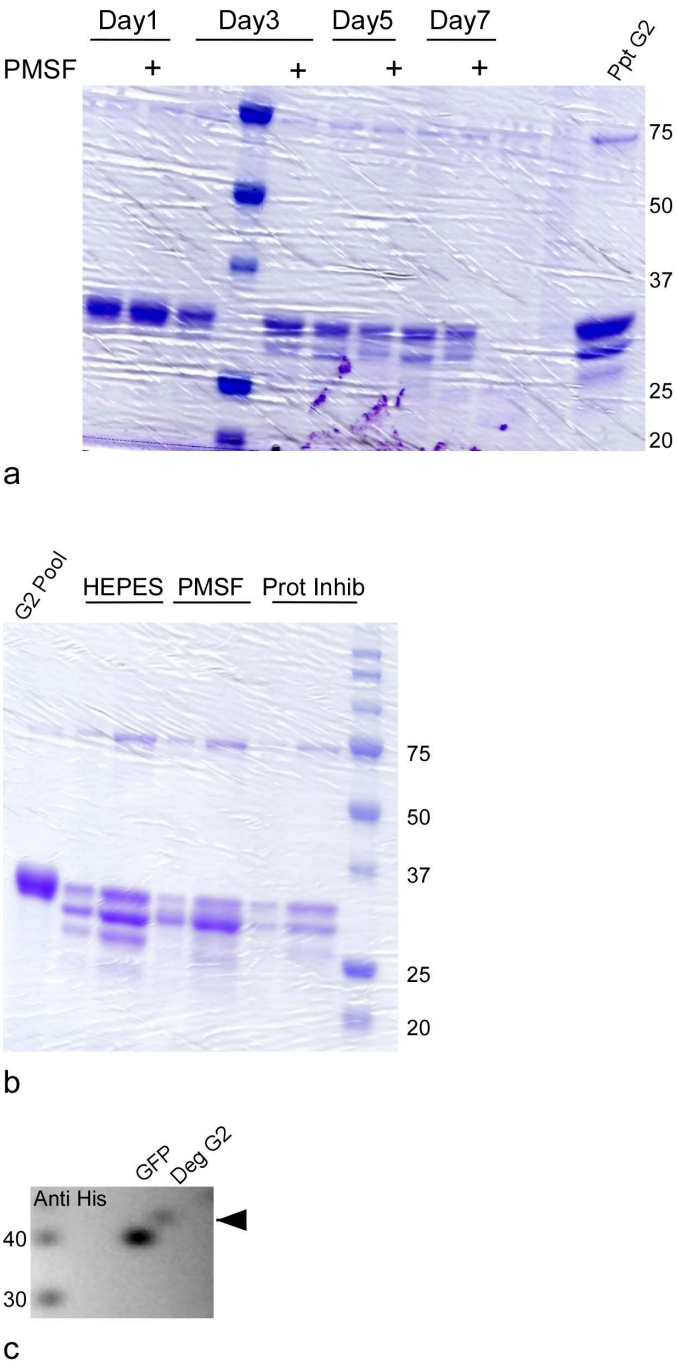


Figure 2.4- Stability of His-Gemin2
(a) His-Gemin2 in HEPES buffer with or without PMSF was incubated at 4°C for 7 days and aliquots were spun down and the soluble protein for each aliquot along with the precipitated protein (ppt) was analyzed by SDS PAGE for degradation. (b) His-Gemin2 was incubated at 4°C for 5 days in HEPES buffer only or with the addition of PMSF and a protease inhibitor cocktail from Sigma and analyzed for degradation by SDS PAGE. (c) Degraded His-Gemin2 in HEPES buffer from (b)

was probed via an anti-his tag antibody along with the control His-GFP. The arrow indicates the full-length His-Gemin2.

2.3.6 Predicted His-Gemin2 cleavage sites and the design of cleavage-resistant Gemin2 constructs

A secondary structure analysis of the N and C termini of His-Gemin2 was conducted to aid in the prediction of His-Gemin2 cleavage sites. The analysis predicted three alpha helices and a beta sheet at the N-terminus and four longer alpha helices at the C-terminus (Figure 2.5a). The location of the cleavage sites were estimated by using the calculated molecular weight of the two main degradation fragments observed in Section 2.3.5 and subtracting that molecular weight from full-length Gemin2. The estimated molecular weight loss of 1.05kDa and 5.01kDa for the two major degradation fragments was then mapped onto the secondary structure of Gemin2 at the N and C termini as the possible cleavage sites. A sequence analysis to identify a protease that was specific for each of the four possible cleavage sites did not yield any significant results. The N-terminal cleavage sites were predicted to be the actual cleavage or degradation sites because they are located in regions lacking defined alpha helical and beta sheet secondary structures while the C-terminal predicted cleavage sites are located in the middle and at the end of an alpha helix (2.5a). Furthermore, the confidence level of the first two N-terminal alpha helices is low relative to the alpha helices at the C-terminus suggesting that the probability of those alpha helices forming in the actual structure is less likely. An *in vivo* study showed that deletion of the first 45 amino acids of Gemin2 had no affect on the interaction of

SMN and Gemin2 (Ogawa 2007). The N-terminal cleavage sites located within predicted random coiled secondary structure, the lack of detection of the N-terminal histidine tag and the lack of SMN binding *in vivo* provide evidence that the N-terminus of His-Gemin2 was degraded *in vitro*. However, the possibility that His-Gemin2 was cleaved at both the N and C termini cannot be ruled out.

In order to create a more stable and soluble Gemin2, two new N terminal deletion proteins were cloned in which 14 and 45 amino acids corresponding to the two predicted cleavage sites were deleted from the N-terminus (Figure 2.5b).

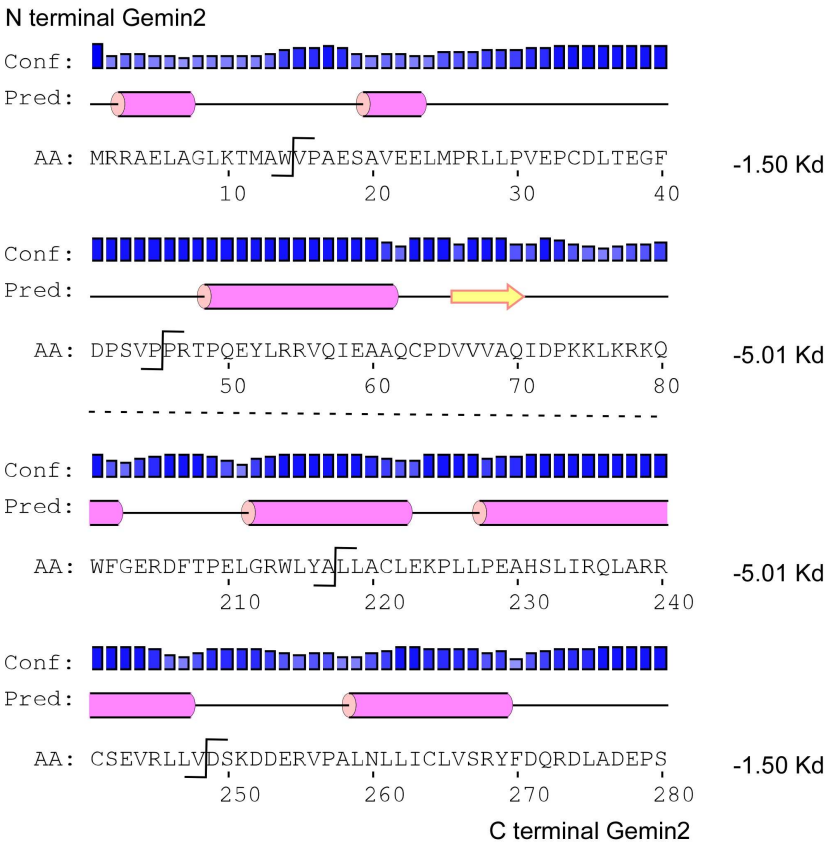


Figure 2.5- Predicted cleavage sites and Gemin2 deletion constructs
(a) The Gemin2 N and C terminal sequences were analyzed via a secondary structure prediction software (Expasy molecular weight tool. Alpha helices (pink) and beta sheets (yellow) predicted with a confidence level (blue) by PSIPRED protein structure prediction server. The lines show the predicted cleavage sites on the Gemin2 sequence that correlate to the two main cleavage fragments (loss of 1.50 and 5.01kDa) seen in SDS PAGE. (b) The 14 and 45 amino acid N-terminal deletion constructs were designed and cloned to create His-N14-Gemin2 and His-N45-Gemin2. The N-terminal sequence of full-length His-G2 is shown below the sequences of the two deletion constructs with the histidine tag red.

2.3.7 Gemin2 deletion construct expression and purification

The two deletion proteins, His-N14-Gemin2 and His-N45-Gemin2, were expressed and purified as was previously done with His-Gemin2 in sections 2.3.2-3 and will be referred to as N14-Gemin2 and N45-Gemin2. Elution fractions from nickel affinity purification showed bands that migrated at the expected size of N14-Gemin2 and N45-Gemin2 of approximately 32kDa and 30kDa respectively, which also migrated the same distance as the two degraded fragments shown previously (Figure 2.6a). The lack of an over-expression band in the soluble lysate along with the lower amount of purified protein showed that the expression of the Gemin2 deletion proteins is significantly less then the full-length His-Gemin2 with the N45-Gemin2 protein expressing at a higher level then N14-Gemin2.

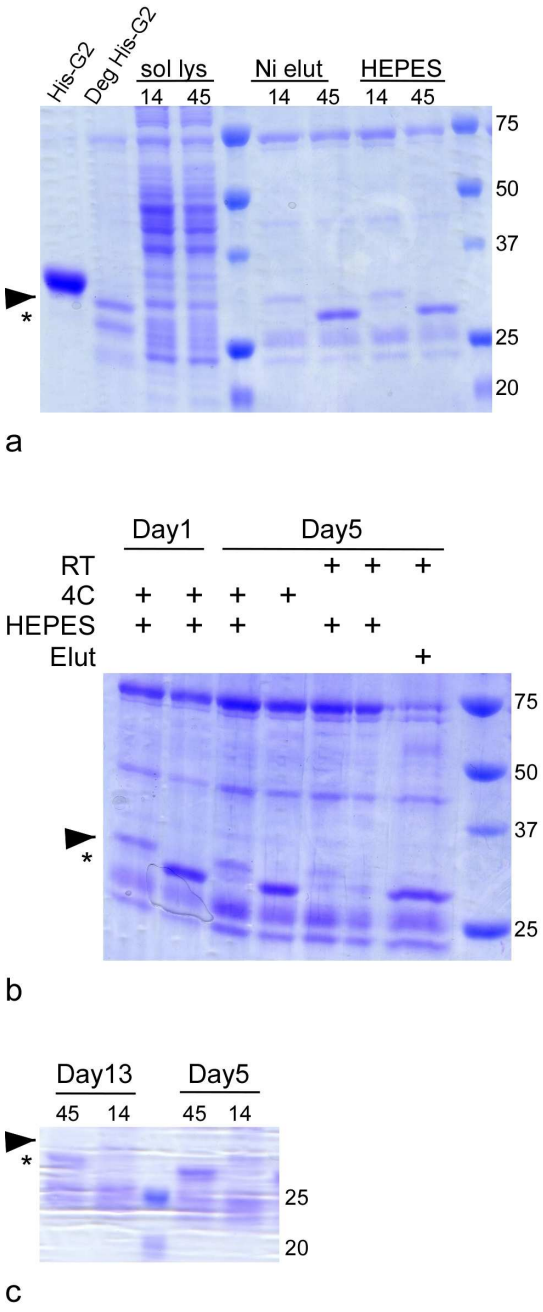


Figure 2.6- Expression and stability of N45-Gemin2
(a) Deletion Gemin2 proteins were expressed and purified similarly to His-Gemin2. Proteins were dialyzed into HEPES buffer after purification and analyzed by SDS PAGE. (b) The stability of protein in HEPES and elution buffer was analyzed at 1 and 5 days after incubation at 4 and 25°C. (c) The stability of the two proteins was analyzed after 5 and 13 days in HEPES buffer. Asterisks indicate the expected molecular weight of the N14-Gemin2 protein and arrows indicate the N45-Gemin2 protein

2.3.8 Comparison of N14-Gemin2 and N45-Gemin2 protein stability

In order to test the stability of these two constructs, each protein was incubated in HEPES or elution buffer at 4 or 25°C for five days and degradation and precipitation were analyzed by centrifugation followed by SDS PAGE. After five days, each of the proteins in HEPES buffer did not precipitate out of solution while N14-Gemin2 underwent slight degradation, as the 25kDa band appears to be more intense (Figure 2.6b). Both proteins were degraded at 25°C in HEPES buffer, but N45-Gemin2 in elution buffer was unexpectedly still present at 25°C. Unlike full-length His-Gemin2, the deletion Gemin2 proteins were more stable in elution buffer then HEPES buffer and the stability at room temperature points to an overall increase in protein stability.

Proteins were incubated at 4°C for over two weeks to determine if the proteins would be stable enough for crystallography trials. After 13 days in HEPES buffer, both proteins were still present, but N45-Gemin2 appeared slightly more stable then N14-Gemin2 as indicated by a larger degradation smear at 25kDa in the N14-Gemin2 sample (Figure 2.6c).

N45-Gemin2 and N14-Gemin2 were successfully expressed and purified. The deletion Gemin2 proteins had lower expression and purification yield relative to His-Gemin2, but the deletion proteins were more stable and soluble. N45-Gemin2 was used for further crystallizing and biochemical characterization experiments because of its higher stability and expression level.

2.3.9 Optimization of N45-Gemin2 stability

N45-Gemin2 was buffer exchanged and concentrated in HEPES and elution buffers to determine the optimal buffer conditions for future experiments. As was predicted from the previous section, N45-Gemin2 is more stable in elution buffer as all the protein in the HEPES buffer was degraded (Figure 2.7a). Either, the increased amount of salt, the buffer composition or possibly the presence of imidazole may stabilize N45-Gemin2. N45-Gemin2 in high salt buffer may favor stable monomers or oligomers instead of aggregation. It has been shown that imidazole stabilizes histidine tag proteins and sometimes facilitates crystallization, so it may be possible that imidazole aids in the solubility and/or stabilization of N45-Gemin2 (Hamilton 2003). It was concluded that elution buffer resulted in a more stable N45-Gemin2 protein.

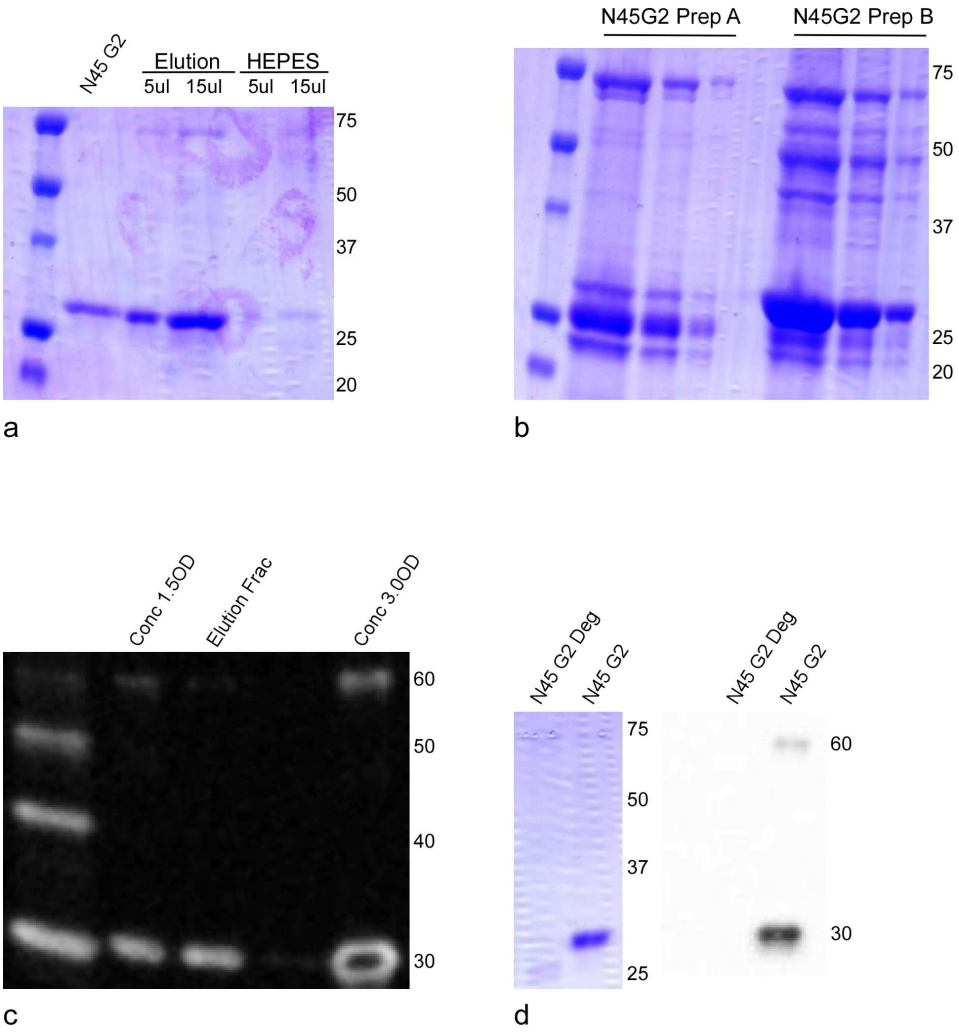


Figure 2.7- Stability of N45-Gemin2

(a) N45-Gemin2 was incubated in elution buffer or HEPES buffer at 4°C for several days and analyzed for stability. (b) Two N45-Gemin2 purified protein samples from different preps (A and B) were concentrated to 1.5 and 3.0 OD_{280} respectively and analyzed for stability. Each prep was purified using the same conditions except that prep B was purified with an optimized imidazole wash buffer of 80mM imidazole whereas prep A used 50mM imidazole. (c). Western blot analysis of prep A, prep B and an elution fraction were transferred to a blot and probed with an anti-his tag antibody. (d) Degraded N45-Gemin2 from prep A and full-length N45-Gemin2 from prep B were run on SDS PAGE and analyzed via western blot (anti-his tag).

Next, N45-Gemin2 was concentrated to ascertain whether the appropriate concentration range and purity for crystallography and biochemical analysis could be obtained. N45-Gemin2 from two separate preparations, prep A, prior to nickel

affinity optimization and prep B, after the optimization of imidazole concentration, was concentrated and analyzed via OD_{280} and SDS PAGE. Prep A and B were concentrated to an OD_{280} of 1.5 and 3.0 respectively (2.4 and 1.2mg/ml). Protein in prep A was degraded from a 30kDa band to a major band at 25kDa while prep B did not (Figure 2.7 b). Also of note, is that the more concentrated N45-Gemin2 in prep B had several higher molecular weight impurities that would not be ideal for crystallization. The degradation of N45-Gemin2 in prep A was likely due to the lower purity of the protein sample. Another possibility is that the degradation of N45-Gemin2 from prep A, which was stored at 4°C for several weeks longer then prep B, resulted from degradation during storage. Interestingly enough, degradation did not occur or occurred at a slower rate in dilute N45-Gemin2 samples suggesting that concentration promoted degradation possibly due to a high concentration of co-purified proteases (data not shown). N45-Gemin2 was able to be concentrated to 2.4 mg/ml, which was higher then the 1.6 mg/ml of His-Gemin2, but concentrated N45-Gemin2 degraded slowly over a several week time span and still contained some impurities. The inefficient concentration and degradation of N45-Gemin2 was not ideal, but determined to be sufficient for future characterization and/or purification steps.

Before characterization of N45-Gemin2, a western blot using an anti-histidine antibody was conducted to confirm the presence of the N45-Gemin2 protein. Full-length N45-Gemin2 protein was detected at the expected molecular weight of 30kDa (Figure 2.7c). In addition, a small portion of N45-Gemin2

migrated as a 60kDa SDS-resistant dimer in the concentrated N45-Gemin2 sample. This suggests that the self-association of Gemin2 is relatively strong as a percentage of intermolecular disulfide bonds were likely still present under SDS PAGE conditions. The observed dimer detected by western blot analysis as well as the fact that Gemin2 was able to be concentrated in a 30kDa molecular weight cut off concentrator without protein loss (data not shown) provided evidence that N45-Gemin2 retained the ability to self-associate. Since, Gemin2 self-association is important for SMN-self association, we predicted that N45-Gemin2 should be able to bind SMN and mimic the native SMN-Gemin2 interactions *in vitro*.

Concentrated N45-Gemin2 continued to degrade during incubation at 4°C after two weeks. Degraded and full-length N45-Gemin2 were analyzed via SDS PAGE and Western Blot to determine if the degradation occurred at the N or C termini. Full-length and degraded N45-Gemin2 are shown at 30kDa and 25kDa respectively on SDS PAGE, but the western blot with an anti-histidine antibody only detected the full-length protein at 30kDa (Figure 2.d). This provides evidence for further N-terminal degradation of the N45-G2 protein over time and/or upon concentration.

In conclusion, N45-Gemin2 was more stable and soluble then His-Gemin2. N45-Gemin2 likely retained the native Gemin2 function of self-association, which is important for future SMN-Gemin2 interaction studies. The N45-Gemin2 protein underwent the same degradation as full-length Gemin2 at a slower rate. The continual degradation of all three Gemin2 constructs may be caused by an unstructured N-terminal region of Gemin2. The recently published

crystal structure of Gemin2 supports this as the first 45 amino acids are highly disordered and serve to fill space in the RNA binding pocket of the Sm proteins D1, D2, E, F and G (Zhang 2011). Furthermore, residues from 45 to 90 in the crystal structure are disordered, but are bound to Sm proteins, suggesting that this interaction stabilizes Gemin2. Another recent paper showing the solution structure of Gemin2 bound to an alpha helix of SMN protein containing residues 26-51 reported that even the small alpha helix region of SMN stabilized the flexible Gemin2 protein (Sarachan 2012). Other recombinant Gemin proteins and Sm proteins show similar degradation patterns *in vitro*, Gemin6, Gemin7 and several Sm proteins were all purified and/or co-expressed with a binding partner (Ma 2005, Zhang 2012). This suggests that in the absence of a binding partner the sm-protein fold of Gemins and Sm proteins may be disordered and thus easily degraded. It is likely that these proteins interact immediately after translation *in vivo* and this interaction could even be part of the regulation of the SMN complex. In agreement with the idea that cleavage and degradation is crucial for the regulation of the SMN complex, Gemin2 and SMN have both been shown to be cleaved by Calpain (Fuentas 2010).

2.3.10 Crystallization of N45-Gemin2

In an attempt to determine the molecular structure of N45-Gemin2, crystallization trials of purified N45-Gemin2 were conducted. Concentrated N45-Gemin2 was screened using the Hampton crystal screen and three conditions produced crystals (Table 2.2). Microcrystals observed in condition 16 are shown in Figure 2.8.

Crystal Screen- #	Well reagents	Temperature
HR2-110 -#16	0.1M HEPES pH 7.5, 1.5M lithium sulfate	16 and 25°C
HR2-110 -#17	0.2M Lithium sulfate, 0.1M Tris HCl pH 8.5, 30% PEG 4000	16°C
HR2-112 -#41	0.01M Nickel (II) Chloride hexahydate, 0.1M Tris HCl pH 8.5, 1.0M Lithium sulfate	16 and 25°C

Table 2.2- Crystallization conditions for N45-Gemin2
Hampton Screen HR2-110 and 112 was used to screen for optimal crystal conditions. The three conditions listed resulted in 2um crystals, all of which contained lithium sulfate as a comment reagent.

Figure 2.8- Crystallization of N45-Gemin2
N45-Gemin2 was crystallized in 0.1M HEPES pH 7.5, 1.5M lithium sulfate using the hanging drop method. Crystals were approximately 2 µm in size.

Crystals of approximately 2 µm were observed using the hanging drop method at room temperature in approximately one day (Figure 2.8). Further attempts to

optimize these initial conditions were unsuccessful. In the process of optimizing the conditions to produce larger crystals, the crystal structure for Gemin2 was published (Zhang 2011). Further optimization of Gemin2 crystals was abandoned and focus was turned towards investigation of the interaction between SMN and Gemin2 proteins.

Chapter 3

EXPRESSION AND PURIFICATION OPTIMIZATION OF SMN

3.1 INTRODUCTION

Since the 1990's when SMN was identified as the protein link to spinal muscular atrophy, intense efforts have been made to understand its function in motor neurons. Biochemical techniques such as crystallography and NMR that give structural clues about the function of proteins have been unsuccessful with full-length recombinant SMN protein. The structure of the Tudor domain of SMN was solved showing the interaction of SMN with Sm-proteins via their arganine-glycine tails. Interestingly, the Tudor domain adopted a partial Sm-fold similar to Sm proteins and Gemins 6 and 7 (Selenko 2001).

Several labs have expressed SMN as a recombinant protein in *E. coli* with a wide variety of fusion tags, but the majority of the protein is insoluble and the small amount of soluble protein forms large oligomers (Takaku 2011,Pellizzoni 1999). SMN with an N-terminal thioredoxin tag and a C-terminal histidine tag (SMN TRX) was refolded from inclusion bodies using dialysis and determined to form large oligomers or soluble aggregates greater then 600kDa by size exclusion chromatography (SEC) (Young 2000b, Nguyen thi 2008). Similarly, Zhang et al was able to co-express soluble recombinant SMN with the Gemin2 protein, but stated that the purified SMN-Gemin2 proteins failed to form protein crystals due to the heterogeneity of the proteins observed on SEC (Zhang 2011). Tukaku et al created a fusion protein by linking SMN to Gemin2 with an amino acid sequence and was able to produce a soluble SMN-Gemin2 fusion protein (Tukaku 2011).

Their results showed that the SMN-Gemin2 fusion protein restored function in SMN knockout chicken cells. However, because the function of chicken SMN has been shown to be different then human SMN, these results may not provide direct evidence that the fusion SMN-Gemin2 protein would change the structure and/or function of the human proteins (Wang 2001).

The majority of the biochemical insights about the SMN complex were determined by experiments in mammalian cells or rabbit raticulocyte lysate (RRL). This suggests that post-translational modifications, chaperone proteins and post-translational binding to other proteins stabilize the SMN protein. In agreement with this is the observation that the phosphorylation state of SMN was crucial for *in vivo* localization and binding to other proteins (Petri 2007). The disadvantage of using partially purified proteins from RRL or mammalian cells is that the protein of interest is bound to several other proteins, which makes the biochemical analysis more complex and indirect. In addition, these techniques are not ideal for expressing the large amounts of pure protein needed for crystallographic and other biochemical studies.

SMN WT has been previously expressed, purified and crystallized in our lab (Seng, unpublished). The following observations led to the optimization of the SMN expression and purification protocol. First, the cell growth of the *E. coli* containing the SMN WT construct was two-fold slower then the exponential growth normally observed in recombinant expressing Bl21 cells. The growth was restored after transformation of the SMN WT plasmid into fresh competent Bl21 cells. These observations indicate that SMN is toxic to the cells even prior to

induction. Second, over-expression of SMN WT at the predicted size of 38kDa-40kDa was not observed after induction, but upon the addition of glucose into the growth media, a small amount of over-expressed protein at 34kDa was observed (Campanaro 2009). Glucose is known to inhibit leaky expression in *E. coli* strains using the lac operon for induction and the over-expressed protein at 34kDa indicates that SMN WT migrates at 34kDa in SDS PAGE. After purification and refolding with nickel affinity chromatography, several low intensity protein bands were observed on a SDS PAGE gel, none of which was in the predicted 38-40kDa range for SMN WT. These results suggest a low expression, purification and/or stability for purified SMN WT.

This chapter describes the optimization of the expression and purification of SMN WT. The yield of recombinant SMN protein was increased by optimizing the expression, purification and stability with the overall aim of characterizing the interaction of recombinant SMN and Gemin2 *in vitro*. SMN WT was expressed employing various conditions, including varying cell strains, induction temperature and time and media additives. Denatured SMN proteins were purified and refolded using nickel affinity chromatography and dialysis. Lastly, purified SMN proteins were incubated and concentrated in several buffer conditions and analyzed via standard and non-reducing SDS PAGE to increase the overall stability of SMN protein.

3.2 METHODS

3.2.1 Constructs

Constructs containing full-length SMN (SMN WT) and SMN lacking exon 7 (SMN D7) were previously cloned into the pRSETc vector with an N-terminal histidine tag (Campanaro 2009). SMN TRX was obtained from another lab (Young 2000b). SMN TRX was constructed by cloning full-length SMN into the pET32 vector, which contains an N-terminal thioredoxin (TRX) tag and a C-terminal histidine tag.

3.2.2 Expression optimization

SMN and SMN D7 constructs were transformed into BL21 and Rosetta PlysS cells. Single colonies were inoculated into 3mL of SOB media with or without 0.5% glucose, 1.0ug/ul of ampicillin, and 0.35ug/uL of chloramphenicol (Rosetta PlysS). Small cultures were incubated at 30°C or 37°C and shaken at 300rpm for 12-16hrs. Small cultures were then inoculated 1:100 into either 50mL or 1.5L large SOB cultures and grown until the OD_{600} reached approximately 0.8- 1.0. Large cultures were induced with 1mM IPTG at either 18°C, 25°C or 37°C and grown for 4-16hrs. Cells were centrifuged at 6000rpm for 10 minutes and frozen at -20°C.

3.2.3 Inclusion body purification and nickel column refolding

Bacterial cell pellets were resuspended (5ml/g cell paste) in PBS with a spec of DNase powder and a final $MgCl_2$ concentration of 20mM. The resuspended lysate either was sonicated or underwent 3-5 freeze/thaw cycles (Rosetta cells) to lyse cells, incubated with 1% Triton X for 5 minutes and spun down at 13,000rpm

for 30 minutes. This inclusion body extraction was repeated with addition of DNase, $MgCl_2$ and 1% Triton X, and again with 2M Urea both in PBS buffer. Purified inclusion bodies were solubilized in 50ml of buffer A (50mM sodium phosphate pH 7.5, 500mM NaCl, 8M Urea) with addition of 20mM BME overnight at 4°C for 12hrs or at room temperature for 2hrs. Solubilized inclusion bodies were clarified again by centrifugation at 13,000rpm for 30 minutes and the supernatant was incubated with nickel beads (Fast flow 6, GE) equilibrated with buffer A for one hour at 4°C. The lysate-bead mixture was then settled by gravity in a column (Biorad) and washed with buffer A + (buffer A containing 20-40mM imidazole). The bound protein was then refolded by addition of 10 column volumes each of a 6-step gradient transitioning from denaturing buffer A to native Buffer B (50mM sodium phosphate pH 7.5, 500mM NaCl, 500mM Sucrose) at the following ratios: 50:10ml, 40:20ml, 30:30ml, 20:40ml, 10:50ml, 0:60mL. The folded protein was then eluted with elution buffer (buffer B containing 250mM imidazole and lacking 500mM sucrose). Eluted fractions were monitored via UV absorbance at 280nm and pooled for SDS PAGE and Western Blot analysis as described in Chapter 2.

3.2.4 Denaturing Nickel Column Purification and Dialysis Refolding

The same process above was repeated excluding the six step gradient and protein was eluted in buffer A containing 250mM imidazole. Denatured protein fractions were pooled based on the absorbance at 280nm, dialyzed for 2hrs each against 2M, 1M, 0.5M, 0.25M Urea in PBS and finally for 8 hrs against PBS at 4°C and then frozen at -20°C.

3.3 RESULTS AND DISCUSSION

3.3.1 SMN constructs

Three plasmid constructs were used to express recombinant SMN WT and SMN D7 proteins. His-SMN WT and His-SMN D7 sequences have been previously cloned into the pRSETc vector with a modified shortened linker sequence containing a single glycine between the histidine tag and the first methionine of the SMN protein (Figure3.1). The SMN WT and D7 sequences are similar but vary in sequence at the C terminus due to an exon seven exclusion and the consequent early translation termination. The third construct, SMN TRX was obtained because of its high stability and solubility (Nguyen thi 2008). The 12kDa thioredoxin tag is often used to increase the solubility of proteins refolded from bacterial inclusion bodies (Shahdev 1998).

His-SMN WT (32kd)
MRGSHHHHHHGMAMSSGGSGGGVPEQEDSVLFRRGTGQSDDSDIWDDTALIKAYDKAVASFKHALKNGDI
CETSGKPKTTPKRKPAKKNKSQKKNTAASLQQWKVGDKCSAIWSEDGCIYPATIASIDFKRETCVVVYTGYG
NREEQNLSDLLSPICEVANNIEQNAQENENESQVSTDESENSRSPGNKSDNIKPKSAPWNSFLPPPPPMPG
PRLGPGKPGLKFNGPPPPPPPPPPHLLSCWLPPFPSGPPIIPPPPPICPDSLDDADALGSMLISWYMSGYHT
GYYMGFRQNQKEGRCSHSLN

His-SMN D7 (31kd)
MRGSHHHHHHGMAMSSGGSGGGVPEQEDSVLFRRGTGQSDDSDIWDDTALIKAYDKAVASFKHALKNGDI
CETSGKPKTTPKRKPAKKNKSQKKNTAASLQQWKVGDKCSAIWSEDGCIYPATIASIDFKRETCVVVYTGYG
NREEQNLSDLLSPICEVANNIEQNAQENENESQVSTDESENSRSPGNKSDNIKPKSAPWNSFLPPPPPMPG
PRLGPGKPGLKFNGPPPPPPPPPPHLLSCWLPPFPSGPPIIPPPPPICPDSLDDADALGSMLISWYMSGYHT
GYYMEMLA

SMN-TRX (55 Kd)

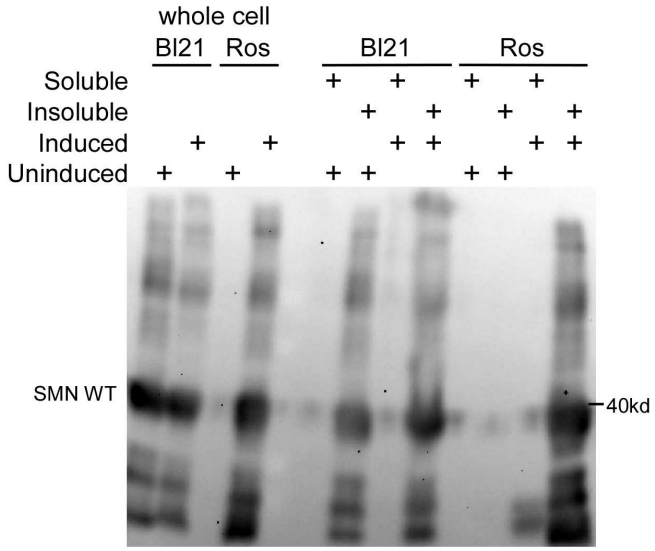
Figure 3.1- SMN constructs
The amino acid sequences of both SMN WT and SMN D7 are shown with the histadine tag (orange) followed by a glycine linker and the start codon of the native protein sequences (red). SMN D7 varies in sequence from WT at the C terminal end (orange). The SMN TRX construct contains an N-terminal Thioredoxin tag (green) and a C-terminal histidine tag (orange).

3.3.2 SMN expression optimization

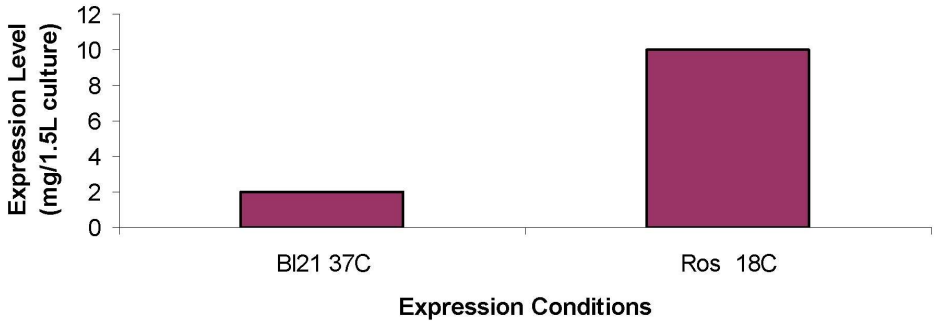
The *E. coli* containing the SMN WT construct was observed to have a two-fold decrease in growth. SMN WT was expressed in Rosetta and Bl21 cells and samples of bacterial cells were analyzed before and after induction. Cells were also lysed and separated by centrifugation to analyze soluble and insoluble samples to determine if the leaky expression of SMN WT caused the decline in the bacterial growth rate. SMN WT in BL21 cells showed a large expression band in both the un-induced and induced whole cell samples whereas the Rosetta cells only showed SMN expression in the induced sample and a weak expression band in the un-induced sample (Figure 3.2a). The leaky expression of SMN WT in Bl21 cells was abolished in Rosetta cells by the extra PlysS plasmid that encodes for T7 Lysozyme. The excess T7 Lysozyme functions to degrade extra T7 RNA polymerase and prevent nonspecific transcription of the target gene controlled by the *lac* operon. The insoluble Bl21 samples showed high SMN expression before and after induction with IPTG, while the soluble fractions showed a small amount of SMN expression. Over-expressed proteins that form inclusion bodies are not toxic to the cell making the small amount of soluble expressed SMN in the BL21 cells the likely culprit for the observed slow bacterial growth in Bl21 cells. The toxicity of the SMN protein along with the known plasmid instability of the pRSETc vector likely resulted in low expression levels of SMN WT (Dumon 2004, Walia 2008). Since, the SMN protein and plasmid are toxic and unstable to the bacterial cells; the cells in the glycerol stock during

growth could have recombined the unstable DNA plasmid, failed to propagate the plasmid during cell division or died due to protein toxicity. This also explains the restoration of SMN expression and normal cell growth after re-transformation of SMN WT plasmid into fresh Bl21 competent cells.

To prove that Rosetta cells increase the long-term stability of SMN expression, SMN WT was expressed from both a Bl21 and Rosetta glycerol stock and the expression was measured by detecting SMN via a western blot. Bl21 cells grown from a glycerol stock showed a low amount of SMN WT expression at 2mg per 1.5L whereas SMN WT expressed in Rosetta cells showed a 5-fold increase at 18°C (Figure 3.2b). Rosetta cells also permanently rescued bacterial growth and have the capacity to increase SMN WT expression to approximately 10-fold to 20mg/1.5L at 37°C. The 37°C growth condition was not chosen due to high amounts of partially translated SMN protein (data not shown). The Rosetta cells increased the expression of SMN WT, stabilized the expression system and also verified that SMN WT migrates at approximately 34kDa in our conditions instead of the predicted 38-40kDa.



a



b

Figure. 3.2. Optimization of SMN expression
(a) SMN was grown in BL21 (DE3) and Rosetta PLysS cells and induced for 3hrs. Whole, soluble and insoluble cell fractions were collected before and after induction and analyzed via SDS PAGE and western blot analysis using an anti-SMN primary antibody. SMN WT protein is labeled just below 40kDa. (b) the SMN WT construct in Bl21 and Rosetta cells was inoculated from a glycerol stock and expressed for 12hrs. Relative expression was estimated by quantification of western blot analysis using purified His-GFP as a control.

3.3.3 Comparison of dialysis and nickel column refolding

SMN TRX a construct received from Dr. Glenn Morris was shown to have a high purification yield using dialysis refolding to refold SMN from inclusion bodies (Young 2000b). A comparison between nickel column refolding and dialysis refolding was made to determine the most efficient technique. SMN WT and TRX proteins were both dialyzed and refolded on a nickel column and analyzed via absorbance at 280nm and SDS PAGE. The purification resulted in similarly pure protein in both dialysis and nickel column refolding (data not shown). The overall yield of each SMN protein was similar for both refolding techniques (Table 3.1). However, the concentration of the dialyzed refolded protein was significantly lower for SMN WT. This was likely due to SMN WT aggregates diffusing through the dialysis bag and nucleating aggregation of other partially folded or native SMN proteins. Furthermore, on-column refolding took an additional hour of purification after nickel affinity chromatography whereas dialysis took an additional 16hrs. SMN TRX had a higher yield using the dialysis method, but was less stable then SMN WT (shown below). In sum, these results show that nickel column chromatography is the most efficient method of refolding SMN from inclusion bodies.

	Dialysis (A280/ml)	**Nickel Column (A280/ml)**	**Total Protein Yield (A280)**
SMN WT	0.08-0.20 (ppt)	0.10-0.30	~ 2.0-4.6
SMN TRX	0.80	0.29	~ 3.0-8.0

Table 3.1- Comparison of refolding techniques
SMN WT and TRX were purified and eluted under denaturing conditions and dialyzed in a 4 step gradient with decreasing urea and increasing PBS or purified and refolded on-column. Purified protein was detected by measuring the

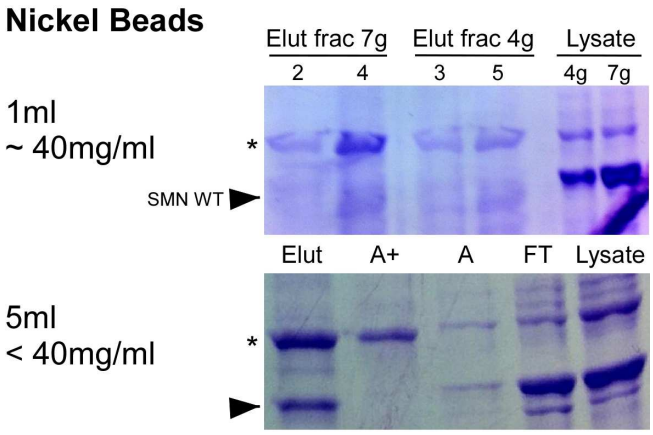
absorbance at 280nm. Ppt indicates precipitation of protein was observed during refolding.

3.3.4 Increased refolding efficiency and purity of SMN WT

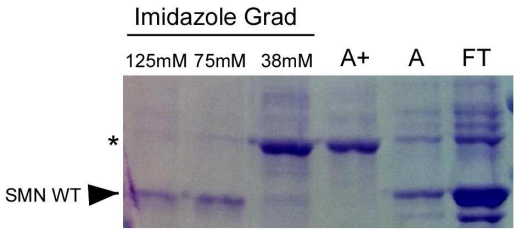
After expression of SMN WT was increased by approximately five-fold, the concentration and total yield of eluted protein did not increase as expected, but instead decreased dramatically. Since, the increase in SMN expression was the only variable changed in the purification process, it was hypothesized that the higher bead binding capacity of the Nickel beads (40mg/ml) induced aggregation during protein refolding on the column. To verify that the higher expression level of SMN WT caused aggregation during refolding, SMN WT was purified using the recommended conditions of 1ml of nickel beads with a 40mg/ml binding capacity as well as 5ml of older nickel beads with a lower bead binding capacity. The elution fraction of the 1ml bead purification contained barely detectable amounts of SMN WT compared to the SMN WT eluted from the 5ml bead elution showing that SMN aggregation during folding caused the decreased yield in SMN (Figure 3.3a). Aggregated SMN WT was not eluted with high imidazole buffer. This was most likely because of a strong hydrophobic interaction between the nickel resin and aggregated protein seen in previous studies and was further confirmed by a significant decrease in column flow rate during refolding (Sweitnicki 2006). In addition, increasing the bead volume to 10ml further increased the overall yield of SMN WT. The elution fraction of the SMN WT purification contained an impurity just above 40kDa (Figure 3.3a). To optimize the purification, SMN WT was purified and eluted under denaturing conditions

with a step gradient of imidazole. SMN WT eluted in the 150mM imidazole fraction while the 40kDa impurity eluted at 38mM imidazole, which showed that an increase in imidazole to 40mM could remove the purity (Figure 3.3 b).

The increase in nickel beads led to an increase in overall yield from approximately 2mg to 4.6 mg per nickel purification. The increase in yield was coupled with an increase in efficiency because instead of using one cell pellet for the purification, only one third of the sonicated and urea-solubilized pellet was used to ensure that SMN WT did not saturate the beads during binding. In addition, the impurity at 40kDa comprised about 50% of the total purified protein and without it the purity of SMN WT increased to greater then 95%.



a



b

Figure 3.3- Optimization of nickel affinity SMN purification

(a) SMN WT was purified via nickel affinity chromatography under denaturing conditions, refolded on-column with a decreasing urea gradient and then eluted with high imidazole buffer. The top panel shows SMN WT purified using 1ml of nickel beads with a high binding capacity of 40mg/ml. *E. coli* lysate and elution fractions from two purifications are indicated (3 and 7 gram pellet). The lower panel shows purification with 5mls of lower binding capacity beads. *E. coli* lysate and elution fractions are indicated as well as flow through, buffer A (Binding buffer), and buffer A+ (Binding buffer containing 20mM imidazole). SMN WT is indicated with an arrow and an impurity with an asterisk (b) Urea-denatured SMN WT lysate was bound to 5mls of low capacity binding nickel beads, washed and eluted with a step gradient of imidazole in urea denaturing buffer. SMN WT is indicated with an arrow.

3.3.5 SMN WT and D7 purification

SMN WT and D7 were purified by nickel chromatography and the purified proteins were assessed for stability one day later. SMN WT and D7 in the eluted fractions migrated as a smear at 34kDa (Figure 3.4a). Both SMN proteins underwent degradation. In addition, both proteins migrated as a 100kDa SDS=resistant trimer band. This band is unlikely to be an impurity because it was not present in SMN purified and eluted under urea denaturing conditions. The observed oligomers at 100kDa in SDS PAGE are in agreement with the low-resolution structure of SMN WT, which showed two trimers forming a pseudo hexamer (Seng, unpublished). Wan et al showed that native SMN in rabbit raticulosyte (RRL) was able to form SDS-resistant dimers and a small amount of trimers via two conserved cysteine residues in the N and C termini (Wan 2008). The trimers may not have been a major product in their study because RRL contains chaperones and other SMN complex proteins that when bound to SMN could have sterically hindered trimer formation and/or the necessary disulfide bonds. However, we cannot rule out the fact that the degradation of our SMN

WT in SDS conditions caused the disulfide linked SDS-resistant trimer. Another interesting difference between our results and the Wan et al results was that SMN D7 did not form SDS-resistant dimers and trimers whereas our recombinant SMN D7 does form trimers. Again, this points to the idea that SMN WT and SMN D7 have similar abilities for disulfide bond formation *in vitro*, but when bound to other proteins in RRL, a minor structural difference or interaction with other proteins may inhibit the optimal orientation of SMN D7 thus inhibiting intermolecular disulfide bond formation. It would be interesting to determine if there is a different sensitivity to disulfide bond formation between SMN WT and D7 *in vitro*, which could discern whether the inability of SMN D7 to form disulfide bonds in RRL was due to intrinsic structural differences or extrinsic protein interactions.

3.3.4 Optimization of SMN WT stability

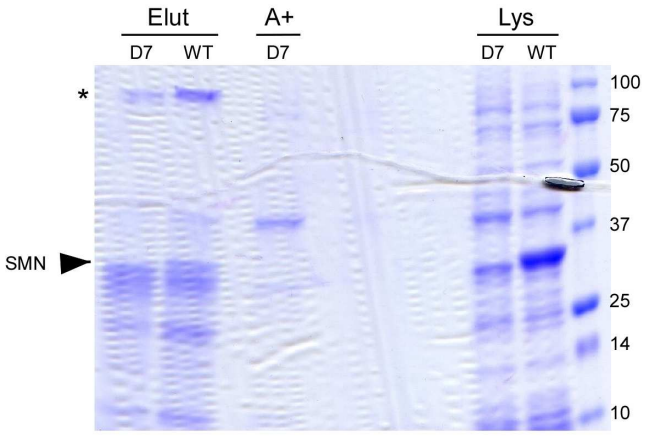
In order to prevent degradation, protease inhibitors were added to the refolding and elution buffers during SMN WT purification. SMN was purified with and without protease inhibitors and analyzed for stability at 4°C. After one day, the SMN WT without protease inhibitor was not present at all and a small amount of SMN WT was present in the purification with protease inhibitors (3.4b). Protease inhibitors did increase the stability of SMN slightly, but even SMN WT with protease inhibitors either degraded and/or precipitated out of solution at 4°C.

As discussed in Chapter 2, the concentration of Gemin2 promoted protein degradation; we also observed the stability of SMN WT upon concentration.

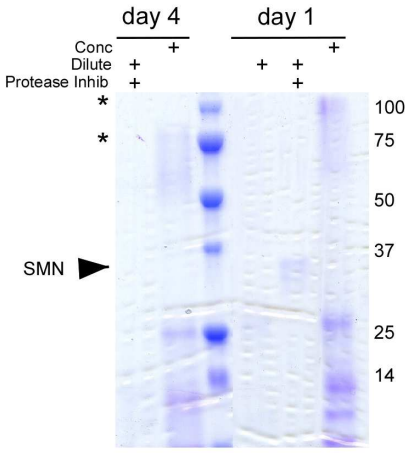
Concentrated SMN WT degraded down to a 25kDa band and the trimer band at 100kDa shifted down to 75kDa (3.4b). The shift from 100kDa to 75kDa of the trimer band corresponds directly to the shift from 34kDa down to 25kDa of the SMN WT monomer band and further provides evidence that SMN WT is forming disulfide-linked trimers. Unexpectedly, concentrated SMN WT was more stable then dilute SMN WT possibly because it formed SDS-resistant trimers.

In an attempt to increase SMN WT stability, protein in elution buffer was concentrated and buffer exchanged into Tris buffer and HEPES buffer, incubated for seven days and analyzed via SDS PAGE. SMN WT in the elution buffer lane migrated as a more intense band at 34kDa and had less degraded protein bands then in the Tris and HEPES buffers immediately after buffer exchange (Figure 3.4c). After seven days of incubation at 4°C, concentrated SMN WT appeared to degrade and/or aggregate in all buffers with SMN in the elution buffer still showing the least amount of degradation. The high salt, BME, or glycerol in the elution buffer may facilitate the stabilization of SMN WT. All three samples migrated as SDS-resistant trimer bands initially after concentration, but the Tris and HEPES buffer samples migrated as two higher intensity bands around 50kDa, which could be disulfide linked SDS-resistant degradation fragments as those buffers did not have BME. Similar to the nucleated aggregation during dialysis, disulfide linked degraded SMN WT fragments may promote SMN WT to aggregate. The degradation of BME over time may be responsible for the overall decrease in SMN stability seen in all three buffers after seven days. Overall,

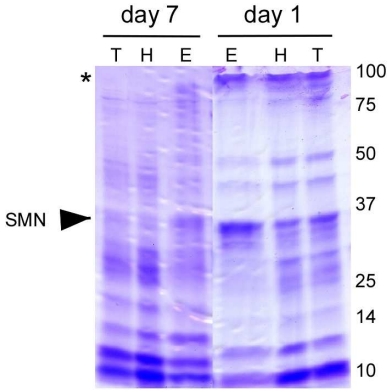
SMN WT was unstable over a week time span, while the presence of BME and protease inhibitors and high salt temporarily increased stability.



a



b



c

Figure 3.4- Stability of SMN

Purified SMN was analyzed via SDS PAGE for stability. SMN WT and D7 are indicated with an arrow and disulfide-linked trimers are indicated with asterisks.

(a) SMN WT and D7 were purified with optimized nickel affinity chromatography conditions. (b) Dilute SMN WT purified with and without protease inhibitors as well as concentrated SMN WT. Samples were incubated for 7 days at 4°C and were analyzed. (c) Purified SMN WT in elution buffer was concentrated and buffer exchanged in either elution buffer containing 50mM sodium phosphate pH 7.4, 500mM NaCl, 5mM BME, Tris buffer (20mM Tris pH 8.5, 250mM NaCl), or HEPES buffer (20mM HEPES pH 7.5, 250mM NaCl, 5% glycerol). Elution buffer is indicated with an E, Tris buffer with T and HEPES buffer with an H. Concentrated samples were analyzed immediately after concentration and 7 days later for stability.

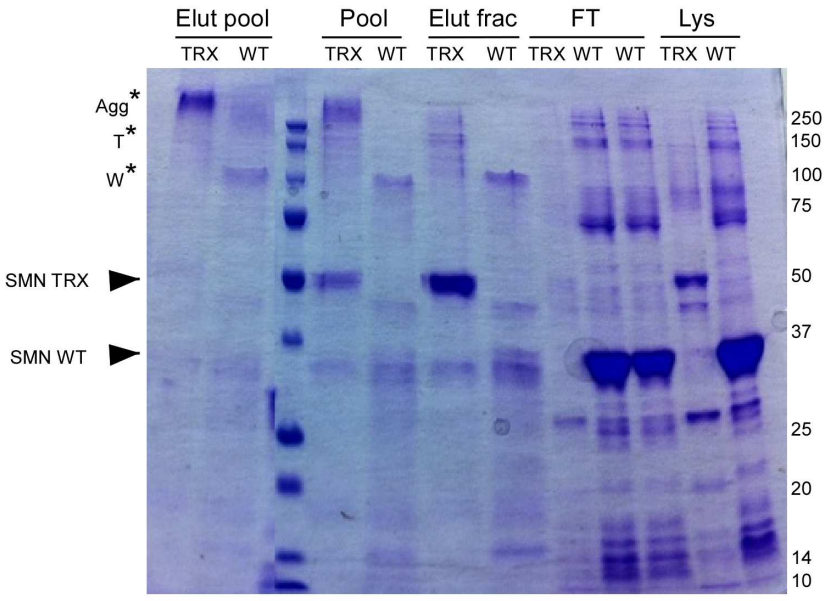
3.3.5 Comparison of the stability of SMN WT and SMN TRX proteins

The stability of SMN TRX and WT were compared during and after purification to determine which protein was suitable for future SMN-Gemin2 characterization studies. SMN TRX and WT were refolded on-column and the process was monitored via SDS PAGE. SMN TRX had a larger elution band then SMN WT even though the SMN TRX over-expression band in the lysate was significantly less then the SMN WT suggesting that SMN TRX refolded more efficiently (Figure 3.4a). SMN WT was present primarily in the elution fraction as the trimer form at 100kDa with a small amount in the monomer form at 34kDa. In the elution fractions, SMN TRX migrated as a slight doublet band at 150kDa that corresponded to SDS-resistant trimers and a larger aggregate or oligomeric band at 250kDa. SMN WT also migrated as a 34kDa doublet band immediately after purification that shifted down to the lower band in the pooled protein. One possibility for the observed doublet bands is that both SMN proteins undergo similar degradation. However, the fact that the SMN TRX construct varies so drastically from the SMN WT construct makes this possibility less likely. Another possibility is that SMN WT and TRX doublet bands are representative of

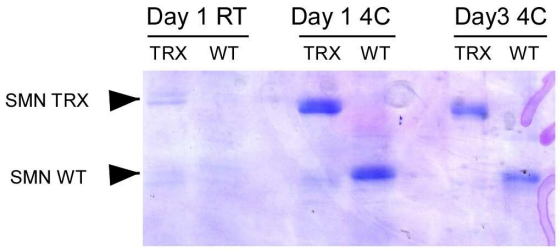
two different protein states that migrate differently in SDS PAGE. Recombinant SMN has been previously noted to have unusual migration in SDS PAGE (Campanaro 2009). The cause of the doublet bands in SMN lacking post-translational modifications is unclear. In the pooled fraction, a large majority of SMN TRX shifted from the 55kDa monomer form to the aggregate or oligomer form at 250kDa. SMN TRX without BME shifted completely to the aggregate band whereas in the SMN WT sample only a small amount of protein shifted to the aggregate and the majority of the monomer shifted to the trimer state. In both the reducing and non-reducing conditions, SMN TRX favored the monomer or aggregate form, whereas SMN WT favored the trimer form. It is possible that the12kDa N-terminal TRX tag inhibits the trimer orientation and disulfide bond formation. The observed aggregation of SMN TRX and WT may be caused by nonspecific disulfide bond formation and/or strong hydrophobic interactions of the predominantly beta sheet SMN protein (Nielsen 2007). SMN TRX had a higher refolding efficiency and total purified protein, but an increase of soluble aggregates or oligomers in SDS PAGE relative to SMN WT suggest SMN TRX may not be able to form the same oligomeric states as SMN WT.

Next, the stability of both proteins after purification was compared. SMN TRX and WT were incubated at RT or 4°C and protein stability was analyzed via SDS-PAGE. Both samples degrade or aggregate at RT after one and slightly after four days (3.4b). According to these results, there is no significant difference in the stability of SMN TRX and WT.

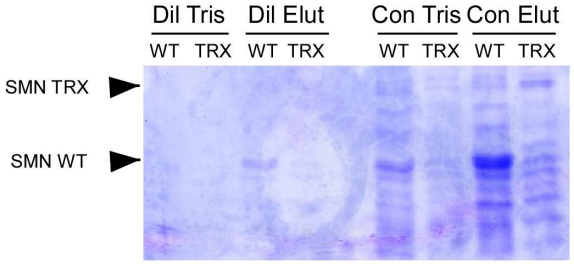
Since both proteins were stable over four days, the stability after protein concentration was determined. SMN WT and TRX were concentrated, and buffer exchanged in either HEPES or Tris buffers. Dilute and concentrated samples were analyzed for stability. Concentrated SMN WT appeared to be more stable then SMN TRX as indicated by the larger band at 34kDa and both proteins were less stable in Tris buffer (3.4c). Because SMN TRX was less stable then SMN WT during protein concentration and SMN TRX favored large aggregates over the ability to form disulfide trimers, SMN WT protein was used for further *in vitro* studies to characterize the SMN-Gemin2 interaction.



a



b



c

Figure 3.5- Stability of SMN WT compared to SMN TRX
The stability and solubility of SMN WT and TRX were compared under various conditions during and after purification. (a) SMN WT and SMN TRX were purified using nickel affinity chromatography. Elution fractions as well as pooled fractions were analyzed on standard and non-reducing SDS PAGE. Both SMN TRX and WT are indicated with arrows. SDS resistant oligomers are indicated with asterisks, a letter W corresponding to WT, T for TRX and A for soluble

aggregate and/or large oligomers. (b) SMN TRX and SMN WT were purified by dialysis in PBS and were incubated at 4°C and RT for 4 days to determine their stability. (c) SMN TRX and SMN WT in PBS dialysis buffer were buffer exchanged and concentrated in elution buffer (50mM sodium phosphate pH 7.4, 500mM NaCl, 5mM BME) and Tris buffer (20mM Tris pH 8.5, 200mM NaCl, 5mM BME).

Chapter 4

CHARACTARIZATION OF GEMIN2 AND SMN PROTEINS

4.1 INTRODUCTION

The self-association domains of SMN as well as the oligomerization of SMN has been studied, but the actual oligomeric state of SMN *in vitro* and *in vivo* is unknown (Young 2000a, Lorson). SMN D7 lacks the ability to form high-order oligomers *in vitro* suggesting that oligomerization is important for the function of SMN. Furthermore, the oligomerization defects of SMN missense mutations in SMA correlate with the severity of the disease phenotype (Lorson 1998).

SMA research has largely been focused on understanding the function of SMN by using *in vivo* cell culture or animal models that use a SMA phenotype to work backward and uncover the disease-causing molecular event. The current studies clearly show that oligomerization is at least part of the molecular event cascade that leads to SMA. Our lab's approach was to start with the SMN protein at the beginning of the molecular cascade. We sought to gain a better understanding about the oligomerization of SMN WT and SMN D7 with the hope that our SMN protein forward approach combined with others SMA phenotype backward approach will eventually close in on the specific molecular event that causes SMA.

SMN-Gemin2 complex interactions have been shown to be important for the function of SMN in snRNP assembly (Fischer 1997). Similar to the knockout of the *smn1* gene in mice, the knock out of *gemin2* is embryonically lethal suggesting each of their functions is crucial for development. Gemin2 and SMN

protein levels were shown to be co-regulated in mice and SMN co-localized with Gemin2 in Hela cells with the highest frequency relative to other SMN-binding proteins (Jablonka 2001, Jablonka 2002, Todd 2010). While SMN has been shown to associate with a variety of proteins, the interdependence and consistent association of SMN and Gemin2 *in vivo* suggests Gemin2 is important for the majority of SMN functions. We thus chose to characterize the SMN-Gemin2 interaction to determine if there is a difference between SMN WT and SMN D7 with the aim of gaining insights into the SMA pathogenesis.

In this chapter, we characterized the oligomerization states of SMN WT and SMN D7 bound to Gemin2 as well as SMN WT and Gemin2 proteins individually. Purified SMN WT and D7, and Gemin2 proteins were characterized by size exclusion column (SEC) chromatography and SDS and Native PAGE gels.

4.2 METHODS

4.2.1 Native PAGE

Protein samples were mixed with 2X native page sample buffer (0.01% Bromophenol Blue, 200mM Tris pH 8.0, 20% glycerol) with or without BME and run on either a 4-20% gradient or 12% discontinuous Tris/Glycine gel (TGX gels Biorad) at 200V on ice until the dye reached the bottom of the gel. Proteins were also run using a McLellan Tris-boric acid continuous gel system. Briefly, samples were mixed by adding one volume of protein sample to two volumes of 0.5X Loading dye (0.5X McLellan gel buffer, 10% glycerol, 0.01% bromophenol blue) on ice. Samples were loaded onto a 6% Tris/boric acid Mclellan gel made and run with the same McLellan gel buffer (10mM Tris, 5mM Boric Acid pH 8.7) for 200V for 40 minutes on ice. For Blue Native (BN) PAGE gels, samples were mixed with 4X Native dye (Invitrogen BN PAGE kit), loaded onto a Bis-Tris 3-12% gel and run with anode buffer in the outer chamber and light cathode buffer in the inner chamber (Invitrogen BN PAGE Kit). The gel was run for 1hr on ice at 100V through the stacking gel and 250V until the dye reached the bottom of the gel. All gels were fixed by heating the gel in 100mL of fixing buffer (40% methanol, 10% acetic acid) in the microwave for 30 seconds and then gently shaking for 20 minutes. This process was repeated for staining (coomassie blue-R stain) and destaining buffers (10% acetic acid) according to the Invitrogen BN PAGE staining protocol.

4.2.2- Size exclusion chromatography

A Superdex 200 5/150 analytical size exclusion tricorn column (GE healthcare) was run using a Pharmacacia FPLC system and monitored via a Biorad Econoview detector coupled to an analogue-digital convertor (Datataq). Standards (BSA, Carbonic anhydrase and B-amyalose or blue dextran-1mg/ml) and concentrated protein samples (SMN WT, SMN TRX, Gemin2, SMN D7 and SMN-G2- concentration range of 0.5-3mg/ml) were run with either phosphate SEC buffer (50mM sodium phosphate, 150mM NaCl pH 7.4) or Tris-reducing SEC buffer (20mM Tris HCl, pH 8.0, 200mM or 500mM NaCl, 1mM EDTA, 5-10mM BME) at a flow rate of 0.2ml/min with a detector sensitivity (Biorad Econoview) of 0.01auf and an anologue-digital converter frequency of 1s.

4.2.3- SEC sample preparation

Frozen protein samples were thawed quickly and immediately 5-10mM BME, 1mM EDTA and 1X protease inhibitors (Gold Biochem) were added to each protein aliquot. SMN was bound to Gemin2 by incubating 1:1 molar ratios of SMN WT and Gemin2 protein at 4°C for one hour. Protein samples were then concentrated in a 500uL 10kDa molecular weight cut off membrane for 20 minutes at 4°C. Protein samples were first buffer exchanged into the appropriate SEC buffer and then a volume of 25uL was injected onto the column.

4.3 RESULTS AND DISCUSSION

4.3.1 Oligomeric states of SMN in vitro

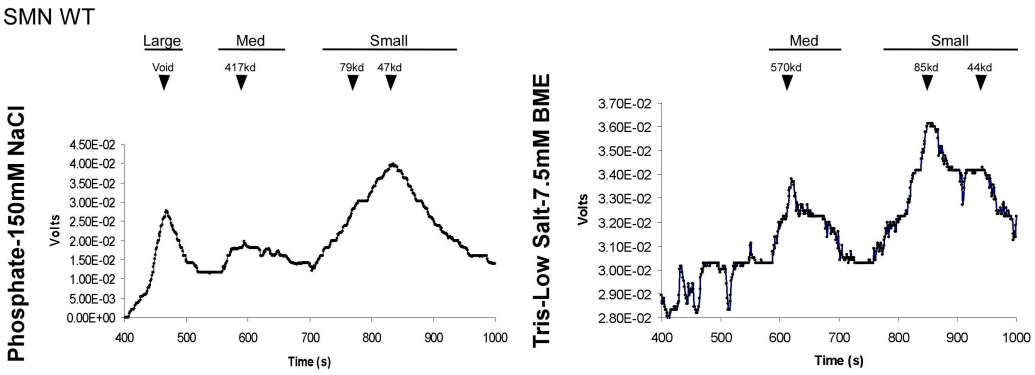
In order to determine the oligomeric state of SMN WT *in vitro*, purified proteins were run on an analytical SEC column that contained a large separation range from 20kDa to 600kDa. SMN WT was first run under standard SEC conditions (50mM sodium phosphate pH 7.4, 150mM NaCl). The SMN WT elution peaks run with standard SEC buffer showed three very broad peaks with one of those peaks being at the void volume of the column (Figure 4.1a left panel). The three peaks suggest that SMN WT did not form one discrete oligomer *in vitro*; rather SMN WT formed a mixture of oligomers that we have labeled in relative size as small, medium and large. The molecular weights of the most prominent peaks were calculated from a set of protein standards and are shown in the chromatogram. The broad peak of each relative oligomeric size suggests that SMN WT formed a mixture of states within each size range. For example, the small oligomeric peaks at 47kDa and 79kDa may be a mixture of monomers, dimers and trimers that cannot be resolved. It is possible that resolved peaks were not observed due to relatively low protein concentration, low detector sensitivity and the self-association of full-length SMN WT with degraded fragments. Further SEC experiments using a column with a smaller separation range could be used to determine the specific lower molecular weight oligomeric states of SMN WT.

4.3.2 The effects of salt and BME concentration on SMN oligomers

Because SMN WT SEC under standard SEC conditions led to persistent column clogging and maintenance, alternative buffer conditions with high and low salt concentrations were employed. Varying BME concentrations were employed as well, since the SDS-resistant oligomers in Chapter 3 indicated that disulfide bond formation may lead to non-specific oligomer formation of SMN. SMN WT run on SEC with Tris buffer containing low salt and 7.5mM BME resulted in small and medium, but no large oligomeric states at the void volume suggesting that BME was in deed able to shift the large oligomeric states by reducing disulfide bonds (Figure 4.1a, right panel). In addition, the more resolved peaks at 44kDa and 85kDa point to SMN WT being in a mixture of monomers and dimers.

SMN WT was next run with SEC buffers containing a range of salt and BME concentrations to gain insight into the formation of SMN oligomeric states. In agreement with the suggestion that increased BME caused smaller SMN WT oligomeric states, none of the buffers containing BME showed large oligomers in SEC. The higher concentrated BME buffers shifted the majority of the medium oligomers to small oligomers with the Tris-high salt buffer predominantly favoring small oligomers at 44kDa and no observed peaks at 85kDa (Figure 4.1b). There also appeared to be a trend of decreased SMN WT small oligomer size with increasing salt concentration. However, more SEC analysis of SMN WT containing buffers with the same BME concentration and varying salt concentration would need to be done to determine the separate contribution of salt and BME concentration to small oligomer formation. The two resolved peaks in

Figure 4.1a (right panel) at 44 and 85kDa were also observed in all buffer conditions with the second larger peak having a molecular weight approximately double of the first peak. Upon SDS PAGE analysis, fractions eluted from SEC in the small oligomer range migrated at 66kDa, which could be a disulfide-linked SMN WT dimer (data not shown). These results provide evidence showing that SMN WT in the small oligomers observed in SEC existed at least in monomers and dimers. The SEC molecular weights of SMN WT monomers and dimers of 44kDa and 88kDa are significantly higher then the actual molecular weights of 34kDa and 68kDa. The discrepancies in molecular weight of SMN WT are likely due to the increased hydrodynamic size of SMN observed in the SEC of the SMN self-association domain (Pellerzzoni 1999, Martin 2012). However, the presence of SMN WT dimers is not in agreement with either the SDS-resistant 100kDa trimers observed in Chapter 3 or our low-resolution SMN WT hexamer structure. Further experiments to determine the actual composition of SMN oligomers and the ratio of those oligomers *in vitro* could provide more insight into the assembly and oligomerization of SMN WT *in vivo*. We propose that SMN WT forms three oligomeric sizes *in vitro* with the small oligomeric size composed of monomers, dimers and/or trimers, the medium size composed of hexamers that we observed in our low-resolution structure and the large size composed of two interacting hexamers that make up a dodecamer (Figure 4.1c).

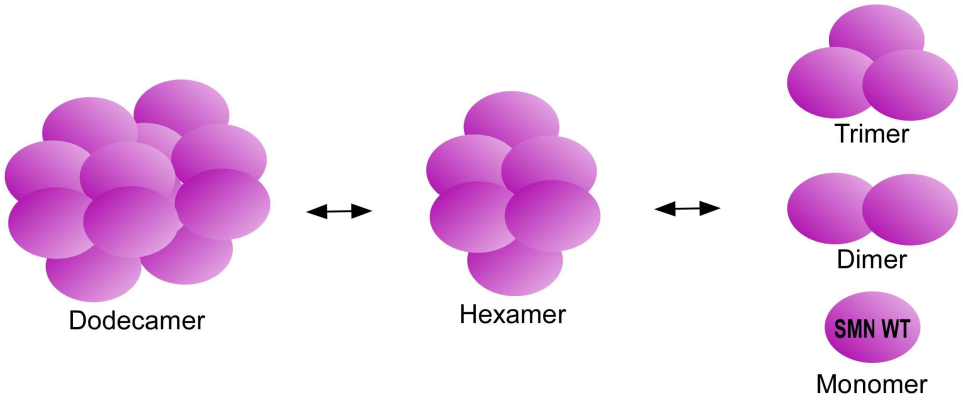


a

Buffer Conditions	Relative SMN Oligomer Size
Phosphate-Low salt	Large, med, small
Tris-High salt-10mM BME	Med*,small
Tris-High salt-5mM BME	Med, small
Tris-Low salt-7.5mM BME	Med, small

* The peak observed was not significantly higher then the baseline

b



c

Figure 4.1- SMN WT oligomerization states
(a) SMN WT in elution buffer were concentrated and buffer exchanged into SEC Tris buffer (20mM Tris pH 8.0) with either low salt (200mM NaCl) or high salt (500mM NaCl) and varying concentrations of BME or Phosphate buffer (50mM sodium phosphate pH 7.4, 150mM NaCl). Protein samples were run on a superdex200 SEC column and eluted protein was detected by absorbance at 280nm (measured in Volts) every second. Predicted molecular weight (kDa) of the protein peaks was calculated based on a set of protein standards. The left panel indicates SMN WT in elution buffer and the right panel indicates SMN WT in Tris Buffer with low salt and 7.5mM BME. Arrows indicate the protein peaks and the predicted molecular weight or the void volume of the column. The relative size of the SMN oligomers was categorized into three sizes, small, medium and large. (b) Relative oligomer sizes of SMN WT run on SEC in varying buffer conditions. (c) Proposed model of SMN WT oligomeric states *in*

vitro from SEC results. SMN WT (purple) may form small oligomers of monomers, dimers and/or trimers as well as larger hexamers and dodecamers.

4.3.3 Oligomeric state of SMN D7 in vitro

The oligomerization state of SMN D7 was examined using SEC with low salt buffer containing BME. Compared to SMN WT that formed both small and medium oligomers, SMN D7 only formed an unresolved small oligomer peak with a molecular weight of 85kDa (Figure 4.2a). This is in agreement with previous data showing that SMN D7 has a decreased ability to oligomerize *in vitro* (Pellerzzoni 2000). According to our results, SMN D7 was able to form monomers, dimers and/or trimers *in vitro* (Figure 4.2b). The dimer or trimer formation of SMN D7 was unexpected, as several studies have shown that SMN D7 has either partial or no ability to self-associate (Lorson 1998, Pellerzzoni 1999, Young 2000b). All three studies contained SMN protein with either an N-terminal GST or TRX tag. We have shown in Chapter 3 that the 12kDa N-terminal TRX tag changed the state of SDS-resistant oligomers observed in SDS PAGE suggesting that the N-terminal tag may inhibit SMN WT self-association likely at the N-terminal self association site of SMN. The inhibition of the N-terminal self-association site by the larger protein tags in these studies could explain the lack of self-association observed between SMN WT and SMN D7 (Figure 4.2c). Our data showed that SMN D7 can self-associate but does not directly prove that SMN WT can self-associate with SMN D7, which is likely to occur in most SMA patients with a high ratio of SMN D7 to SMN WT. Since it has been recently shown that the C-terminal region of SMN D7 alone did not

form dimers or trimers, our results indicate that the SMN D7 self-association observed is likely due the interaction of the SMN D7 N-terminal self-association site and thus we predict that SMN WT and SMN D7 are likely able to self-associate as well. However, in vivo, it is likely that SMN WT containing both the N-terminal and C-terminal self-association has a higher affinity for SMN WT compared to SMN D7. SEC analysis of our SMN D7 combined with SMN WT in a smaller separation range able to resolve monomers to trimers could verify that SMN WT and SMN D7 self-associate, which is the opposite of the current view of the SMN D7-SMN WT interaction. Determining if SMN WT-D7 self-associate *in vivo* and the functional consequences of the association will be very important in understanding the molecular function of SMN.

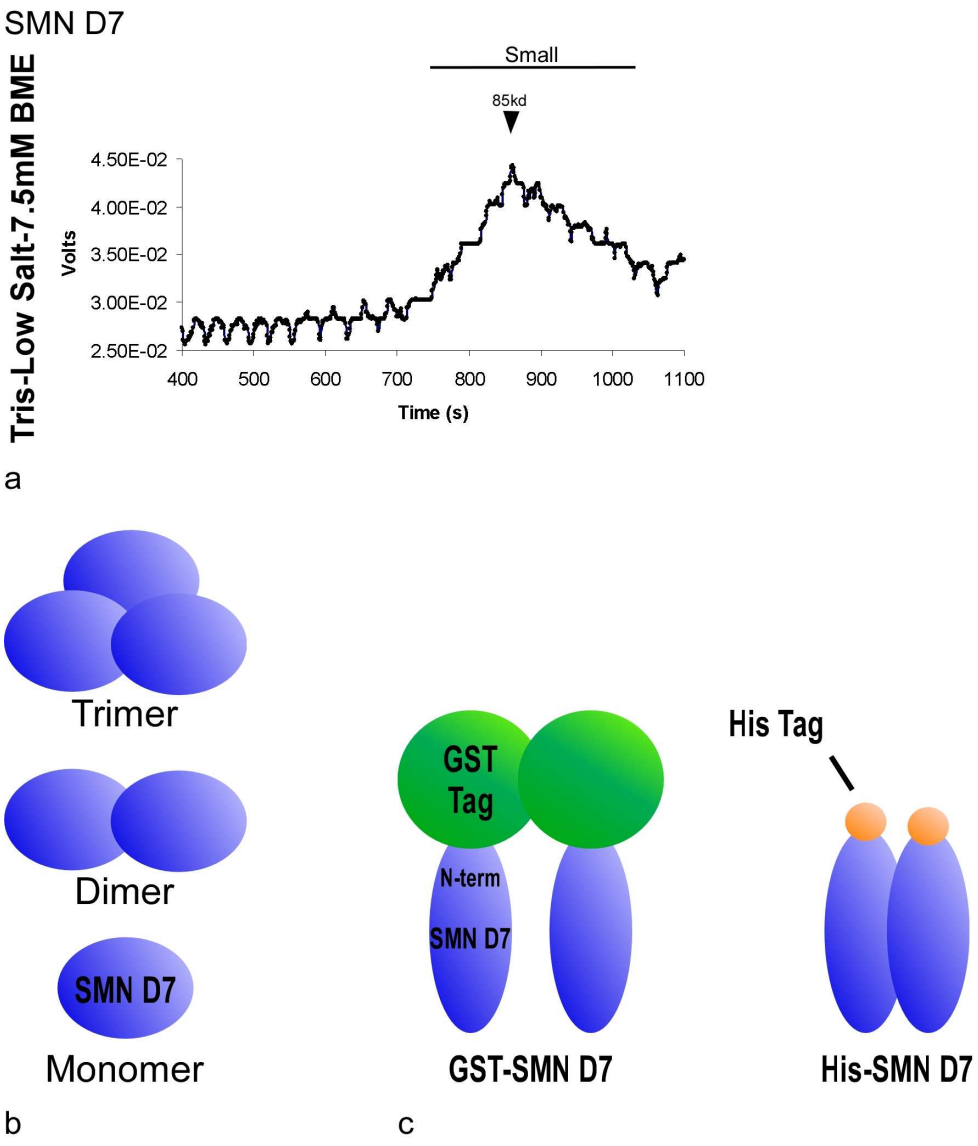


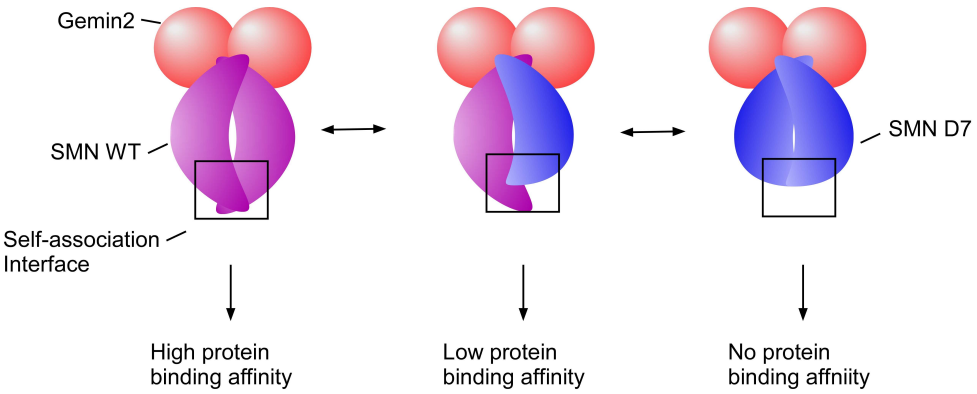
Figure 4.2- SMN D7 oligomerization states
(a) SMN D7 in elution buffer was concentrated and buffer exchanged into SEC buffer (20mM Tris pH 8.0, 200mM NaCl, 7.5 mM BME). Protein samples were run on a superdex 200 SEC column and eluted protein was detected by absorbance at 280nm (measured in Volts) every second. Predicted molecular weight (kDa) of the protein peaks was calculated based on a set of protein standards. (b) Proposed oligomeric states of SMN D7 from SEC results (blue). (c) Proposed N-terminal self-association of SMN D7 (blue). SMN D7 with a GST tag (green) does not self-associate whereas SMN D7 with a histidine tag (orange) does self-associate.

In summary, SMN WT formed a mixture of oligomeric states while SMN D7 only

formed smaller oligomeric states. Surprisingly, under our conditions of low salt

and BME, SMN D7 was able to self-associate in a similar manner as SMN WT. The current molecular view of the SMN D7-SMN WT interaction is that SMN D7 cannot self-associate with itself or SMN WT and this lack of self-association is responsible for the loss of function of SMN in SMA. Our data indicates a model in which SMN WT and SMN D7 and other missense mutated SMN proteins are able to self-associate, but that the self-association interface created by the SMN D7 and missense mutated SMN proteins prevents binding of proteins essential for function (Figure 4.3a). The function and the specific protein that binds to the SMN self-association interface is currently unknown, but the function likely involves a specific step in the assembly of splicing or β-actin RNPs. In agreement with our model, Martin et al showed that several SMN mutant proteins with mutations in the C-terminal oligomerization domain do not effect the C-terminal self-association of SMN *in vitro* (Martin 2012). Our model also explains the highly variable SMA phenotypes as the observed phenotype would vary based on two factors, the oligomerization interfaces present and the *smn2* copy number (4.3b). The oligomerization interfaces of each of three possible combinations of SMN, SMNWT-SMN WT, SMN WT-SMN D7 and SMN D7-SMN D7 will be referred to from here on as WT-WT, WT-D7, and D7-D7. The three SMN interfaces would have a range of protein binding affinities with the WT-WT interface having the highest relative binding affinity to an unknown protein and D7-D7 having no binding affinity. In 95% of SMA patients lacking a functional SMN WT protein from the *smn1* gene due to a gene deletion in exon 7, the predominant interfaces present would be WT-D7 and D7-D7 creating a lower

protein binding affinity, which would then inhibit the function of SMN and result in the SMA phenotype (Figure 4.3b). The variability seen in SMA phenotype severity ranging from mild to severe would result from variability in the copy number of the *smn2* gene, which would change the ratio of WT-D7 and D7-D7 oligomer interfaces. For example an additional *smn2* copy number expressing 90% SMN D7 and 10% SMN WT would increase the amount of SMN D7-SMN D7 interfaces dramatically as well as some WT-D7 and WT-WT interfaces. However, since D7-D7 interfaces have no protein binding, the SMA phenotype should be less severe as a slight increase in the WT-D7 interfaces with partial protein binding affinity and WT-WT with high protein binding affinity would result in higher SMN function. This is in agreement with the general trend observed in SMA phenotypes, in which a higher *smn2* copy number correlates to a less severe phenotype (Wirth 2006b). Although there is a general trend relating severity (Type I-IV) with *smn2* copy number, phenotypic variability also exists within each disease type (Wirth 2000b). In our model this may be accounted for by SMA patients that have missense mutations in the C-terminal region of SMN, which would likely change the oligomer interface to varying degrees depending on the specific mutation as well as whether the mutation was homozygous or heterozygous.



a

People	*smn1*:*smn2*	Predicted Self-associations	Protein binding
Healthy	2:2	Mostly WT-WT and some WT-D7, D7-D7	High affinity
SMA patients	0:2-4	Mostly D7-D7 and some WT-WT, WT-D7	Low affinity

b

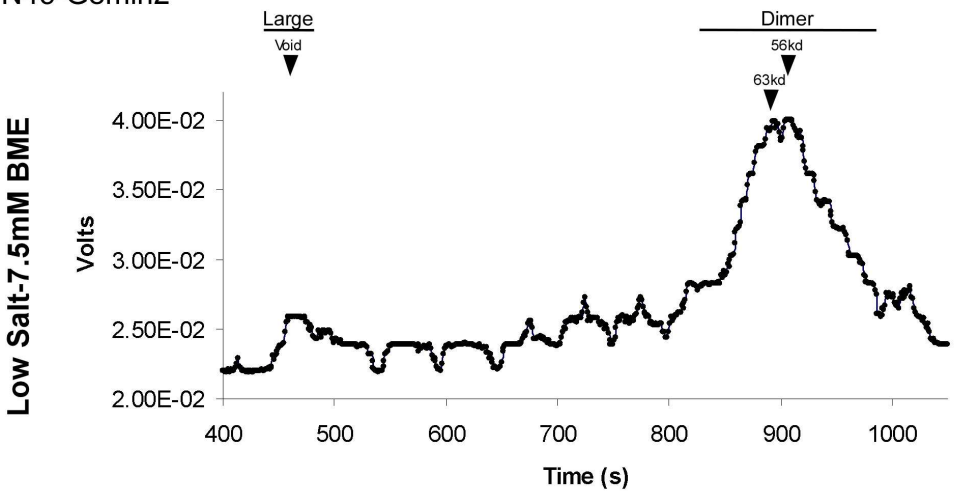
Figure 4.3- Model of the self-association of SMN in SMA
(s) SMN is proposed to self-associate in three possible combinations, SMN WT-SMN WT, SMN WT-SMN D7, and SMN D7-SMN D7 aided by Gemin2 dimerization (red). SMN WT is shown in purple and SMN D7 is shown in blue. Upon self-association, a new C-terminal self-association interface is created that binds an unknown protein with an unknown function. The relative protein binding affinity is high in SMN WT-SMN WT interfaces, low in SMN WT-SMN D7 interfaces and abolished in SMN D7-SMN D7 interfaces. (b) The relationship between *smn* copy number, predicted SMN self-association sites and protein binding affinity.

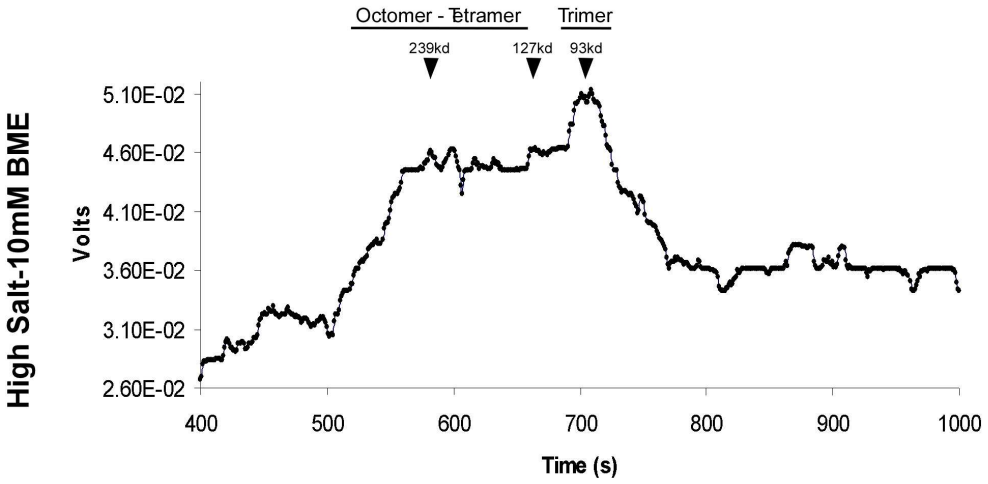
4.3.4 Sensitivity of N45-Gemin2 oligomerization state to salt and reducing agent concentration

Native Gemin2 has been shown to predominantly form dimers with a minor amount of monomers and trimers. SEC of N45-Gemin2 was performed under varying buffer conditions to ascertain if N45-Gemin2 forms similar oligomers as native Gemin2. N45-Gemin2 formed a dimer at 63kDa under Tris-low salt-7.5mM BME conditions with a minor amount of large oligomers at the void volume (Figure 4.4a). The SEC molecular weight of 63kDa was close to the actual molecular weight of 58kDa and the second 56kDa peak observed in SEC is

likely the degraded 25kDa N45-Gemin2 observed in SDS PAGE. Both high salt buffer conditions resulted in an array of N45-Gemin2 oligomer states (Figure 4.4b, c). Although the increased BME and salt concentration abolished the large oligomer peak seen in the low salt buffer, it caused the N45-Gemin2 to form a range of high-order oligomers. These results may explain the small crystals observed in N45-Gemin2 crystallography trials in Chapter 2, in which the protein was concentrated in buffer with high concentration of salt and BME. The fast forming microcrystals were likely caused by increased nucleation resulting from the increased self-association of N45-Gemin2 in high salt and BME buffer conditions. N45-Gemin2 formed a myriad of oligomerization states under varying buffer conditions. The low concentrated salt and BME buffer condition was chosen for SMN-Gemin2 interaction studies because N45-Gemin2 dimers are consistent with the dimer formation of native Gemin2 and the void volume oligomerization was abolished resulting in less aggregation and SEC column maintenance.

N45-Gemin2


Low Salt-7.5mM BME
Large
Void
Dimer
56kd
63kd
4.00E-02
3.50E-02
3.00E-02
2.50E-02
2.00E-02
Volts
400
500
600
700
800
900
1000
Time (s)


High Salt-10mM BME
Octomer - Tetramer
Trimer
239kd
127kd
93kd
5.10E-02
4.60E-02
4.10E-02
3.60E-02
3.10E-02
2.60E-02
Volts
400
500
600
700
800
900
1000
Time (s)

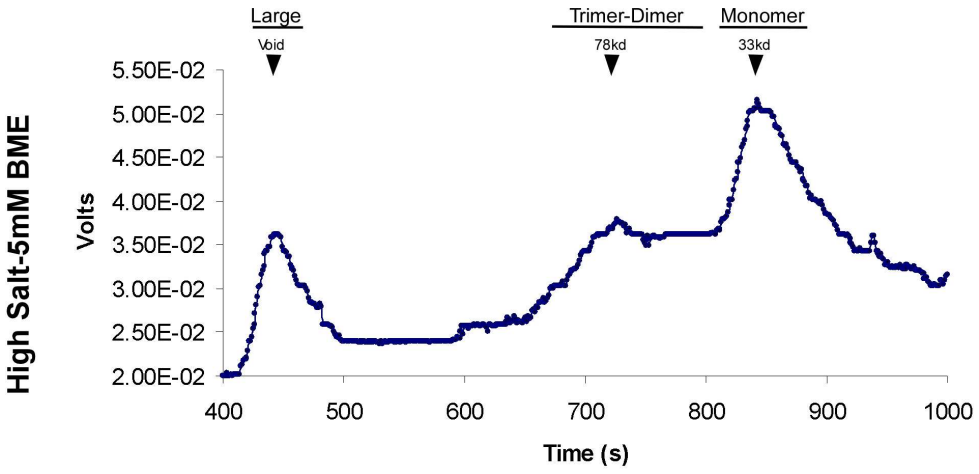

High Salt-5mM BME
Large
Void
Trimer-Dimer
78kd
Monomer
33kd
5.50E-02
5.00E-02
4.50E-02
4.00E-02
3.50E-02
3.00E-02
2.50E-02
2.00E-02
Volts
400
500
600
700
800
900
1000
Time (s)

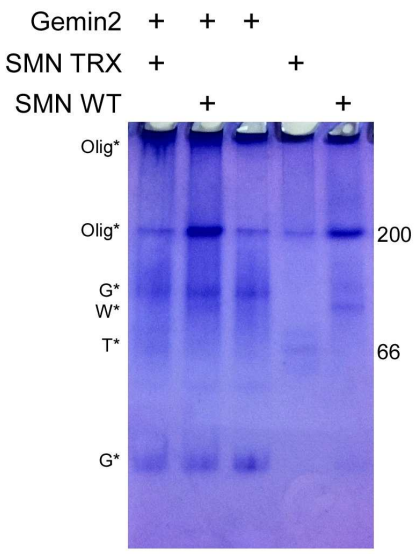
Figure 4.4- Gemin2 oligomerization states
N45-Gemin2 in elution buffer was concentrated and buffer exchanged into Tris buffer (20mM Tris pH 8.0) with either low salt (200mM NaCl) or high salt (500mM NaCl) and varying concentrations of BME. Protein samples were run on a superdex200 SEC column and eluted protein was detected by absorbance at 280nm (measured in Volts) every second. Predicted molecular weight (kDa) of the protein peaks was calculated based on a set of protein standards. Arrows indicate the protein peaks and the predicted molecular weight or void volume. The predicted Gemin2 oligomer(s) are indicated above the peaks. (a) Gemin2 SEC was run in Tris buffer with low salt and 7.5mM BME. (b) Gemin2 in Tris buffer with high salt and 10mM BME. (c) Gemin2 in Tris buffer with high salt and 5mM BME.

4.3.5 SMN bound to Gemin2 forms large oligomers and/or soluble aggregates

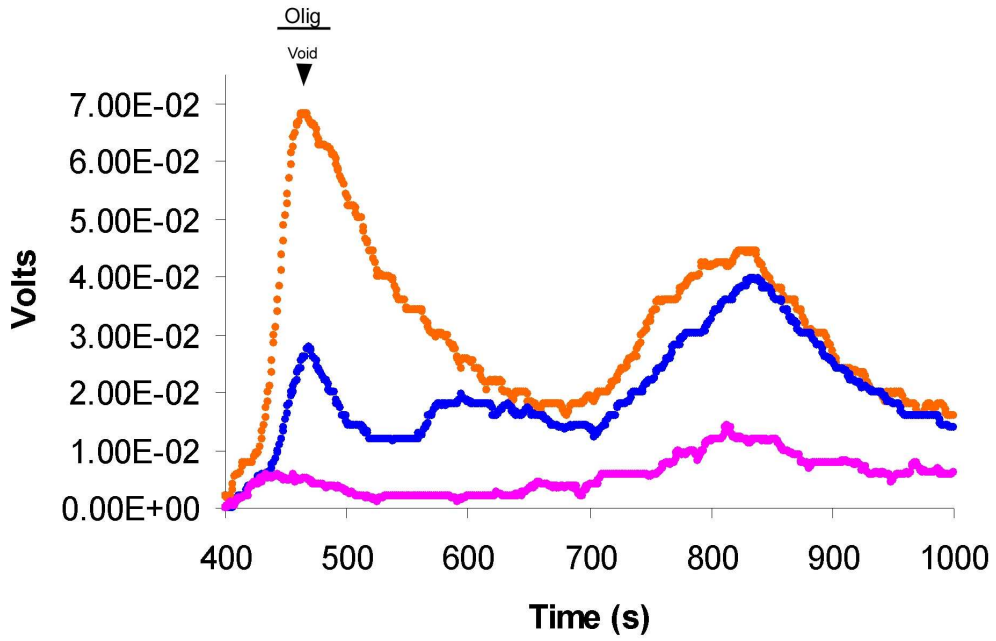
SMN WT and TRX were each incubated with or without N45-Gemin2, concentrated and analyzed via Native PAGE to verify a SMN-Gemin2 interaction. SMN WT, SMN TRX and N45-Gemin2 all three showed bands that migrated at the same distance as β-amylase (200kDa) (Figure 4.5a), which were likely large oligomers that cannot migrate further into the gel, as SMN TRX is twice the size of SMN WT and N45-Gemin2. SMN WT bound to N45-Gemin2 resulted in an intense oligomer band at 200kDa as well as a large amount of oligomer and/or aggregate protein in the well of the native PAGE gel. This suggested that SMN-Gemin2 forms large oligomers and /or soluble aggregates *in vitro*. SMN TRX also migrated as both the oligomer band at 200kDa as well as in the well, but the larger oligomer/aggregate was more prevalent. The native PAGE results indicate that both SMN WT and SMN TRX interact with Gemin2; however, it was difficult to draw any conclusions about the details of the interaction and the oligomers from this data because the control protein oligomers migrate at the same distance as the SMN-Gemin2 oligomers. Furthermore, SMN-Gemin2 was

more concentrated as some SMN WT and SMN TRX aggregated during concentration.

To gain further insight into the SMN-Gemin2 interaction, SMN WT was bound to N45-Gemin2 and run on SEC. SMN WT and Gemin2 alone were run on SEC as controls. SMN-Gemin2 eluted from the column as large oligomers or soluble aggregates at the void volume as well as at the same elution time as SMN WT and Gemin2 (Figure 4.5b). SEC of SMN-Gemin2 did not result in a discrete peak, but instead resulted in a broad peak at the void volume, suggesting SMN-Gemin2 formed an oligomer that is larger then 600kDa or a variety of unresolved large oligomers similar to SMN WT. SMN WT-Gemin2 formed large oligomers, which prevented the characterization of the SMN-Gemin2 interaction



a



b

Figure 4.5- The oligomerization state of SMN bound to Gemin2.
(a) Gemin2, SMN WT and SMN TRX alone or combined were concentrated and run on a 4-20% Native Tris gel. The asterisks indicate protein bands with the corresponding letter indicating the protein (G-Gemin2, W-SMN WT, T-SMN TRX, Olig-oligomer). 200kDa and 66kDa indicate the migration distance of b-amylase and BSA in native page. (b) SEC of SMN WT (blue), Gemin2 (pink), and SMN WT + Gemin2 (orange) buffer exchanged in and run with 50mM

sodium phosphate pH 7.4, 150mM NaCl. Arrow indicates the oligomers which eluted at the void volume.

4.3.6 The SMN/Gemin2 complex is sensitive to disulfide bond formation

We have previously shown that disulfide bonds may be important for SMN WT oligomer formation in Chapter 3. Since, the SMN-Gemin2 complex forms similar large and possibly even varying oligomeric states as SMN WT does, we next sought out to determine if SMN-Gemin2 oligomers *in vitro* were mediated by disulfide bond formation. SMN WT bound to Gemin2 and concentrated in elution buffer (S +G Elut) as well as SMN WT-Gemin2 previously concentrated in Tris buffer (S+G old) were both run on a SDS PAGE gel under reducing and non-reducing conditions. The SMN WT-Gemin2 in Tris buffer under non-reducing conditions completely shifted up to a molecular weight around 200kDa, whereas in the freshly prepared SMN WT-Gemin2, the SMN WT shifted up to a smear at 200kDa, but the Gemin2 did not (Figure 4.6a). This indicates that the SMN WT-Gemin2 interaction in the old sample was linked via disulfide bonds most likely induced by the gradual degradation of BME in the buffer. Another interesting point is that the degraded protein bands migrated with the oligomer at 200kDa in the old SMN WT-Gemin2 sample, whereas the smaller degradation bands in the freshly prepared SMN WT-Gemin2 sample did not. This suggested that buffer conditions without BME promoted non-specific disulfide bonds between full-length and degraded SMN WT proteins. The determination of non-specific disulfide bonds was important because the degradation of SMN WT and Gemin2 cannot be fully inhibited even in the

presence of protease inhibitors. Furthermore, in the non-reducing SMN WT-Gemin2 sample that was not boiled during sample preparation, the majority of SMN WT shifted primarily up to the 100kDa oligomer band observed in Chapter 3 instead of the large 200kDa oligomer when the sample was boiled. Either SMN WT forms a larger oligomer made up of the small 100kDa oligomers that is strongly associated and cannot be dissociated by SDS unless heat is applied or SMN WT 100kDa oligomers aggregate in the presence of SDS and heat. The later is less likely because the protein was able to enter the gel, suggesting it was soluble.

To further assess the role of disulfide bonds in SMN-Gemin2 complex formation, SMN WT and N45-Gemin2 were mixed at three different molar ratios (1:1, 1:3, and 3:1) and run on a SDS PAGE gel. The presence of the higher 100kDa oligomeric SMN WT band decreased as N45-Gemin2 increased in molar ratio relative to SMN WT (Figure 4.6b). This provided evidence that a structural change occurred upon SMN WT binding to Gemin2 that prevented the disulfide bond formation of SMN WT 100kDa oligomers. In addition, because the intensity of the full-length SMN WT and Gemin2 bands in the monomer form were calculated to be present at a 1:1 ratio with the excess SMN WT in the 100kDa form (data not shown), SMN WT and Gemin2 likely bind at a 1:1 ratio.

Similar to SMN WT, the SMN-Gemin2 interaction is sensitive to disulfide bond formation. SMN WT may undergo a change in structural confirmation and/or oligomerization state upon binding to Gemin2.

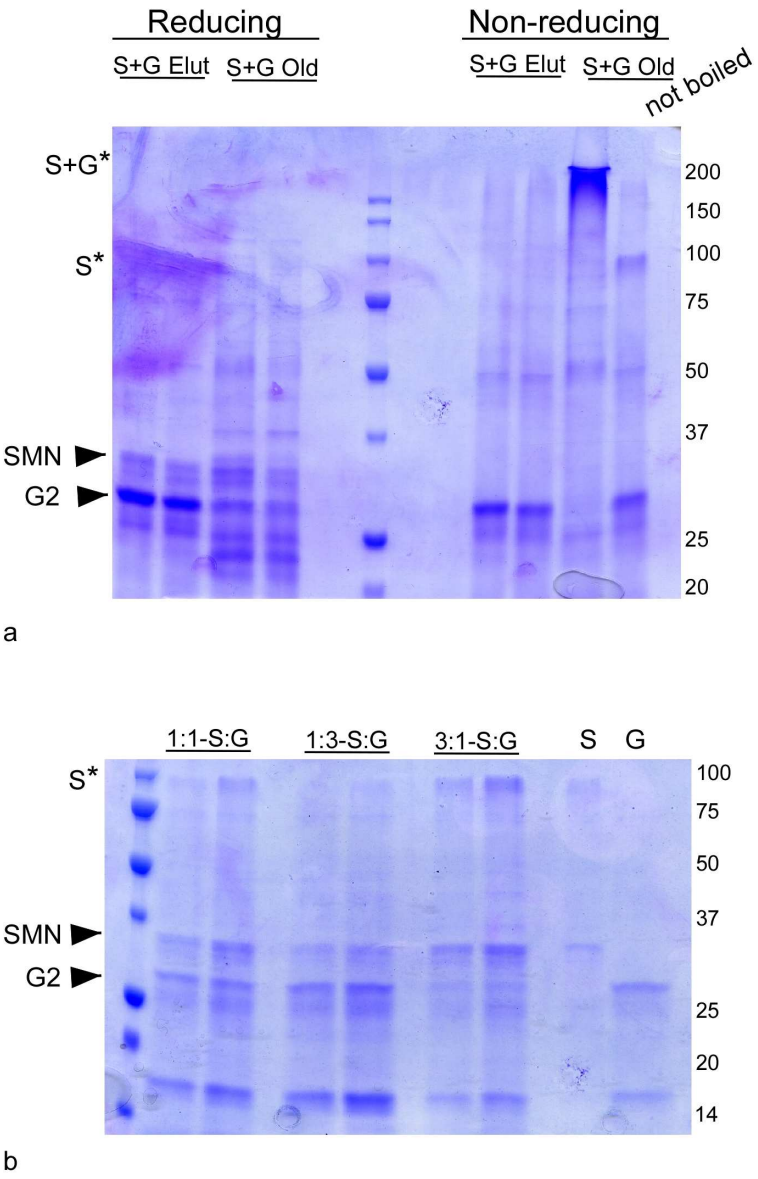


Figure 4.6- Sensitivity of the SMN-G2 complex to disulfide bond formation.
(a) SMN WT was mixed with Gemin2 at equal molar ratios and concentrated in elution buffer after purification (S+ Elut). SMN WT and Gemin2 were mixed at equal molar ratios and buffer exchanged into 20mM Tris pH 8.0, 200mM NaCl, 5mM BME and incubated at 4°C for several weeks (S+G old). All samples were run on an SDS PAGE gel with (reducing) or without (non-reducing) BME in the loading dye. The last sample was not boiled. (b) SMN WT and Gemin2 were mixed at 1:1, 1:3 and 3:1 ratios and concentrated. All S+G samples along with SMN WT (S) and Gemin2 (G) alone were run on an SDS PAGE gel. For both gels, arrows indicate SMN WT and Gemin2 and asterisks indicate oligomers of either SMN WT (S*) or SMN-G2(S+G*).

4.3.7 Comparison of the SMN-Gemin2 interaction between SMN WT and SMN D7

SMN D7 and SMN WT bound to N45-Gemin2 were run on SEC to determine if the SMN-Gemin2 interaction is different between SMN WT and the SMN D7. SMN WT-Gemin2 showed two major peaks at 140kDa and at the void volume (Figure 4.7a). The 140 kDa peak was consistent with the addition of the SMN WT peak and the Gemin2 peak and based on the predicted dimer state of Gemin2 and the proposed dimer state of SMN WT, it was likely a heterodimer of two SMN proteins and two Gemin2 proteins. SMN D7-Gemin2 unexpectedly formed heterodimers and larger oligomers, although the ratio of the heterodimer peak to the oligomer peaks was higher when compared to SMN WT (Figure 4.7b). A SMN D7-Gemin2 peak of 91kDa was also observed. This 91kDa peak was slightly larger then the SMN D7 only control of 85kDa. With the resolution of the SEC chromatogram, it is unclear if the 91kDa peak was SMN D7 or another oligomeric state of SMN-Gemin2.

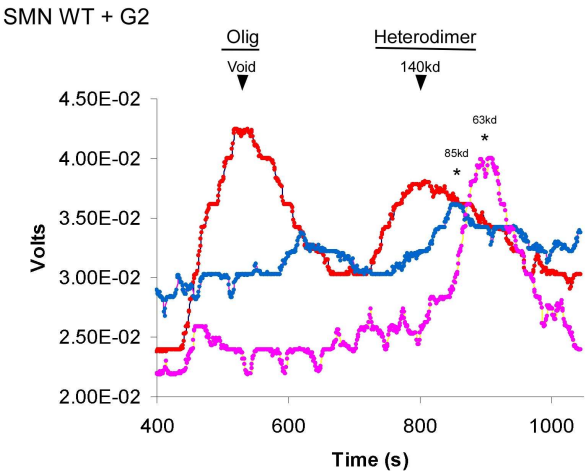
SMN D7 bound to Gemin2 formed similar oligomeric states as SMN WT bound Gemin2, but the equilibrium of the states were different as SMN WT-Gmin2 favored higher-order oligomers and SMN D7-Gemin2 favored the formation of a heterodimer.

SDS PAGE of the input SEC protein samples was run to verify the purity and ratio of SMN and Gemin2 proteins. All SMN WT-Gemin2 SEC samples migrated as monomers in SDS PAGE (Figure 4.7c). The SMN D7 in the protein only control sample migrated as 100kDa oligomers whereas half of the SMN D7 in the SMN D7-Gemin2 sample migrated as 100kDa oligomers. The SDS PAGE

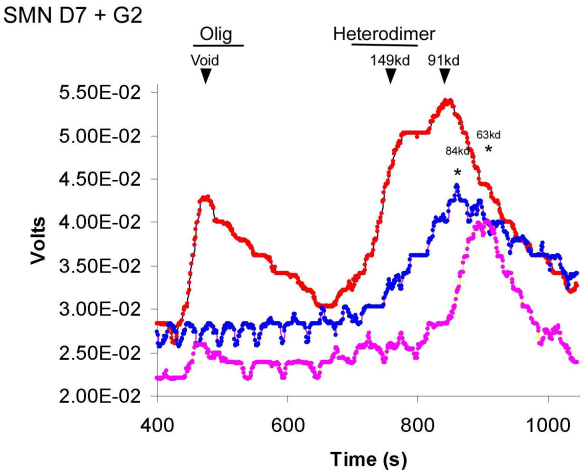
gel suggested that SMN D7 in the SMN D7-Gemin2 sample existed in two oligomeric states. Again, it is difficult to draw firm conclusions without identification of the proteins that eluted from column. However, the second oligomeric state of SMN D7 that migrated at 100kDa on the SDS PAGE gel may correspond to the additional 91kDa peak observed in the SEC chromatogram of SMN D7-Gemin2. If this was, indeed true, it suggests that disulfide-linked SMN D7 was inhibited from binding Gemin2. Further experiments to duplicate these results would need to be done for verification followed by the exploration of the disulfide bond sensitivity difference between SMN WT and D7 and the effect on Gemin2 binding. SMN WT and Gemin2 concentrated down to a 1:1 ratio that was more concentrated then SMN WT supporting the 1:1 binding ratio and higher stability of SMN WT-Gemin2. SMN D7 and Gemin2 did not concentrate down to a 1:1 ratio, which could be due to either inefficient binding of SMN D7 to Gemin2 or insufficient full-length SMN D7 as SMN D7 on SDS PAGE also showed a degradation product at 25kDa.

SMN WT and SMN D7 both bound Gemin2 at a 1:1 ratio suggesting that SMN-Gemin2 binding alone does not lead to SMA. In accordance with our results that insufficient Gemin2 binding does not cause the SMA phenotype, a SMA missense mutation D44V, located in the Gemin2 binding domain of SMN, was originally shown to disrupt Gemin2-SMN binding *in vitro*, but recently Sarachan et al discovered the disruption in Gemin2 binding was facilitated by detergent in the experimental system (Sarachan 2012). SMN D7 bound to Gemin2, but did not form higher-order oligomers like SMN WT. Our data

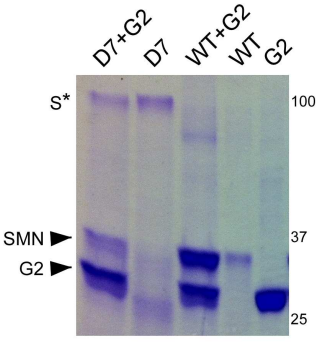
indicated SMN WT-Gemin2 oligomers are predominantly facilitated by the oligomerization of SMN WT while Gemin2 enhances oligomerization, as SMN D7 did not form higher-order oligomers unless it was bound to Gemin2.



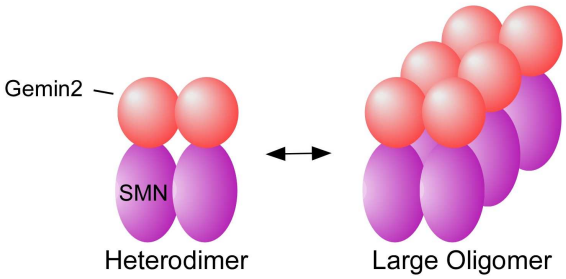
a



b



c



d

Figure 4.7- Comparison of the Gemin2-SMN interaction between SMN WT and D7.
(a) SMN WT and D7 were each mixed with equal molar ratios of Gemin2 and buffer exchanged into Tris buffer (20mM Tris pH 8.0, 200mM NaCl, 7.5mM BME). Protein samples were run on a superdex 200 SEC column and eluted protein was detected by absorbance at 280nm (measured in Volts) every second. Predicted molecular weight (kDa) of the protein peaks was calculated based on a set of protein standards. The chromatogram of SMN WT alone (blue) alone, Gemin2 alone (pink) and SMN WT/Gemin2 (orange) with arrows indicating the peaks of SMN-Gemin2 and asterisks indicating peaks of SMN WT and Gemin2 protein controls. The predicted SMN/Gemin2 oligomers are indicated above the peaks. (b) The SEC chromatogram of SMN D7 (blue) alone, Gemin2 (pink alone) and SMN D7/Gemin2 (orange) with arrows indicating the peaks of SMN-Gemin2 and asterisks indicating peaks of SMN D7 and Gemin2 protein controls. (c) SDS PAGE analysis of the input protein samples loaded onto the SEC column. Arrows indicate SMN WT-Gemin2 and asterisks indicate oligomers of SMN (S*). (d) The predicted model of SMN-Gemin2 oligomers from SEC results. SMN-Gemin2 form heterodimers and large oligomers composed of heterodimers subunits.

We propose a model of the oligomerization of SMN-Gemin2 in the SMA disease. The first part of the model predicts the equilibrium of SMN-Gemin2 oligomers arising from the mixture of SMN WT and SMN D7 (Figure 4.7d). The equilibrium is between the two SMN-Gemin2 states we observed in SEC, heterodimers and the large oligomers. As discussed in Section 4.3.3 the mixture of SMN WT and SMN D7 *in vivo* likely leads to three possible combinations of the self-association of SMN, WT- WT, WT-D7 and D7-D7. We predict that WT-WT favors SMN-Gemin2 large oligomers, WT-D7 favors an equal mixture of large oligomers and heterodimers, and D7-D7 favors heterodimers. We predict that the large oligomers and heterodimers of SMN-Gemin2 each have a specific function. In deed, it has been shown *in vitro* that only higher-order oligomers of SMN WT bound to SmB protein suggesting a function in snRNP assembly or sequestration and storage of RNP proteins (Pellerzzoni 1999). The function(s) of

the heterodimer is unknown, but the assembly of the SMN complex is likely to be one of these functions as it is unlikely that all the proteins in the SMN complex would be able to bind to SMN-Gemin2 large oligomers. Assuming that each WT-WT, WT-D7 or D7-D7 correlate to a specific equilibrium of large oligomers and heterodimers and each of those oligomeric states have a specific function, the amount of SMN WT and D7 produced from the *smn1* and *smn2* genes would modulate those two functions. The last part of the model is to correlate SMN-Gemin2 oligomerization modulation to the SMA phenotype. To do this, we estimated the three combinations of SMN WT and D7 in individuals that are healthy, SMA carriers, and SMA patients. These approximations were based on observed relative *smn* gene copies as well as the higher affinity WT-WT versus WT-D7 self-associations (Young 2000b, Diep Tran 2001, Kao 2006). Healthy individuals with two of each *smn* gene copy are predicted to have WT-WT SMN-Gemin2 large oligomers, D7-D7 SMN-Gemin2 heterodimers and some WT-D7 SMN-Gemin2 containing both oligomeric forms. SMA carriers are predicted to have WT- WT SMN-Gemin2 large oligomers and slightly more WT-D7 and D7-D7 heterodimers as they have more copies of *smn2*. SMA patients are predicted to have a majority of D7-D7 and D7-WT SMN-Gemin2 heterodimers because they predominantly have SMN D7 being produced from the *smn2* gene. We predict that healthy individuals and SMA carriers do not show a neurodegenerative phenotype because they have enough WT-WT SMN Gemin2 to form large oligomers and perform the function associated with those oligomers, whereas SMA patients with SMN D7 only form SMN-Gemin2 heterodimers and

therefore lose the large oligomer function. One interesting observation in agreement with this model is several recent genetic studies correlating the prevalence of neurodegenerative diseases with *smn* gene copy numbers. These studies showed that individuals with either a homozygous deletion of *smn2* or a lower ratio of *smn2* copies relative to *smn1* copies have a higher incidence of sporadic lower motor degradation (LMD) disease and sporadic amytrophic lateral sclerosis (ALS) (Blauw 2012, Lee 2012, Kim 2010, Corcia 2009). In our model, the higher *smn2*:*smn1* gene copy ratio of these LMD and ALS patients would results in a more WT-WT SMN. The increased amounts of WT-WT SMN interfaces would then lead to an equilibrium shift to SMN-Gemin2 large oligomers and a neurodegenerative phenotype by limiting or abolishing the function of the SMN-Gemin2 heterodimers. Only 10-14% of the population was projected to have homozygous deletions of the *smn2* gene, which was significantly higher then the projected incidence of ALS and LMD in individuals lacking the *smn2* gene (Workman 2009). However, these initial genetic studies had a small population size and more genetic studies would give more insight into the correlation between smn copy number and neurodegenerative diseases. This model is a simplified explanation of the effects of SMN-Gemin2 oligomerization in SMA. Also of note, is that the SMN-Gemin2 oligomerization modulation model is partially exclusive from the SMN self-association model presented in Section 4.3.3. The SMN self-association model is based on the assumption that SMN D7-D7 self-association interfaces do not have protein binding affinity and therefore a loss of function occurs. However, the SMN-Gemin2 oligomerization

modulation model is based on the assumption that each oligomeric state of SMN-Gemin2 has a specific function. We currently do not know whether one or several functions of SMN lead to the neurodegenerative SMA phenotype in motor neurons and not other cell types. Research on the functions SMN and SMN-Gemin2 oligomers including the mixture of SMN WT and SMN D7 will lead to a better understanding of the molecular mechanism of SMA as well as providing specific drug targets to speed up the SMA drug discovery process.

Chapter 5

CONCLUSION

5.1 CONCLUSION

SMA is a fatal neurodegenerative disease caused by low levels of SMN WT protein, which currently does not have a treatment. We determined that SMN WT forms a mixture of oligomers whereas SMN D7 does not. We characterized the interaction of SMN-Gemin2 and found that the proteins associated with a 1:1 ratio. SMN D7 was unexpectedly able to self-associate *in vitro*, which was enhanced upon binding to Gemin2. We predict two models that link both SMN self-association interfaces and SMN-Gemin2 oligomerization to the SMA disease.

5.2 FUTURE WORK

The future work of this project would be to continue the *in vitro* characterization of SMN oligomerization and the SMN-Gemin2 interaction as well as reproducing the SMN-Gemin2 interaction results and verifying the peaks of the SEC chromatogram via western blot analysis. An understanding of the specific oligomeric states would optimize the experimental system and further allow for the characterization of SMA mutant proteins as well as mixtures of SMN mutant, SMN D7 and SMN D7 proteins, which more closely resembles the molecular environment of SMA *in vivo*. Another important future work is to model the SMN-Gemin2 interaction using our low-resolution structures of SMN WT and D7, the published Gemin2 structure, and our results about the Gemin2-SMN interaction.

REFERENCES

Bachand, F. (2002) The Product of the Survival of Motor Neuron (SMN) Gene is a Human Telomerase-associated Protein, *Molecular Biology of the Cell*, **13**, 3192-3202.

Battle, D.J., *et al.* (2007) SMN-independent subunits of the SMN complex. Identification of a small nuclear ribonucleoprotein assembly intermediate, *J Biol Chem*, **282**, 27953-27959.

Battle, D.J., *et al.* (2006) The SMN complex: an assembly machine for RNPs, *Cold Spring Harb Symp Quant Biol*, **71**, 313-320.

Battle, D.J., *et al.* (2006)b The Gemin5 protein of the SMN complex identifies snRNAs, *Mol Cell*, **23**, 273-279.

Baumer, D., *et al.* (2010) The role of RNA processing in the pathogenesis of motor neuron degeneration, *Expert Rev Mol Med*, **12**, e21.

Blauw, H.M., *et al.* (2012) SMN1 gene duplications are associated with sporadic ALS, *Neurology*, **78**, 776-780.

Buhler, D., *et al.* (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy, *Hum Mol Genet*, **8**, 2351-2357.

Burghes, A.H. and Beattie, C.E. (2009) Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?, *Nat Rev Neurosci*, **10**, 597-609.

Burglen, L., *et al.* (1995) [Identification of the gene determining spinal muscular atrophy: perspectives], *Arch Pediatr*, **2**, 505-507.

Burnett, B.G., *et al.* (2009) Regulation of SMN protein stability, *Mol Cell Biol*, **29**, 1107-1115.

Burnett, B.G. and Sumner, C.J. (2008) Targeting splicing in spinal muscular atrophy, *Ann Neurol*, **63**, 3-6.

Campanaro, B.M. (2009) Insights into the molecular weight and post-translational modifications of the survival of motor neuron protein. PhD Thesis. Arizona State University, Tempe, pp. 222.

Capon, F., *et al.* (1995) De novo deletions of the 5q13 region and prenatal diagnosis of spinal muscular atrophy, *Prenat Diagn*, **15**, 93-94.

Carissimi, C., *et al.* (2006) Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly, *J Biol Chem*, **281**, 8126-8134.

Charroux, B., *et al.* (2000) Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli, *J Cell Biol*, **148**, 1177-1186.

Corcia, P., *et al.* (2009) The importance of the SMN genes in the genetics of sporadic ALS, *Amyotroph Lateral Scler*, **10**, 436-440.

Crawford, T.O. and Pardo, C.A. (1996) The neurobiology of childhood spinal muscular atrophy, *Neurobiol Dis*, **3**, 97-110.

Diep Tran, T., *et al.* (2001) The gene copy ratios of SMN1/SMN2 in Japanese carriers with type I spinal muscular atrophy, *Brain Dev*, **23**, 321-326.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. (2004) The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3), *Protein Expr Purif*, **37**, 203-206.

Eggert, C., *et al.* (2006) Spinal muscular atrophy: the RNP connection, *Trends Mol Med*, **12**, 113-121.

Fallini, C., Bassell, G.J. and Rossoll, W. (2012) Spinal muscular atrophy: the role of SMN in axonal mRNA regulation, *Brain Res*, **1462**, 81-92.

Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis, *Cell*, **90**, 1023-1029.

Friesen, W.J., *et al.* (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets, *Mol Cell*, **7**, 1111-1117.

Fuentes, J.L., Strayer, M.S. and Matera, A.G. (2010) Molecular determinants of survival motor neuron (SMN) protein cleavage by the calcium-activated protease, calpain, *PLoS One*, **5**, e15769.

Giesemann, T., *et al.* (1999) A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with smn in nuclear gems, *J Biol Chem*, **274**, 37908-37914.

Grimmler, M., *et al.* (2005) Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization, *Hum Mol Genet*, **14**, 3099-3111.

Hamamoto, S., *et al.* (2006) Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo, *J Virol*, **80**, 5670-5677.

Hamilton, S., *et al.* (2003) Effect of imidazole on the solubility of a his-tagged antibody fragment, *Hybrid Hybridomics*, **22**, 347-355.

Helmken, C., *et al.* (2000) An essential SMN interacting protein (SIP1) is not involved in the phenotypic variability of spinal muscular atrophy (SMA), *Eur J Hum Genet*, **8**, 493-499.

Hua, Y., *et al.* (2011) Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model, *Nature*, **478**, 123-126.

Humphrey, E., Fuller, H.R. and Morris, G.E. (2012) Current research on SMN protein and treatment strategies for spinal muscular atrophy, *Neuromuscul Disord*, **22**, 193-197.

Jablonka, S., *et al.* (2001) Co-regulation of survival of motor neuron (SMN) protein and its interactor SIP1 during development and in spinal muscular atrophy, *Hum Mol Genet*, **10**, 497-505.

Jablonka, S., *et al.* (2002) Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death, *Proc Natl Acad Sci U S A*, **99**, 10126-10131.

Jana, S. and Deb, J.K. (2005) Strategies for efficient production of heterologous proteins in Escherichia coli, *Appl Microbiol Biotechnol*, **67**, 289-298.

Jones, K.W., *et al.* (2001) Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin, *J Biol Chem*, **276**, 38645-38651.

Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli, *Curr Opin Biotechnol*, **6**, 494-500.

Kao, H.Y., *et al.* (2006) Determination of SMN1/SMN2 gene dosage by a quantitative genotyping platform combining capillary electrophoresis and MALDI-TOF mass spectrometry, *Clin Chem*, **52**, 361-369.

Kim, J., *et al.* (2010) Association between survivor motor neuron 2 (SMN2) gene homozygous deletion and sporadic lower motor neuron disease in a Korean population, *Ann Clin Lab Sci*, **40**, 368-374.

Le, T.T., *et al.* (2005) SMNDelta7, the major product of the centromeric survival motor neuron (SMN2) gene, extends survival in mice with spinal muscular atrophy and associates with full-length SMN, *Hum Mol Genet*, **14**, 845-857.

Lee, J.B., *et al.* (2012) Homozygous SMN2 deletion is a major risk factor among twenty-five Korean sporadic amyotrophic lateral sclerosis patients, *Yonsei Med J*, **53**, 53-57.

Lee, M.B., *et al.* (2005) The DEAD-box protein DP103 (Ddx20 or Gemin-3) represses orphan nuclear receptor activity via SUMO modification, *Mol Cell Biol*, **25**, 1879-1890.

Lefebvre, S., *et al.* (1995) Identification and characterization of a spinal muscular atrophy-determining gene, *Cell*, **80**, 155-165.

Lefebvre, S., *et al.* (2002) A novel association of the SMN protein with two major non-ribosomal nucleolar proteins and its implication in spinal muscular atrophy, *Hum Mol Genet*, **11**, 1017-1027.

Leung, A.K. and Nagai, K. (2005) Gemin 6 and 7 lend a hand to snRNP assembly, *Structure*, **13**, 833-834.

Liu, Q., *et al.* (1997) The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins, *Cell*, **90**, 1013-1021.

Lorson, C.L. and Androphy, E.J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN, *Hum Mol Genet*, **9**, 259-265.

Lorson, C.L., *et al.* (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy, *Proc Natl Acad Sci U S A*, **96**, 6307-6311.

Lorson, C.L., *et al.* (1998) SMN oligomerization defect correlates with spinal muscular atrophy severity, *Nat Genet*, **19**, 63-66.

Lorson, M.A., *et al.* (2008) Identification and characterisation of a nuclear localisation signal in the SMN associated protein, Gemin4, *Biochem Biophys Res Commun*, **375**, 33-37.

Ma, Y., *et al.* (2005) The Gemin6-Gemin7 Heterodimer from the Survival of Motor Neurons Complex Has an Sm Protein-like Structure, *Structure*, **13**, 883-892.

Mailman, M.D., *et al.* (2002) Molecular analysis of spinal muscular atrophy and modification of the phenotype by SMN2, *Genet Med*, **4**, 20-26.

Martin, R., *et al.* (2012) The Survival Motor Neuron Protein Forms Soluble Glycine Zipper Oligomers, *Structure*.

McAllister, G., Amara, S.G. and Lerner, M.R. (1988) Tissue-specific expression and cDNA cloning of small nuclear ribonucleoprotein-associated polypeptide N, *Proc Natl Acad Sci U S A*, **85**, 5296-5300.

McAndrew, P.E., *et al.* (1997) Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number, *Am J Hum Genet*, **60**, 1411-1422.

Meister, G., *et al.* (2000) Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins, *Hum Mol Genet*, **9**, 1977-1986.

Meister, G., *et al.* (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs, *Nat Cell Biol*, **3**, 945-949.

Mourelatos, Z., *et al.* (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs, *Genes Dev*, **16**, 720-728.

Nguyen thi, M., *et al.* (2008) A two-site ELISA can quantify upregulation of SMN protein by drugs for spinal muscular atrophy, *Neurology*, **71**, 1757-1763.

Nielsen, M.M., *et al.* (2007) Unfolding of beta-sheet proteins in SDS, *Biophys J*, **92**, 3674-3685.

Ogawa, C., *et al.* (2007) Gemin2 plays an important role in stabilizing the survival of motor neuron complex, *J Biol Chem*, **282**, 11122-11134.

Otter, S., *et al.* (2007) A comprehensive interaction map of the human survival of motor neuron (SMN) complex, *J Biol Chem*, **282**, 5825-5833.

Paushkin, S., *et al.* (2002) The SMN complex, an assemblyosome of ribonucleoproteins, *Curr Opin Cell Biol*, **14**, 305-312.

Pellizzoni, L., Charroux, B. and Dreyfuss, G. (1999) SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins, *Proc Natl Acad Sci U S A*, **96**, 11167-11172.

Pellizzoni, L., *et al.* (2001) A functional interaction between the survival motor neuron complex and RNA polymerase II, *J Cell Biol*, **152**, 75-85.

Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly, *Science*, **298**, 1775-1779.

Petri, S., *et al.* (2007) Dephosphorylation of survival motor neurons (SMN) by PPM1G/PP2Cgamma governs Cajal body localization and stability of the SMN complex, *J Cell Biol*, **179**, 451-465.

Pillai, R.S., *et al.* (2003) Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing, *Genes Dev*, **17**, 2321-2333.

Rossoll, W. and Bassell, G.J. (2009) Spinal muscular atrophy and a model for survival of motor neuron protein function in axonal ribonucleoprotein complexes, *Results Probl Cell Differ*, **48**, 289-326.

Rossoll, W., *et al.* (2002) Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons?, *Hum Mol Genet*, **11**, 93-105.

Samuelson, J.C. (2011) Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations, *Methods Mol Biol*, **705**, 195-209.

Sarachan, K.L., *et al.* (2012) Solution structure of the core SMN-Gemin2 complex, *Biochem J*.

Schmauss, C., *et al.* (1989) A comparison of snRNP-associated Sm-autoantigens: human N, rat N and human B/B', *Nucleic Acids Res*, **17**, 1733-1743.

Selenko, P., *et al.* (2001) SMN tudor domain structure and its interaction with the Sm proteins, *Nat Struct Biol*, **8**, 27-31.

Seng, C. (unpublished) The Structure of SMN. Arizona State University, Tempe Arizona.

Sharpe, N.G., *et al.* (1989) Isolation of cDNA clones encoding the human Sm B/B' auto-immune antigen and specifically reacting with human anti-Sm auto-immune sera, *FEBS Lett*, **250**, 585-590.

Swietnicki, W. (2006) Folding aggregated proteins into functionally active forms, *Curr Opin Biotechnol*, **17**, 367-372.

Takizawa, Y., *et al.* (2010) GEMIN2 promotes accumulation of RAD51 at double-strand breaks in homologous recombination, *Nucleic Acids Res*, **38**, 5059-5074.

Talbot, K., *et al.* (1997) Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism?, *Hum Mol Genet*, **6**, 497-500.

Thompson, T.G., *et al.* (1995) A novel cDNA detects homozygous microdeletions in greater than 50% of type I spinal muscular atrophy patients, *Nat Genet*, **9**, 56-62.

Todd, A.G., *et al.* (2010) SMN, Gemin2 and Gemin3 associate with beta-actin mRNA in the cytoplasm of neuronal cells in vitro, *J Mol Biol*, **401**, 681-689.

Todd, A.G., *et al.* (2010b) Analysis of SMN-neurite granules: Core Cajal body components are absent from SMN-cytoplasmic complexes, *Biochem Biophys Res Commun*, **397**, 479-485.

Tsai, L.K. (2012) Therapy Development for Spinal Muscular Atrophy in SMN Independent Targets, *Neural Plast*, **2012**, 456478.

Walia, R., Deb, J.K. and Mukherjee, K.J. (2008) Stability studies with different vector backbones utilizing the T7 expression system inEscherichia coli, *Journal of Chemical Technology & Biotechnology*, **83**, 1120-1125.

Wan, L., *et al.* (2008) Inactivation of the SMN complex by oxidative stress, *Mol Cell*, **31**, 244-254.

Wang, J. and Dreyfuss, G. (2001) Characterization of functional domains of the SMN protein in vivo, *J Biol Chem*, **276**, 45387-45393.

Wirth, B., Brichta, L. and Hahnen, E. (2006) Spinal muscular atrophy: from gene to therapy, *Semin Pediatr Neurol*, **13**, 121-131.

Wirth, B., *et al.* (2006b) Mildly affected patients with spinal muscular atrophy are partially protected by an increased SMN2 copy number, *Hum Genet*, **119**, 422-428.

Workman, E., Kolb, S.J. and Battle, D.J. (2012) Spliceosomal small nuclear ribonucleoprotein biogenesis defects and motor neuron selectivity in spinal muscular atrophy, *Brain Res*, **1462**, 93-99.

Workman, E., *et al.* (2009) A SMN missense mutation complements SMN2 restoring snRNPs and rescuing SMA mice, *Hum Mol Genet*, **18**, 2215-2229.

Xiao, J., *et al.* (2011) Discovery, synthesis, and biological evaluation of novel SMN protein modulators, *J Med Chem*, **54**, 6215-6233.

Young, P.J., *et al.* (2000) The relationship between SMN, the spinal muscular atrophy protein, and nuclear coiled bodies in differentiated tissues and cultured cells, *Exp Cell Res*, **256**, 365-374.

Young, P.J., *et al.* (2000b) The exon 2b region of the spinal muscular atrophy protein, SMN, is involved in self-association and SIP1 binding, *Hum Mol Genet*, **9**, 2869-2877.

Zhang, H., *et al.* (2006) Multiprotein complexes of the survival of motor neuron protein SMN with Gemins traffic to neuronal processes and growth cones of motor neurons, *J Neurosci*, **26**, 8622-8632.

Zhang, R., *et al.* (2011) Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly, *Cell*, **146**, 384-395.