Functional Characterization of Snurportin-1 C-Terminal Domain in The Context of a Novel Muscular Dystrophy

by

Burak Sarıbaş

A Dissertation Submitted to the
Graduate School of Health Sciences
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Cellular and Molecular Medicine

KOÇ
ÜNİVERSİTESİ

March 18, 2024

Functional Characterization of Snurportin-1 C-Terminal Domain in The Context of a Novel Muscular Dystrophy

Koç University

Graduate School of Health Sciences

This is to certify that I have examined this copy of a master's thesis by

Burak Sarıbaş

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assist. Prof. Nathalie Escande Beillard (Advisor)

Prof. Tuğba Bağcı Önder

Prof. Arzu Çelik Fuss

Date: 18.03.2024

The most genuine guide for everything in the world, whether for material gains, spirituality, or success, is knowledge and science. Seeking guidance outside of knowledge and science is negligence, ignorance, and a sign of misguidance.

- Mustafa Kemal Atatürk

In dedication to Türkiye, my greatest inspiration. This thesis is a heartfelt acknowledgment of the resilient spirit that has shaped me and my academic journey.

ABSTRACT

Functional Characterization of Snurportin-1 C-Terminal Domain in The Context of a Novel Muscular Dystrophy

Burak Sarıbaş

Master of Science in Cellular and Molecular Medicine

March 18,2024

SNUPN gene encodes for a nuclear import adaptor protein Snurportin 1 (SPN1) which participates in the nuclear transport of small nuclear ribonucleoproteins (snRNPs), by specifically binding to their m3G-cap. Nucleocytoplasmic shuttling is a highly versatile process that plays a major role in the maturation of pre-mRNA before translation, by facilitating the formation of the spliceosome complex through the transport of snRNPs.

The spliceosome is a large RNA-protein complex that facilitates the removal of the introns from nuclear pre-mRNA. Mutations in proteins involved in the nucleocytoplasmic machinery can impair spliceosome formation, leading to abnormal splicing, and are reported to be causative for various genetic disorders, including muscular dystrophies. Muscular Dystrophies includes a variety of genetically and phenotypically heterogeneous disorders characterized by progressive muscle weakness, often manifesting in infancy or early childhood. To this date, muscular dystrophies have been associated with more than 85 genes, however, 40% of the cases do not have a definitive molecular diagnosis.

Our group recruited eighteen patients from fifteen unrelated families diagnosed with muscular dystrophy. All patients presented with muscular weakness, sporadically accompanied by either neurological defects and/or cataracts. By performing whole exome sequencing we uncovered novel germline homozygous or compound heterozygous variants in the *SNUPN* gene across all patients. Notably, no pathogenic variants of this gene have previously been linked to muscular defects. We hypothesized that *SNUPN* mutations result in structural defects in SPN1 that impair the nucleocytoplasmic transport of the UsnRNPs.

Interestingly, all the SPN1 pathogenic variants are localized in the poorly defined C-terminal region of the protein. This region is known to be taking part in intramolecular interactions. The primary aim of this study was to develop tools and conduct in vitro assays to further characterize the function of the C-terminal region of SPN1 and ultimately investigate the protein's ability to carry out intermolecular interactions via its C-terminal region.

To address this question, we first generated *SNUPN* wild-type and mutant DNA constructs in the pCS2+ mammalian expression vector by molecular cloning and site-directed mutagenesis. Next, biochemical experiments were carried out by overexpressing WT and mutants of SPN1 in HEK293T/HeLa cells to compare their levels, localization, and interactions. Through these experiments, we aimed to unravel the importance of the C-terminal region of the protein and the impact of the mutations on SPN1's function in self-oligomerization. We hypothesize that pathogenic *SNUPN* variants may disrupt self-oligomerization of the C-terminal region of SPN1 thereby altering the cellular function of the protein. This may ultimately impair the nucleocytoplasmic shuttling of the UsnRNPs' leading to spliceosome defects and the observed phenotype in patients. Overall, this study aims to highlight the essential role of the uncharacterized C-terminal region of the SPN1 protein.

ÖZETÇE

Yeni Bir Musküler Distrofi Bağlamında Snurportin-1 C-Terminal Bölgesinin İşlevsel Karakterizasyonu

Burak Sarıbaş

Hücresel ve Moleküler Tıp

March 18, 2024

SNUPN geni, nükleer taşıma adaptör protein Snurportin 1 (SPN1)'i kodlar, bu protein küçük nükleer ribonükleoproteinlerin (snRNP'ler) nükleer taşımasına katılır ve özel olarak m3G-cap'lerine bağlanarak bu süreci sağlar. Nükleositoplazmik taşıma, önce çeviriden önce pre-mRNA'nın olgunlaşmasında önemli bir rol oynayan ve snRNP'lerin taşınması yoluyla spliceozom kompleksi oluşumunu kolaylaştıran son derece çeşitli bir süreçtir.

Spliceozom, nükleer pre-mRNA'dan intronların çıkarılmasını kolaylaştıran büyük bir RNA-protein kompleksidir. Nükleositoplazmik makineye dahil olan proteinlerdeki mutasyonlar, spliceozom oluşumunu bozabilir, anormal mRNA kırpılmasına yol açabilir, bu sebeplerden ötürü kas distrofisi dahil çeşitli genetik bozuklukların nedeni olarak bildirilmiştir. Kas distrofisi, genellikle bebeklik veya erken çocukluk döneminde ortaya çıkan, ilerleyen kas zayıflığı ile karakterize edilen genetik ve fenotipik olarak heterojen bozuklukları içeren bir grup hastalıktır. Bugüne kadar, kas distrofisi 85'ten fazla genle ilişkilendirilmiş olmasına rağmen, vakaların %40'ında kesin bir moleküler teşhis bulunmamaktadır.

Grubumuz, kas distrofisi tanısı konmuş on beş ayrı aileden toplam on sekiz hastayı çalışmamıza dahil etti. Tüm hastalar kas zayıflığı ile birlikte, sporadik nörolojik bozukluklar ve/veya kataraktın eşlik ettiği semptomlar gösterdi. Yapılan tüm ekzom dizileme sonucunda, tüm hastalarda *SNUPN* geninde yeni homozigot veya bileşik heterozigot varyantlar ortaya çıktı. Önemli bir şekilde, bu genin daha önce kas defektleri ile ilişkilendirilen patojenik varyantları bulunmamaktadır. Bu verilerden yola çıkarak *SNUPN* mutasyonlarının, SPN1'in yapısal bozukluklara neden olduğu ve UsnRNP'lerin nükleositoplazmik taşımasını bozarak spliceozom defektlerine yol açtığı hipotezi geliştirildi.

İlginç bir şekilde, tüm SPN1 patojenik varyantları proteinin kötü tanımlanmış C-terminal bölgesinde lokalizedir. Bu bölgenin intra-moleküler etkileşimlere katıldığı bilinmektedir. Bu çalışmanın temel amacı, SPN1'in C-terminal bölgesinin fonksiyonunu daha fazla karakterize etmek ve nihayetinde proteinin C-terminal bölgesi aracılığıyla inter-moleküler etkileşim yeteneğini araştırmaktır.

Bu soruyu yanıtlamak için, önce *SNUPN* doğal tip (WT) ve mutant DNA konstrüksiyonlarını pCS2+ memeli anlatım vektöründe moleküler klonlama ve yönlendirilmiş mutajenez kullanarak oluşturduk. Ardından, biyokimyasal deneyler, WT ve SPN1'in mutantlarını HEK293T/HeLa hücrelerinde aşırı ifade ederek, düzeylerini, lokalizasyonlarını ve etkileşimlerini karşılaştırmak üzere gerçekleştirildi. Bu deneyler aracılığıyla, proteinin C-terminal bölgesinin önemini ve mutasyonların SPN1'in kendisini oligomerizasyonunda nasıl bir etki yarattığını anlamayı amaçladık. Patojenik *SNUPN* varyantlarının, SPN1'in C-terminal bölgesinin kendi kendine oligomerizasyonunu bozarak proteinin hücresel fonksiyonlarını değiştirebileceğini ve bu durumun UsnRNP'lerin nükleositoplazmik taşınmasını bozarak hastalardaki gözlemlenen fenotipe yol açabileceğini hipotez etmekteyiz. Genel olarak, bu çalışma SPN1 proteininin karakterize edilmemiş C-terminal bölgesinin temel rolünü vurgulamayı amaçlamaktadır.

AKNOWLEDGEMENTS

Pursuit of the knowledge paired with dedication is a passport to success. As I enter to new chapter of my life, it was the driving force behind my achievements and the cornerstone of my journey. However, I would never be able to finalize my work without the guidance and support of following people. Thus, I would like to give my most sincere regards and appreciation to them.

I would like to express my profound thanks and regards that will last forever to my advisor, mentor, and role model Dr. Nathalie Escande Beillard. Rarely do kindness and brilliance converge, yet in Dr. Beillard's case, they are not only harmoniously combined but also complemented by genius. Her guidance and unwavering support throughout my journey have enabled me to overcome my weaknesses. I have gained invaluable lessons through the kindness and brilliant supervision of Dr. Nathalie. Her guidance has not only enhanced my scientific knowledge but also shaped my personal character, revealing the person I aspire to become in the future. I would consider myself fortunate if I can embody the qualities, she has instilled in me. I will be forever grateful to her for having faith in me and giving me the opportunity to join her amazing team.

I also would like to thank all my thesis jury members; Prof. Tuğba Bağcı Önder and Prof. Arzu Çelik Fuss for their critical evaluation and valuable discussions of this work.

I am grateful to all the members of the N.E.B. lab. The experience of working, learning, and sharing laughter together with them felt like a privilege.

A special acknowledgement and heartfelt regards go to Marwan Ahmad Nashabat for his continuous support, help and great advices at bench-side as well as daily life. He was a big brother to me rather than a colleague. I would like to thank him for always being there for me whenever I needed it.

I would like to extend my heartfelt gratitude to Dr. Nasrinsadat Nabavizadeh for her unwavering guidance, support, and for making possible one of the most significant scientific opportunities to enhance myself. I hope that in life, she will encounter individuals as kind as she is, who will make everything easier for her.

I want to express my gratitude to all my lab mates and friends; Hilal Pırıl Saraçoğlu, Seyide Ecesu Uygur, and Mert Kaya, Ece Çepni, Beyza Yavuzcan for sharing genuine laughter and offering support both at the bench side, throughout our daily life and in our academic journey. Their companionship has been invaluable.

Throughout my academic journey, I had the privilege of meeting amazing friends who supported me to a great extent and helped me realize the path I would like to move forward on. For their beautiful friendship and our genuine laughter together, I would like to thank İpek Yılmaz Düzgün, Ariana Sivaslıoğlu, Gizem Söyler and Mehmet Eren Çiftçi.

"We are the scientist we must help each other." One of the greatest quotes that I heard from one of the greatest scientists that I have ever met and a moto that I aspire to live by. Prof. Bruno Reversade is a remarkable scientist that leads by example, in terms of scientific inquisitiveness, research ethics, leadership and most of all humility, generosity and kindness. Being guided by Dr. Bruno is an opportunity that I will forever be grateful for as it was an experience that not only provided me with diverse scientific skills but also greatly aided in my personal development. His unique perspective on all things science and otherwise push me to think beyond limitations, dream big, and to most of all remain humble throughout this journey. I hope to one day be a person that's as kind, wise and generous as Dr. Bruno is. Thank you for believing in me and giving me the chance to be a part of your brilliant team.

I would like to extend a special acknowledgment, my heartfelt thanks, and sincere regards to Dr. Emmanuelle Szenker-Ravi, a great scientist, mentor, helpful, and a good friend who spreads laughter wherever she goes. As a candidate aspiring to be a scientist, you would understand that a good training and a great support is strongly correlated with your daily morale and success. Under her training, I believe I experienced best of both. Also, I would like to express all my thanks to current and former Reversade lab members.

I wish to extend my heartfelt gratitude to Prof. Hülya Kayserili for her exceptional guidance, which has enabled one of the most significant scientific opportunities in my academic journey. I am forever indebted to her for her trust and support. Additionally, I would like to express thanks to the entire Koç University 6th floor Medical Genetics team, particularly Dr. Umut Altunoğlu and Gül Leyla Kılıç

With endless and unconditional support people can achieve everything in life if they work enough for it. This is why I would like to dedicate my last words to express my heartfelt gratitude and forever lasting respect to my family who supported me through my failures and success. They were blessings along my journey, and the qualities they instilled in me have shaped the person I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CADD	Combined Annotation Dependent Depletion
CB	Cajal Body
CBC	Heterodimeric Cap Binding Protein
CDS	Coding Sequence
CK	Creatine Kinase
CRM1	Exportin 1 Protein
DANN	Deleterious Annotation of Genetic Variants Using Neural networks
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EMG	Electromyography
HRP	Horseradish peroxidase
IBB	Importin Beta Binding
m3G	Trimethylguanosine
m7G	7-methylguanosine
MCAP	Mendelian Clinically Applicable Pathogenicity
PBS	Phosphate-buffered saline
PHAX	Phosphorylated Adaptor for RNA Export Protein
PVDF	Polyvinylidene Difluoride
SIFT	Sorting Intolerant from Tolerant
SMN1	Survival of Motor Neuron 1 Protein
snRNA	small-nuclear RNA
snRNP	small nuclear ribonucleoproteins
SPN1	Snurportin1 Protein
TGS1	Trimethylguanosine Synthase 1
WT	Wild-Type
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside

Chapter 1:
INTRODUCTION

1.1 Snurportin 1 Protein

Majority of the eukaryotic organisms share the ability to transport RNA and protein molecules between different compartments of the cell such as the cytoplasm and nucleus (Ding & Sepehrimanesh, 2021; Görlich & Mattaj, 1996). Nuclear translocation of substrates occurs via factors that aid in the recognition of substrates and mediate their transport via karyopherins (Yang et al., 2023). The usage of adaptors enhances the variety of the cargoes that can be imported into the nucleus (Yang et al., 2023).

SNUPN gene located on chromosome 15 (15q24.2), encodes the Snurportin 1 (SPN1) protein which is an essential nuclear import adaptor of the spliceosomal small nuclear ribonucleoproteins (snRNPs) (Huber et al., 1998). snRNPs are RNA-protein complexes found in the spliceosome which facilitate the removal of the introns from pre-mRNA for their maturation (Gruss et al., 2017). The structure of the SPN1 protein is characterized by three functional domains: an N-terminal importin-β1 binding domain (IBB), a core trimethylguanosine (m3G-cap) binding domain, and a poorly defined C-terminal region that is reported to be responsible for interacting with Exportin 1 (CRM1) (Ospina et al., 2005) (Figure 1.1). Importin-β1 is a part of the Importin-β family of karyopherins that functions in the nuclear transport of proteins from the cytoplasm to the nucleus (Luo et al., 2013). The IBB domain (residues 1-65) plays a crucial role in facilitating the nuclear import of the SPN1 protein in RanGTP independent manner (Mitrousis et al., 2008; Ospina et al., 2005). This domain is highly conserved across higher eukaryotic species. Thus, mutations that cause structural impairments in the IBB domain of the SPN1 decrease the proteins` nuclear import function (Ospina et al., 2005).

UsnRNP molecules contain a 2,2,7-trimethyl-guanosine cap (m3G-cap) in their structure which allows for their specific recognition by SPN1 (Fischer et al., 2011). The core m3G-cap binding domain (residues 162-300) of the SPN1 is highly specific for the recognition of the m3G-cap (Piecyk et al., 2015; Strasser et al., 2005). Moreover, this

domain has not been reported to share any sequence similarities with any other protein (Strasser et al., 2005). m3G-cap binding pocket formed by this domain harbors two significant tryptophan residues Trp107 and Trp276 which facilitate the specific recognition of the m3G-cap of the UsnRNPs`. Furthermore, mutations on these residues are reported to reduce the recognition and binding of the SPN1 to its ligand (Strasser et al., 2005).

To this day, the C-terminal region of the SPN1 (residues 301-360) remains uncharacterized. Also, attempts to obtain the crystalized complete structure of SPN1 including the C-terminal, either alone or in complex with m3G-cap oligo and/or importin-β1 have been unsuccessful (Strasser et al., 2004). This highlights the intrinsically disordered nature of the C-terminal region of the protein. Moreover, *in silico* structural prediction of the SPN1 protein suggests that the C-terminal region of the protein harbors an alpha-helix between the residues Gly303 and Glu322 and is mainly composed of non-polar residues (Nashabat et al., 2024) (Figure 1.2).

Alpha-helices are the most prevalent structural elements in proteins and are crucial in determining their overall structure and functionality (Haimov & Srebnik, 2016). An alpha helix is a type of secondary structural element that resembles a rod-like structure with side chains that radiate outward in a helical array. Its central section is made up of a tightly coiled main chain (Sefidkar et al., 2022). It has been well established that interactions between hydrophobic side chains inside an a-helix, in addition to each residue's inherent helical tendency, can impact the stability of the structure (Creamer & Rose, 1995). Moreover, due to their inter- and intramolecular interactions via hydrogen bonds, alpha helices were suggested to play essential roles in oligomerization, assembly, and functioning of the protein through coiled-coil interactions (Liu et al., 2006; Zhang et al., 2015).

Notably, intramolecular interactions between the N- and C-terminus regions of the SPN1 have previously been reported by Matera *et al.*, (2005). Their findings proposed that loss of the C-terminal region of the protein increased the affinity of the SPN1 to importin-β1. Furthermore, mutations on either region of the protein have the probability

of disrupting the interaction between the two regions, consequently, resulting in a decrease in the efficacy of nuclear import (Ospina et al., 2005).

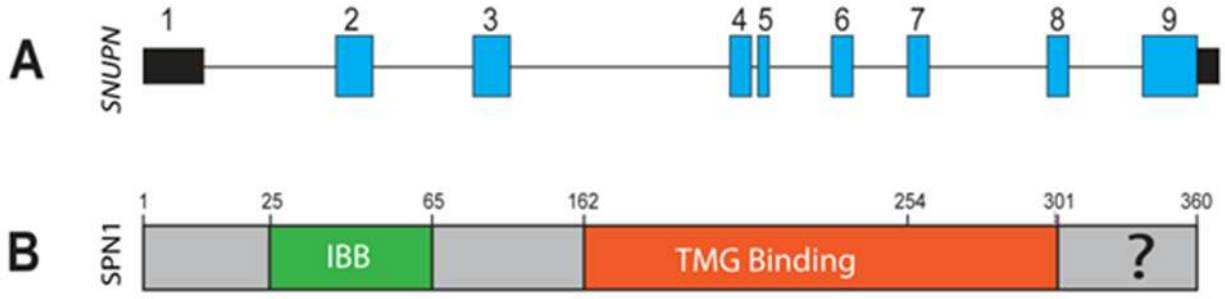

Figure 1.1: The diagram shows the SNUPN gene and SPN1 protein. A) Diagram illustrates the nine exons that make up the Homo sapiens *SNUPN* (MIM607902). B) SPN1 protein, consist of two conserved domains; Importinβ-1 binding domain (IBB) (1-25) (green) and trimethylguanosine (m3G)-cap-binding domain (TMG) (162-165) (bright orange), and uncharacterized C-terminal region (301-360) indicated with question mark.

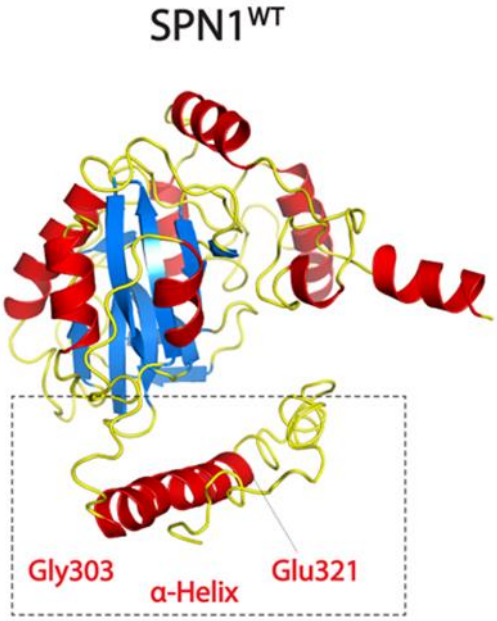

Figure 1.2: In silico predictions of wild-type (WT) SPN1 protein. Predicted 3D structure of the wild-type SPN1 protein, C-terminus highlighted within a dashed border. Colors of the complete protein structure are based on the secondary structures with alpha helices in red, beta sheets in blue, and loops in yellow (*adapted from* (Nashabat et al., 2024)).

1.2 Spliceosome and Biogenesis of Small Nuclear Ribonucleoproteins

RNA splicing is an essential process that occurs post-transcription for most of the eukaryotic genes. The sections of eukaryotic genes that code for proteins are frequently interrupted by non-coding "introns". Thus, to form mature messenger RNAs (mRNAs), these introns must be removed from the premature mRNA (pre-mRNA), and the protein-coding portions termed "exons" must be connected. pre-mRNA splicing is a critical stage in eukaryotic gene expression (Wilkinson et al., 2020). The spliceosome is a small-nuclear RNA (snRNA)-protein complex that functions in the removal of the introns from the pre-mRNA (Will & Lührmann, 2011). Biogenesis of the snRNP's starts with the transcription of the U1, U2, U4, U5, and relatively less abundant U11, U12, and U4atac snRNAs in the nucleus via RNA polymerase II (RNA Pol II) (Narayanan et al., 2002).

These non-coding and non-polyadenylated transcripts acquire m7G-cap transcriptionally and they are exported to the cytoplasm by a complex composed of Phosphorylated adapter RNA export protein (PHAX), The heterodimeric cap-binding complex (CBC), Exportin 1 (CRM1) in Ran-GTP dependent manner following their transcription (Fischer et al., 2011; Matera et al., 2007)

In the cytoplasm, dissociation of the snRNA export complex serves as a signal for the snRNA maturation and formation of functional snRNP's (Fischer et al., 2011). In the cytoplasm, the formation of the stable snRNPs is orchestrated via the survival motor neuron (SMN) protein complex (Raimer et al., 2017). Following the assembly of the Sm-core domain (B/B', D1, D2, D3, E, F, and G) which is crucial for the stability of the snRNP, snRNAs undergo two major modifications; trimming of the 3' extensions, and hypermethylation of the m7G-cap to m3G-cap (TMG) (Fischer et al., 2011; Urlaub et al., 2001). Hypermethylation of the m7G-cap to m3G-cap is carried out by the Tgs1 protein and serves as a nuclear import signal (Narayanan et al., 2002; Raimer et al., 2017). Nuclear import of the snRNPs is facilitated by SPN1 protein which binds specifically to the m3G cap of the snRNPs and importin-β1 to form nucleocytoplasmic import complex (Lott & Cingolani, 2011; Narayanan et al., 2002; Ospina et al., 2005).

Following the assembly of the nucleocytoplasmic import complex, snRNPs are imported to the nucleus to be localized in Cajal bodies (CBs) (Fischer et al., 2011; Karijolich & Yu, 2010). CBs are membraneless nuclear bodies which are known to play significant roles in RNA metabolism, biogenesis of ribosomes, maintenance of the telomeres, splicing, and maturation of the snRNPs (Sawyer et al., 2016). Within the CBs, firstly, nucleocytoplasmic import complex dissociates, and specific ligands are then recruited to facilitate the maturation of respective snRNPs (Fischer et al., 2011; Sleeman & Lamond, 1999). Additionally, snRNPs also undergo post-translational modifications, such as pseudouridylation and methylation of specific nucleotides in these nuclear bodies for their maturation, and mature snRNPs form the spliceosome complex for the removal of the introns from pre-mRNA (Kiss, 2004; Machyna et al., 2013).

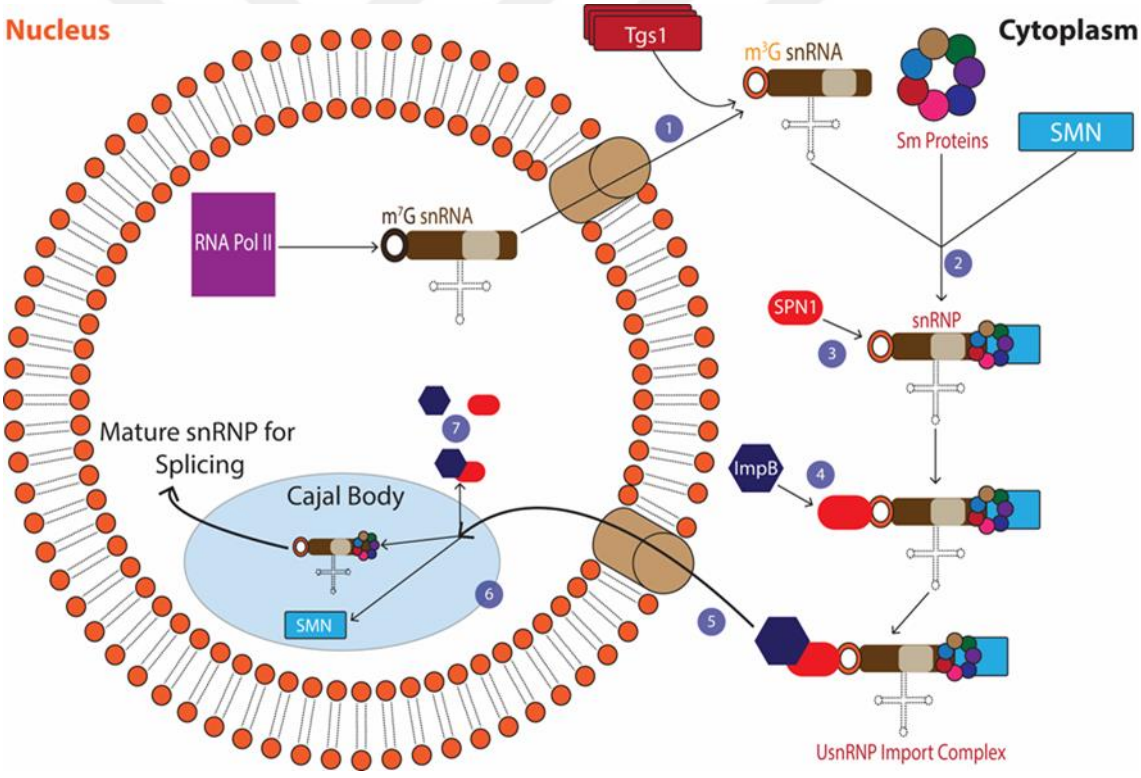

Figure 1.3: Schematic representation of biogenesis of the UsnRNP's. snRNAs are encoded in the nucleus by RNA polymerase II (RNA Pol II, in purple). (1) Upon their transport to the nucleus, the m^7G cap of the snRNAs is hypermethylated by Tgs1 to m3G-cap. (2) In cytoplasm m3G-capped snRNAs, Sm proteins, and the SMN complex get together to form snRNPs. (3) The nuclear import adaptor SPN1 protein specifically recognizes and binds to the m3G-cap of snRNP's. (4) Importinβ-1 (Impβ-1) binds to the SPN1, and (5) transports the UsnRNP complex into the nucleus. (6) In the nucleus, complex dissociates, and UsnRNP's are localized in Cajal bodies where they mature to form spliceosomes.

1.3 Spliceosomal Defects and Muscular Dystrophies

RNA splicing and alternative splicing events are among the key features of the human transcriptome, as more than 90% of protein-coding genes in humans have multiple isoforms (Pistoni et al., 2010). Improper splicing has been well established as one of the most prevalent causes of hereditary diseases, which highlights the significance of proper maturation of snRNPs for the formation of the splicing machinery (López-Bigas et al., 2005; Pistoni et al., 2010; Scotti & Swanson, 2016; Wang & Cooper, 2007). Various diseases, such as cancer, retinitis pigmentosa, and numerous muscular dystrophies, have been associated with splicing mutations and splicing factors including U2 snRNP, U4 snRNP, and SMN gene (Love et al., 2023).

Muscular dystrophies (MDs) constitute genetically and phenotypically heterogeneous groups of inherited disorders that are characterized by progressive muscular weakness and wasting (Mercuri & Muntoni, 2013). The pathology of the MDs encompasses the degeneration, necrosis of the muscular fibers, and replacement of the affected tissue with fatty and fibrous tissues. These disorders exhibit a broad range in terms of severity, pattern of the affected muscles, and the age of onset (Emery, 2002). Moreover, MDs are one of the most prevalent among hereditary diseases, with 40% of the cases lacking a definitive molecular diagnosis.

Impairments in the function of pre-messenger RNA splicing regulators, and spliceosomal components could lead to aberrant splicing of several pre-mRNAs (Pistoni et al., 2010). Diseases like FSHD or DM1 have already been associated with the development of muscular diseases (Pistoni et al., 2010). Significantly, mutations in the survival of the motor neuron 1 (SMN1) gene, which has an essential role in the biogenesis of snRNPs lead to spinal muscular atrophy (SMA1) (Faustino & Cooper, 2003; Pistoni et al., 2010).

1.4 Snurportin 1 Protein and Muscular Dystrophies

Identification and characterization of the function of essential proteins in RNA metabolism provide valuable insights for understanding the cause and progression of muscular dystrophies. SPN1 protein plays a vital role in the assembly of the splicing machinery. Despite its integral function, the detailed understanding of the protein's endogenous role is still unknown and its association with human disease has never been reported.

Moreover, the intrinsically disordered nature of its uncharacterized C-terminus region poses challenges in investigating SPN1 function and interactions (Ospina et al., 2005; Strasser et al., 2005). SPN1's central function in the pathway of biogenesis of the snRNPs, along with its interaction with major proteins such as SMN, suggests that defects in SPN1 might play a significant role in the physiopathology of muscular dystrophies.

Notably, our team identified a total of nine unique recessive pathogenic mutations in the SNUPN gene in eighteen families who present progressive muscular weakness (Nashabat et al., 2024). The majority of these mutations are localized in the uncharacterized C-terminal region, indicating a mutational hotspot. Eighteen affected individuals display a range of muscular weakness, in relation to the severity of the mutations which highlights the importance of the C-terminal domain.

In summary, the uncharacterized C-terminus of the SPN1 is mainly composed of non-polar residues which are predicted to harbor an alpha-helix. Furthermore, C-terminus is reported to be capable of intermolecular interactions (Nashabat et al., 2024; Ospina et al., 2005). This thesis aims to investigate the role of the uncharacterized C-terminal domain in self-interaction in the context of a novel muscular dystrophy, by generating transient overexpression constructs of the wild-type and homozygous pathogenic mutants that impair or disrupt the C-terminus of the SPN1.

Chapter 2:
MATERIAL AND METHODS

2.1 Cloning PCR

To generate the overexpression constructs the whole open reading frame (ORF) of human *SNUPN* and the cDNA fragments containing Δ 254-360 of the *SNUPN* gene were amplified by iProof High-Fidelity PCR kit (Biorad, catalog # 1725330) using complement primer sequences with additional Kozak consensus sequence on N-terminal, and BamHI and XhoI restriction enzyme sites on N-terminal and C-terminal respectively. To distinguish the overexpressed proteins from the endogenous proteins, constructs were tagged with either the Flag epitope tag or the Myc epitope tag. Tag sequences were integrated at the N-terminal of the protein after the start codon. For generating wild-type and mutant *SNUPN* Δ254-360 fragment constructs (m1, m6, and m7) *SNUPN* wild-type and mutant constructs were used as a template. Amplification reactions were prepared with various combinations of primers to create constructs with desired tags. Primers are listed in Table 2.1. Reaction setup and thermal cycling setup are described in Table 2.2 and Table 2.3 respectively. Subsequently, 5 µL from the PCR samples were mixed with the 1 µL of 6X Gel Loading Dye (New England Biolabs, catalog # B7024A) and then loaded in precast 1% Agarose gel, prepared with SyberTMSafe Stain (1:10000, Invitrogen, catalog #S33102) in 1X TAE buffer. Samples were run under 120V for 60 minutes and visualization was performed with Gel Doc XR+ Molecular Imager (BIORAD, catalog # 1708195EDU).

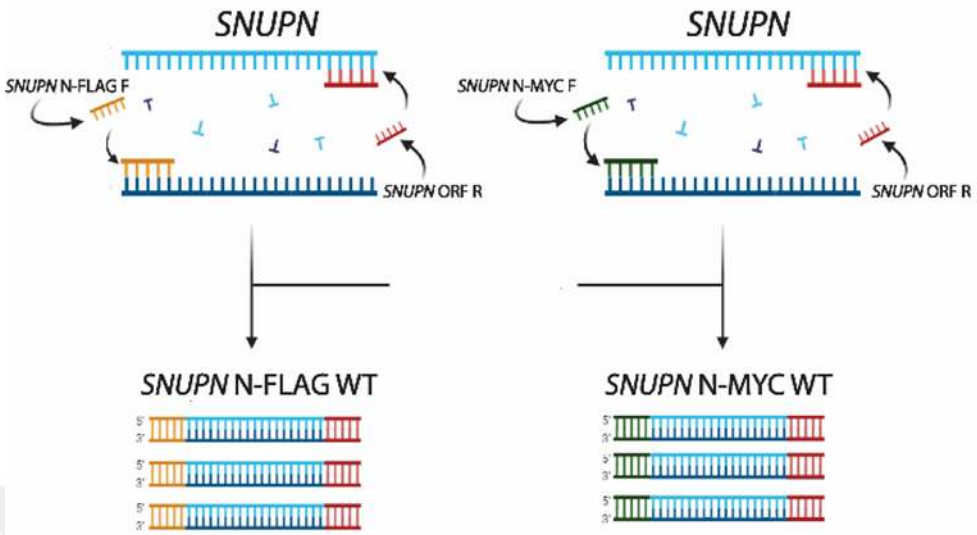

Figure 2.1: Schematic diagram* illustrating *the cloning PCR reaction. Generation of the Flag (orange) and Myc (green) epitope-tagged versions of the *SNUPN* WT amplicons.

Table 2.1: List of Primers used in Cloning Experiment

Gene	Label	Strand	Sequence (5'-3')
SNUPN	*SNUPN*-ORF-F	Forward	ATCGGATCCGGGAAGATGGAAGAGTTGAGTAGGCCCT
SNUPN	*SNUPN*-ORF-R	Reverse	TAGCTCGAGTTAATTCTCCATGAGGCATCCAGG
SNUPN	*SNUPN* ORF N-Flag-F	Forward	ATCGGATCCGGGAAGATGGACTACAAGGACGACGAT GACAAGGAAGAGTTGAGTCAGGCCCT
SNUPN	*SNUPN* ORF N-Myc-F	Forward	ATCGGATCCGGGAAGATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGGAAGAGTTGAGTCAGGCCCT
SNUPN	*SNUPN*Δ254-360 Myc-F	Forward	ATCGGATCCATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGGTAGATGGACTTCTCTTCTACCACA
SNUPN	*SNUPN*Δ254-360 Flag-F	Forward	ATCGGATCCATGGACTACAAGGACGACGATGACAAGGTAGATGGACTTCTCTTCTACCACA

Table 2.2: PCR Reaction Setup

Component	**Volume**	**Final Concentration**
5X IProof HF Buffer	4 µL	1X
dNTP mix	0,4 µL	200 µM each
Forward Primer	1 µL	0.5 µM
Reverse Primer	1 µL	0.5 µM
DNA template	1 µL	1 ng
iProof DNA polymerase	0.5 µL	0.05 U/ µL
ddH20	12,1 µL	

Table 2.3: Thermal Cycling Protocol

Step	Temperature	Duration	Cycle
Initial Denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	10
Annealing	55°C	30 seconds	
Extension	72°C	30 seconds	
Denaturation	98 °C	10 seconds	25
Annealing	63°C	30 seconds	
Extension	72°C	30 seconds	
Final Extension	72°C	10 minutes	1
Hold	4°C	∞	

2.2 Gel Extraction

The identified bands were extracted from the agarose gel by using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, catalog # 740609.50) according to the manufacturer's instructions for Gel Extraction from agarose gels.

2.3 Polyadenylation of Blunt-ended PCR Fragment

iProof™ High-Fidelity DNA Polymerase generates blunt-ended PCR fragments. To facilitate TA cloning of the constructs into the pGEMT-Easy vector, dATP (New England Biolabs, catalog #N0440S) was added at the 3' end of the purified PCR products via Taq DNA Polymerase with ThermoPol Buffer (New England Biolabs, catalog #M0267L), reaction setup can be visualized in Table 2.4. After incubating the reaction mixture for 90 minutes, PCR samples were purified by using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, catalog #740609.50) according to the manufacturer's instructions for PCR purification.

Table 2.4: Polyadenylation Reaction Setup

Component	Volume	Final Concentration
10X ThermoPol Buffer	5 µL	1X
Taq DNA polymerase	1 µL	5 Units/50 µL
dATP	1 µL	2 mM
PCR product	5 µL (500 ng)	500 ng/50 µL
Water	To 50 µL	

2.4 pGEM-T Easy Vector Ligation

Polyadenylated constructs were cloned into pGEM®-T Easy (Promega, catalog #A1360) cloning vector kit. The ligation reaction was prepared according to the manufacturer's instructions with a 1:3 vector to insert ratio; the reaction setup can be visualized in Table 2.5. Samples were incubated at room temperature for 1 hour, and later on, samples were transformed into DH5α chemically competent cells, and the remaining reaction was stored at -20°C.

Table 2.5: pGEM-T Easy Ligation Reaction Setup

Component	Volume	Final Concentration
2X Rapid Ligation Buffer, T4 DNA Ligase	5 µL	1X
pGEM®-T Easy Vector	1 µL	50 ng/µL
PCR product	2 µL	55 ng/µL
T4 DNA Ligase	5 µL	3 Weiss units/µl
Water	To 10 µL	

2.5 Transformation to Chemically Competent DH5α Cells

Before starting, the desired number of plates were placed in an incubator at 37 °C upside down and left to dry. 5 µL of ligation product was transformed into 50 µL of DH5α competent cells and mixed via gentle flicking a few times, then mixtures were incubated for 30 minutes on ice. Later on, the mixture heat shocked in a water bath exactly at 42 °C for 30 seconds and subsequently incubated on ice for 2 minutes. 950 µL of LB medium were added to transformed cells and mixtures were incubated for 1 hour at 37°C with 225 rpm shaking. Following the incubation, samples were centrifuged for 2 minutes at 7000 rpm and 900 µL of the supernatant was aspirated, and the cell pellet was resuspended in the remaining 100 µl. 100 µL of transformed cell culture was plated onto LB/Agar plates with ampicillin (100 ug/mL, GOLDBIO, catalog #A-301-25), and, 40 µL of 2% 5-Bromo-4-chloro-3-indolyl b-D-galactopyranoside (X-Gal) (Glentham Life Sciences, catalog # 7240-90-6) in dimethyl sulfoxide (DMSO). Each mixture was well spread on plates and plates were placed into the incubator at 37°C, overnight.

2.6 Small Scale Plasmid DNA Isolation

A single colony was inoculated from the plate and grown overnight at 37°C on a shaker in 5 mL of LB media supplemented with ampicillin (100 ug/mL) for selection. On the next day, 2 mL from the liquid culture was centrifuged at 11.000 rcf for 30 seconds to pellet the cells. Plasmid DNA was isolated using the NucleoSpin Plasmid Kit (MACHERY-NAGEL, catalog #740588.50) according to the manufacturer's protocol, and stored at -20°C.

2.7 Cycle Sequencing of the Constructs

The DNA sequence of the constructs was checked with the use of BigDye Terminator v3.1 Cycle Sequencing Kit (catalog #4336917) according to the manufacturer's instructions with the use of primer sequences of SP6 and T7 promoter region-specific primers (Table 2.6), reaction and thermal cycling setups can be visualized in Table 2.7 and Table 2.8.

Table 2.6: List of Sequencing Primers

Label	**Sequence (5'-3')**
SP6 Sequencing Primer	ATTTAGGTGACACTATAG
T7 Sequencing Primer	TAGCTCGAGTTAATTCTCCATGAGGCATCCAGG

Table 2.7: BigDye Terminator v3.1 Cycle Sequencing Reaction Setup

Component	Volume	Final Concentration
BigDye™ Terminator 3.1 Ready Reaction Mix	1 µL	
BigDye™ Terminator v1.1 & v3.1 5X Sequencing Buffer	2 µL	1X
SP6 Sequencing Primer	0.5 µL	0.5 µM
T7 Sequencing Primer	0.5 µL	0.5 µM
Template	1 µL	200 ng/ µL
Water	To 10 µL	

Table 2.8: BigDye Terminator v3.1 Cycle Sequencing Thermal Cycling Setup

Step	Temperature	Duration	Cycle
Initial Denaturation	95°C	3 minutes	1
Denaturation	95°C	30 seconds	35
Annealing	50°C	15 seconds	
Extension	60°C	30 seconds	
Hold	4°C	∞	1

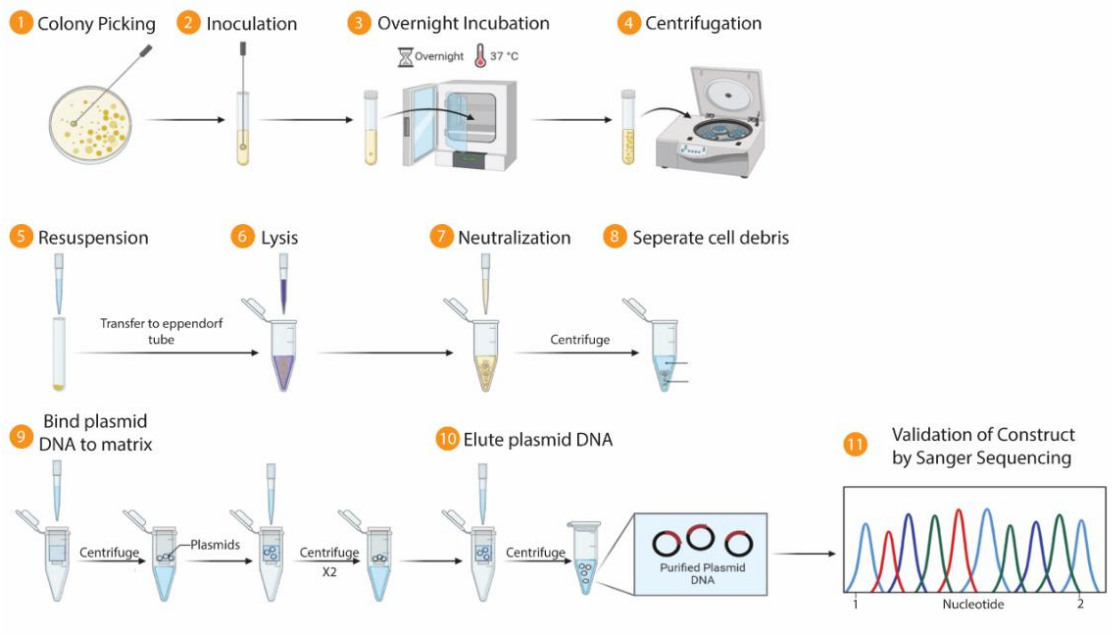

Figure 2.2: Illustration of isolation and validation of the SNUPN plasmid DNA constructs. Schematic illustrates the steps performed for isolation of the plasmid constructs (1-10), and the technique of validation (11) after obtention of the purified plasmid DNA. Schematic is illustrated by biorender.

2.8 Restriction Digestion

Restriction digestion was performed to isolate the insert from the pGEM® -T Easy vector and allow proper matching between the inserts and the pCS2+ vector during ligation. pCS2+ vector and constructs were digested, with BamHI (New England Biolabs, catalog #R0136S) and XhoI enzymes (New England Biolabs, catalog #R0146S). The reaction setup can be visualized in Table 2.9. Samples were incubated at 37°C for 20 minutes. After the digestion, 5 µL from the samples migrated under 120V for 60 minutes in precast 1% Agarose gel, prepared with Syber™Safe Stain (1:10000, Invitrogen, catalog

#S33102) in 1X TAE buffer. Visualization was performed with Gel Doc XR+ Molecular Imager (BIORAD, catalog # 1708195EDU). Inserts isolated via gel extraction and the pCS2+ vector digestion reaction were purified with NucleoSpin Gel and PCR Clean up kit (MACHEREY-NAGEL, catalog # 740609.50) according to the manufacturer's protocol.

Table 2.9: Restriction Digestion Reaction Setup

Component	Volume	Final Concentration
10X rCutSmart™ Buffer	5 µL	1X
BamHI	1 µL	20 units/ 50 µL
XhoI	1 µL	20 units/ 50 µL
DNA	1 µg	20 ng/ µL
Water	To 50 µL	

2.9 Subsequent Cloning into pCS2+ Mammalian Expression Vector

Digested *SNUPN* constructs were cloned into the digested pCS2+ expression vector via the Quick Ligation™ kit (New England Biolabs, catalog #M2200). The ligation reaction contains 100 ng vector DNA, 80 ng insert (according to 1:3, vector: insert ratio), 2X Quick Ligase Buffer, and Quick Ligase Enzyme, to a final reaction volume of 20 µl, reaction setup can be visualized in Table 2.10. The ligation reaction was incubated at room temperature for 10 minutes, and then it was used for transformation in DH5α chemically competent cells as described earlier. Following the transformation, plasmid DNA was isolated from positive clones via the small-scale plasmid isolation and extraction method described earlier. Next DNA sequence of the plasmids was validated via Cycle Sequencing.

Table 2.10: Quick Ligation Reaction Setup

Component	**Volume**
Vector	2 µL (100ng)
Insert	4 µL (80 ng)
Quick Ligase	1 µL
Quick Ligase Reaction Buffer 2X	2 µL
Water	To 20 µL

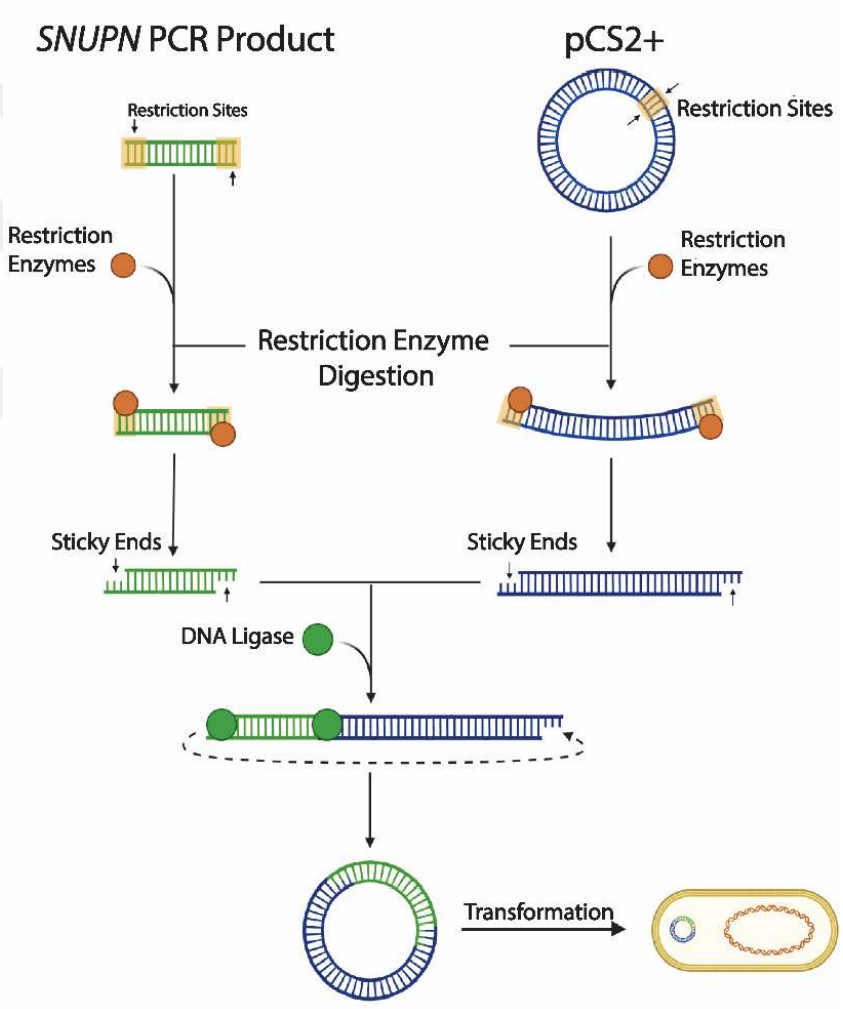

Figure 2.3: Schematic of molecular cloning. Integration of the *SNUPN* WT inserts into pCS2+ mammalian expression vector, followed by their transformation into chemically competent DH5-α bacteria. Schematic is illustrated by biorender.

2.10 Large Scale Plasmid DNA Isolation and Extraction

5 nanogram (ng) from the validated construct is transformed into DH5α cells as described earlier. A single colony was inoculated from the plate and grown for 8 hours at 37°C on a shaker in 5 mL of LB media supplemented with ampicillin (100 ug/mL). 1 mL from the liquid culture was transferred to the 200 mL LB media supplemented with ampicillin (100 ug/mL). On the next day, bacteria cells were harvested via centrifugation at 4.000 for 15 minutes at 4°C. Plasmid DNA was isolated by using NucleoBond Xtra Maxi Plus kit (MACHERY-NAGEL, catalog #740416.50) according to the manufacturer's protocol, and stored at -20°C.

2.11 Site-Directed Mutagenesis

Site-directed mutagenesis was used to substitute or delete specific nucleotides to create patient mutations. The *SNUPN* wild-type construct was used as a template to generate mutant constructs by Quick-Change II XL Site-Directed Mutagenesis kit (Agilent, catalog #200522). All primer sequences used for site-directed mutagenesis are shown in Table 2.11, reaction and thermal cycling setup can be visualized in Table 2.12. cycling parameters can be visualized in Table 2.13. Subsequently, after the thermal cycling reaction was incubated on ice for 2 minutes and the PCR reaction was digested with 1 µl of Dpn I (10 U/µl) for 1 hour at 37°C for the digestion of the parental DNA. Following, 45 µl of XL1-Blue Supercompetent Cells provided in the kit thawed on ice and 2 µl of the β-ME mix provided with the kit was added to the cells, and the mixture was incubated on ice for 10 minutes. 2 µl of the Dpn I-treated DNA was transferred to the mixture and reaction was incubated on ice for 30 minutes, subjected to heat-pulse for 30 seconds in a water bath at 42°C, and incubated on ice for 2 minutes subsequently. 500 µL of pre-heated LB broth was added later and the mixture was incubated for 1 hour at 37°C with 225 rpm shaking. After the incubation, 250 µL of the mixture was plated onto LB/Agar plates with ampicillin (100 ug/mL), and the plate was incubated at 37°C overnight. Isolation, extraction, and validation of the plasmid DNA were carried out with the protocols described earlier.

Table 2.11: The List of Primers Used in Site-Directed Mutagenesis Experiment

Gene	Label	Strand	Sequence (5'-3')
SNUPN	*SNUPN*-M1 (926 T>G) F	Forward	CACCAGCTCCAGCAGAGTATGGAGCACAAGAAG
SNUPN	*SNUPN*-M1 (926 T>G) R	Reverse	CTTCTTGTGCTCCATACTCTGCTGGAGCTGGTG
SNUPN	*SNUPN*-M6 (902-903DEL) F	Forward	GACCACCAAGCCAGACTGCTGGGCAC
SNUPN	*SNUPN*-M6 (902-903DEL) R	Forward	GTGCCCAGCAGTCTGGCTTGGTGGTC
SNUPN	*SNUPN*-M7 (899-900DEL) F	Reverse	GACCACCAAGCCAGTATGCTGGGCACCA
SNUPN	*SNUPN*-M7 (899-900DEL) R	Forward	TGGTGCCCAGCATACTGGCTTGGTGGTC

Table 2.12: Site-Directed Mutagenesis Reaction Setup

Component	Volume	Final Concentration
10X Reaction Buffer	5 µl	1X
dsDNA template	1 µL	10 ng/50 µL
Forward primer	125 ng	125 ng/50 µL
Reverse primer	125 ng	125 ng/50 µL
dNTP mix	1 µL	
QuikSolution	3 µL	
PfuUltra HF DNA polymerase	1 µL	2.5 U/µl
Water	To 50 µL	

Table 2.13: Site-Directed Mutagenesis Thermal Cycling Setup

Step	Temperature	Duration	Cycle
Initial Denaturation	95 °C	1 minute	1
Denaturation	95 °C	50 seconds	18
Annealing	60°C	50 seconds	
Extension	68°C	5 minutes	
Final Extension	68°C	7 minutes	1
Hold	4°C	∞	1

2.12 Cell Culture and Transient Transfection

HEK293T cells (ATCC, catalog no. CRL-3216, RRID: CVCL_0063) were cultured in high glucose DMEM (Sigma, Catalog # D6429) supplemented with 10% fetal bovine serum (FBS) (Biowest, catalog #S1600-500) and 1% penicillin/streptomycin (Biowest, catalog #L0022-100) and they were maintained in a humidified 5% CO^2 atmosphere at 37°C. All the cell lines were also tested negative for Mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza, catalog #LT07-118). Transfection with the expression constructs was carried out with a Lipofectamine 2000 kit (Invitrogen, catalog #11668019) according to the manufacturer's protocol. Non-transfected respective cell lines which contain the same number of cells were cultured alongside the transfected cells for negative control.

2.13 Whole Protein Extraction from Cultured Cells

48 hours after transfection of the cells cell medium was removed and cells were treated with pre-warmed 4 ml 0.05% Trypsin-EDTA 0.53 mM (Wisent, catalog #325-542-EL) and incubated for 5 minutes at 37°C. Trypsin was deactivated by the addition of the 5 mL cell culture medium mentioned above, samples were transferred into a 15 mL falcon tube and centrifuged at 1500 rpm for 5 minutes. Supernatants were removed and the pellet was resuspended with 5 mL of filtered ice-cold PBS and centrifuged at 12.000 rpm for 5 minutes at 4°C. After the removal of the supernatant, total protein from the cells was extracted by using cold 350 µL of 1X RIPA lysis buffer (Millipore, catalog #20-188) supplemented with cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche, catalog #11873580001). Samples were incubated on ice for 30 minutes, and later on, centrifuged at 17.000 rcf for 30 minutes at 4°C. Each supernatant is transferred into a sterile low-protein binding centrifuge tube and stored at -20°C.

2.14 Immunoprecipitation

The whole cell lysate was isolated as described earlier and Anti-Flag M2 beads (Sigma-Aldrich, catalog #A2220) were subjected to one wash with PBS and 3 times with 1X RIPA lysis buffer. Next 20 µL aliquot of Anti-Flag M2 beads (Sigma-Aldrich, catalog #A2220) was collected via centrifugation at 6.000 rpm for 90 seconds at 4°C. Beads were resuspended in cell lysate and, the mixture was incubated on a rotator overnight at 4°C. On the following day, the mixture was washed 6 times with 1X RIPA lysis buffer. Elution of the proteins was carried out with the addition of 50 µL of 2X Laemmli loading buffer with 60 mM DTT, and incubation of the samples at 95°C for 10 minutes. In the final step, beads were removed with centrifugation at 6.000 rpm for 90 seconds at 4 °C, and the supernatant was collected into a sterile-low protein binding Eppendorf tube. Analysis of the samples was carried out by western blot and HRP conjugated anti-mouse IgG light chain specific secondary antibody (Abbkine, catalog #A25012) and HRP conjugated anti-rabbit IgG light chain specific secondary antibody (Abbkine, catalog # A25022) were used in the visualization.

2.15 Western Blotting

For western blotting, the samples were reduced by Laemmli sample buffer (Biorad, catalog #161-0747) containing dithiothreitol (DTT) and denatured for 5 min at 95°C. Then, the Precision Plus Protein Dual Color Standard (Bio-Rad, catalog # 1610374) and equal volumes of protein samples were loaded on 4–20 % Criterion™ TGX™ Precast Midi Protein Gels (Bio-Rad catalog # 5671093) for SDS-PAGE gel electrophoresis and run at 80 V for 2 hours. Following, proteins were transferred on an Immun-Blot® Low Fluorescence PVDF Membrane (Bio-Rad, catalog #1620261) using Trans-Blot® Turbo™ Transfer System (Bio-Rad, catalog #1704150). Therefore, Trans-Blot Turbo Midi-size Transfer Stacks (nitrocellulose membranes) were immersed in Trans-Blot Turbo 1X Transfer buffer (Bio-Rad, catalog #10026938), one nitrocellulose membrane was placed as the bottom layer inside of the cassette and, a Trans-Blot Turbo Midi-size PVDF membrane equilibrated with first submersion in 100% methanol and then 1X transfer buffer. PVDF membrane was placed on top of the nitrocellulose membrane and gel was placed on the PVDF membrane. The second wetted nitrocellulose membrane was placed on top of the gel to serve as the top ion reservoir stack. Potential air bubbles were

removed by gentle rolling with the blot roller. The transfer was performed by the administration of the 2.5A-25 V for 7 minutes. Later on, membranes were blocked with 5 % milk/TBST for 1 hour at room temperature and then probed with the primary antibodies diluted in 5 % milk/TBST overnight at 4 °C (Table 2.14). After 3 washes in TBST, membranes were incubated with the respective HRP conjugated anti-mouse (1:1000, Abbkine, catalog # A21010) and anti-rabbit (1:1000, Abbkine, catalog # A21020) secondary antibodies in 5 % milk/TBST. After 3 washes in TBST, the signal was revealed by the Pierce™ ECL Western Blotting Substrate (Thermo Fisher Scientific, catalog #32106) and the SuperSignal West Chemiluminescent Substrates (Thermo Fisher Scientific, catalog #34580/34096/A38554). In the end, membranes were developed by ChemiDoc MP Imaging System (Bio-Rad, catalog # 17001395).

Table 2.14: Primary Antibodies Used in Western Blot

Antibody	Host	Class	Assay and dilution	Company	Catalog #
Anti-SPN1	Rabbit	Polyclonal	Immunoblot: 1/1000	Proteintech	15358
Anti-GAPDH	Mouse	Monoclonal	Immunoblot: 1/500	Santa Cruz	47724
Anti-Myc	Mouse	Monoclonal	Immunoblot: 1/1000	Sigma	05-724
Anti-Flag	Rabbit	Polyclonal	Immunoblot: 1/1000	Cell Signaling	2368

Chapter 3:
RESULTS

3.1 Preliminary Clinical and Functional Data

3.1.1 Clinical and Molecular Analysis

The first patient (II:1), was a female from Kosovo born from a non-consanguineous family, her development was normal until the early stages of her childhood (Figure 3.1A). At the age of 5, she developed an abnormal gait with mildly anterior pelvic tilt, experienced frequent falls, and her symptoms worsened progressively over time. Clinical examinations of her limb muscles reflected a generalized muscle weakness (Figure 3.1A). Clinical observations were compatible with a diagnosis of muscular dystrophy (MD). However, routine molecular genetic examinations failed in the identification of pathogenic variants linked to any muscular dystrophy. Therefore, our team carried out a whole exome sequencing (WES) and uncovered a novel germline homozygous pathogenic mutation in *SNUPN* (*MIM607902*) (*m*1, Chr15: c.926T>G, p.lle309Ser). This homozygous variant had never been reported in public databases such as gnomAD, ExAC, or Exome Variant Server. Except for de novo mutation p.Asp255Glu affiliated with a disease yet of uncertain pathogenicity, there had not been an association of any homozygous *SNUPN* variant with a disorder (Cappi et al., 2016).

Segregation analysis on unaffected parents (F1-I:1 and F1-I:2) and brother (F1-II:2) confirmed revealed that this unique mutation segregated with the disease, which suggests an autosomal recessive inheritance pattern. Furthersmore, *in silico* analysis of pathogenicity of the *m*1 mutation with prediction tools such as CADD, DANN, MCAP, FATHMM-MKL, SIFT, polyPhen-2, and MutationTaster anticipated to be deleterious. Additionally, the mutated amino acid residue (Ile309) of SPN1 was highly conserved across vertebrate species, which contributes to the prediction of the possible pathogenicity (Figure 3.2C). Later on, 17 more cases from 14 unrelated families originating from Europe (F2), USA (F3), Switzerland (F4), Iraq (F5 and F12), Iran (F6 and F7), Macedonia (F8), Romania (F9), Guatemala (F10), another from Kosovo (F11), Columbia (F13 and F14), and Germany (F15) were identified through GeneMatcher (Sobreira et al., 2015) or

Centogene (Rostock, Germany) in-house database. All the cases were diagnosed with muscular dystrophy features. Routine molecular genetic examinations failed in the identification of the genetic cause in all patients of the cohort. Therefore, trio or single WES was carried out, which revealed recessive pathogenic mutations in the *SNUPN* gene that confirmed Mendelian recessive inheritance. Overall, our team identified four germline homozygous mutations which also consist of the one in patient F1 (*m*1: c.926T>G; p.lle309Ser (F9 and F11)) and three novel (*m*6: c.902_903delAT; p.Tyr301Cysfs*29 (F4, F13, F14); *m*7: c.899_900delAC; p.Asp300Valfs*30 (F5-F7, F12); *m*9: c.164G>A; p.Arg55Gln (F10) as well as four compound heterozygous mutations (*m*2: c.760-1G>A and *m*3: c.958delA; p.Met320ter (F2); *m*4: c.848C>G; p.Ser283ter; and *m*5: c.922C>T; p.Gln308ter (F3); *m*8: c.787C>T; p.Gln263ter and m1: c.926T>G; p.lle309Ser (F8); *m*1: c.926T>G; p.lle309Ser and *m*4: c.848C>G; p.Ser283ter (F15)) (Figure 3.1B).

In summary, eighteen affected individuals exhibit common muscular defects with varying degrees of severity (refer to Table 15). Overall, our results provide compelling clinical and genetic evidence that *SNUPN* mutations are causally related to a unique subtype of muscular dystrophy disorder (Nashabat et al., 2024).

Figure 3.1: Identification of patients with muscular dystrophy carrying biallelic SNUPN variants.

A) Images of the affected patients: Affected girl from family F1 (F1-II: 1:8 years old), exhibiting aberrant posture with left dissymmetry and having trouble raising her arms (Left). Affected individuals from family 4 (F4-II:2; 16 years old) (middle) and family 12 (F12-II:2; 8 years old) (right) are confined in a wheelchair and rely on artificial breathing apparatuses illustrating the profound severity of their disability. B) Pedigree`s of fifteen families showing autosomal recessive inheritance. Double lines denote a consanguineous marriage, symbols filled with black represent affected, and, crossed symbols indicate deceased individuals. The triangle symbol represents miscarriages. Compound heterozygous variations are displayed according to their paternal origin and *SNUPN* variant coordinates are supplied (Nashabat et al., 2024).

Table 3.1: Patient's clinical features associated SNUPN mutations

Clinical Features	Number of Patients Tested	Percentage
Muscular Weakness	18/18	100%
Elevated CK Levels	16/17	94%
Myopathy (EMG)	11/11	100%
Neurological Defects	7/9	78%
Bilateral Cataract	6/18	33%
Spine Malformation	12/16	75%
Resp. Insufficiency	11/18	61%

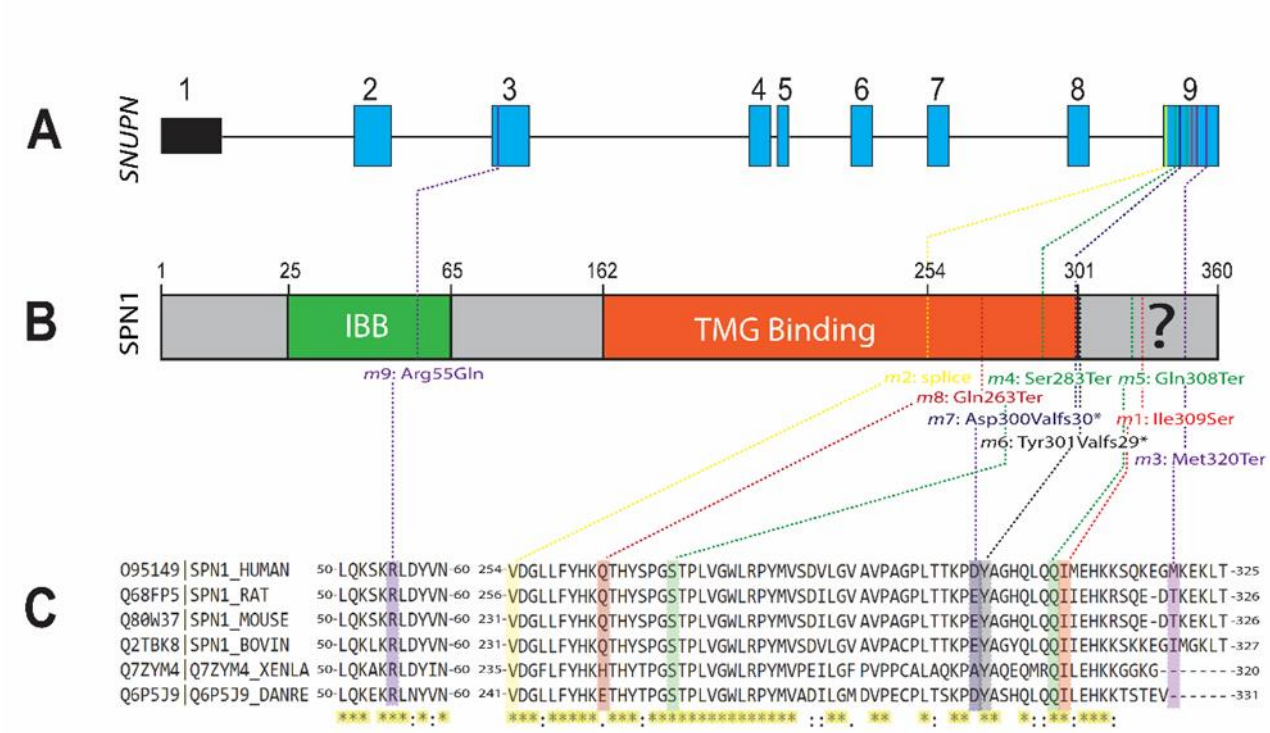

Figure 3.2: Schematic of SNUPN gene and SPN1 protein with identified mutations. A) Illustration of the genomic structure of Homo sapiens *SNUPN* (*MIM607902*) which consists of nine exons. B) Schematic diagram of the SPN1 protein composed of two conserved domains: Importinβ-1 binding domain (IBB) (green) and trimethylguanosine (m3G)-cap-binding domain (TMG) (bright orange). Nine pathogenic germline variants are depicted on both *SNUPN* genomic and SPN1 protein schematics. The majority of the pathogenic variants (*m*1-*m*8) are localized in the C-terminus region between residues 254-320, whereas solely the *m*9 pathogenic variant is found in the N-terminus region. The question mark highlights the uncharacterized C-terminal domain (Nashabat et al., 2024).

3.1.2 Analysis of SNUPN and SPN1 Levels in Patient Fibroblast

To begin investigating the potential pathogenic effects of *SNUPN* mutations, primary fibroblasts were isolated from three affected patients (F1$^{m1/m1}$, F2$^{m2/m3}$, F4-II:2$^{m6/m6}$, F4-II:3$^{m6/m6}$) who carry the homozygous mutation located at the uncharacterized C-terminal domain of *SNUPN*. We then performed quantitative PCR (RT-qPCR) and western blotting to compare mutant and wild-type endogenous Snurportin-1 mRNA and protein levels in mutant and control fibroblast lines. Results of the RT-qPCR experiment revealed no statistically significant difference between the wild-type and mutant (F1$^{m1/m1}$, F2$^{m2/m3}$, F4-II:2$^{m6/m6}$ and F4-II:3$^{m6/m6}$) mRNA levels for *SNUPN* gene in fibroblasts (Figure 3.3). However, analysis of SPN1 level in the whole protein lysate by western blot using anti-

Snurportin-1 antibody showed the presence of two bands with lower kDa and less intensity which corresponds to the truncated SPN1 mutant variants in patients with compound heterozygous mutation ($F2^{m2/m3}$) (Figure 3.4 A, Lane 3). Furthermore, quantification of SPN1 levels in independent protein extracts using four to six western blots analysis revealed a significant decrease in the total level of SPN1 in all the mutants compared to three WTs (Figure 3.4A and 3.4B).

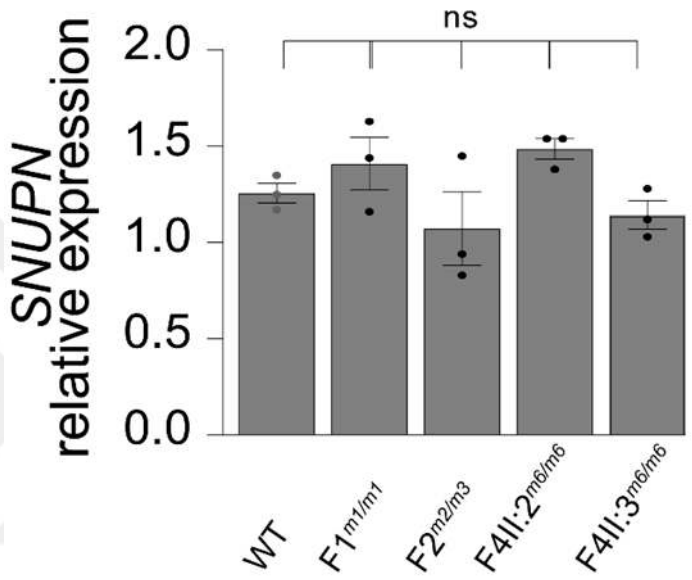

Figure 3.3: RT-qPCR graph of SNUPN mRNA levels in patient fibroblasts. *SNUPN* expression analysis shows no significant difference between the three wild-type (WTs) and four mutants ($F1^{m1/m1}$, $F2^{m2/m3}$, F4-II:$2^{m6/m6}$, F4-II:$3^{m6/m6}$ in fibroblast. Fold change relative to WTs is plotted as mean ± SEM, (n=3 biological replicate). ns, non-significant (two-tailed unpaired t-test) (Nashabat et al., 2024).

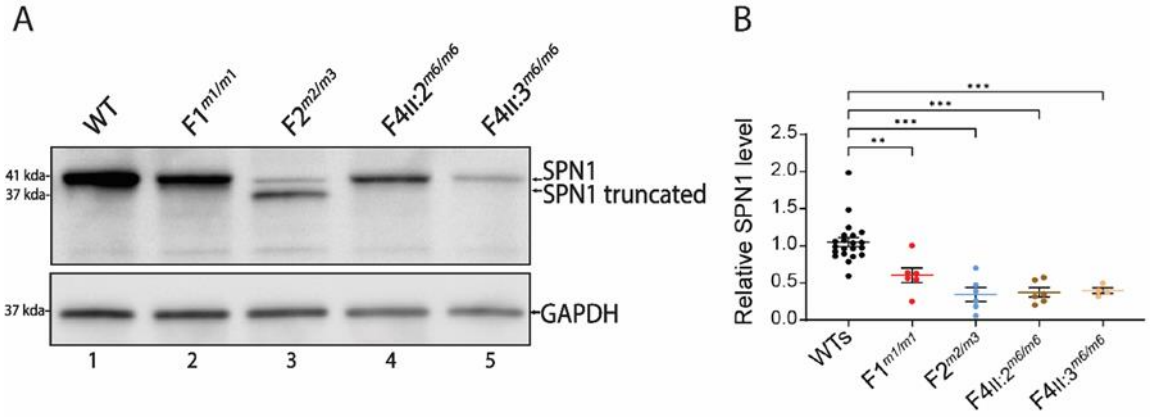

Figure 3.4: Western blot analysis of SPN1 levels in mutants. A) Western blot result of SPN1 protein levels in patients F1$^{m1/m1}$, F2$^{m2/m3}$, F4-II:2$^{m6/m6}$, F4-II:3$^{m6/m6}$ compared to WT fibroblasts. GAPDH served as a loading control. B) Quantification of SPN1 protein levels in patients F1$^{m1/m1}$, F2$^{m2/m3}$, F4-II:2$^{m6/m6}$, F4-II:3$^{m6/m6}$ compared to WTs (n=3) fibroblasts using four to six independent western blots. In the analysis, each dot indicates one quantification. Fold change relative to WTs is plotted as mean ± SEM. ns, non-significant; **P < 0.01; ***P < 0.001 (two-tailed unpaired t-test) (Nashabat et al., 2024).

3.1.3 In silico Predictions of SPN1 Mutants 3D Structure

SPN1 residues between 254 and 320 harbor the majority of the pathogenic variants (m1-m8) which are expected to alter either the TMG binding domain or the C-terminus region which have not been characterized yet (Figure 3.2B). Marwan Ahmad Nashabat, PhD in the lab investigated the structural impacts of the pathogenic variants on SPN1 by using prediction models generated with I-TASSER server (2015) and PyMOL Molecular Graphics System computational tools based on SPN1 reference proteins (PDB 3GJX and 1XK5) (Fig. 3.5)

Structural prediction reveals a well-defined alpha helix in the C-terminal domain between Gly303 and Glu322 in WT SPN1 protein (Nashabat et al., 2024). Notably, this region is predicted to be partially (*m*1) or completely (*m*6 and *m*7) disrupted in mutant SPN1 proteins (Figure 3.5).

Alpha helices are well known to play a pivotal role in the structural stability of multiple protein subunits (Yakimov et al., 2016). We thus hypothesized that structural alterations at the C-terminus of the mutant SPN1 proteins might cause an impairment in their ability to form oligomeric complexes. This hypothesis was further supported by a

previous study showing that N- and C-terminus domains of SPN1 protein can form an intramolecular interaction (Ospina et al., 2005).

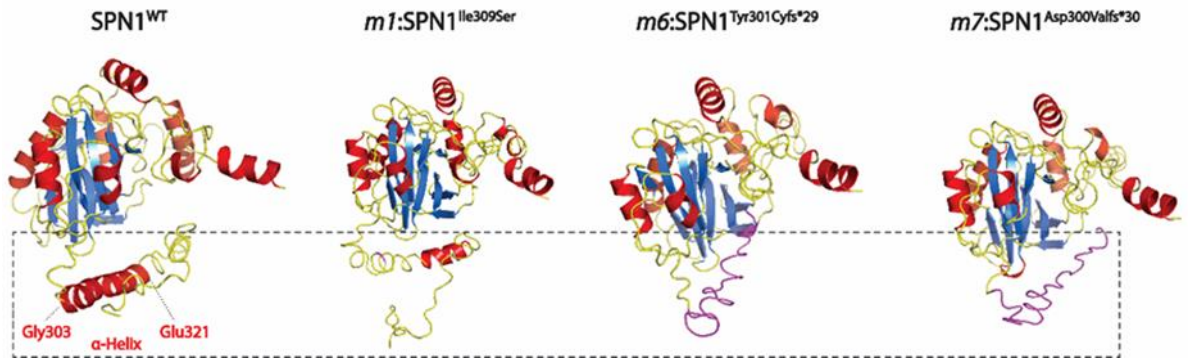

Figure 3.5: In silico predictions of wild-type (WT) and mutant* SPN1 *proteins. Predictions for the 3D structure of the wild-type SPN1 protein and SPN1 models for homozygous pathogenic variants (*m*1, *m*6, and *m*7) affecting C-terminus (shown in the dotted box). Colors of the complete protein structure are based on the secondary structures with alpha helices in red, beta sheets in blue, and loops in yellow. Magenta color was used to represent mutated amino acids in m6 and m7 patients' models (Nashabat et al., 2024).

3.1.4 SPN1 C-Terminal Domain Mutations Impair Self-Interaction in Patient Fibroblast

To investigate our hypothesis regarding the self-interaction of the endogenous SPN1 protein, we analyzed cytoplasmic fractions extracted from WT, F1$^{m1/m1}$, F2$^{m2/m3}$, F4-II:3$^{m6/m6}$ mutant fibroblasts using western blot. We excluded the usage of reducing agents (Dithiothreitol (DTT)) and blotted with an SPN1-specific antibody the entire membrane. Normally, DTT is commonly used in biochemistry to destabilize alpha helix structures by reducing or breaking disulfide bonds, thereby unfolding the 3D structure of the protein. By strategically excluding DTT in our experiment, we aimed to preserve SPN1 3D conformation and detect its potential self-oligomerization.

Our data illustrated in Figure 3.6 revealed:

(1) A band around 41 kDa in all the samples corresponding to the SPN1 monomer,

(2) The higher band at around 150 kDa that was decreased in the missense mutant (F1$^{m1/m1}$) and completely absent in compound heterozygous and frameshift mutants (F2$^{m2/m3}$ and F4-II:3$^{m6/m6}$ respectively) that display more severe C-terminal disruption.

(3) However, the observed band at 100 kDa provided inconsistent results between independent experiments. Therefore, we are not yet able to explain its presence and specificity.

These preliminary data provided strong support for our hypothesis regarding the self-oligomerization potential of the SPN1 protein. We concluded that due to its size, the band around 150 kDa likely corresponds to an SPN1 tetramer. Notably, this oligomer formation appeared to be hindered in all the samples carrying mutations affecting the C-terminal.

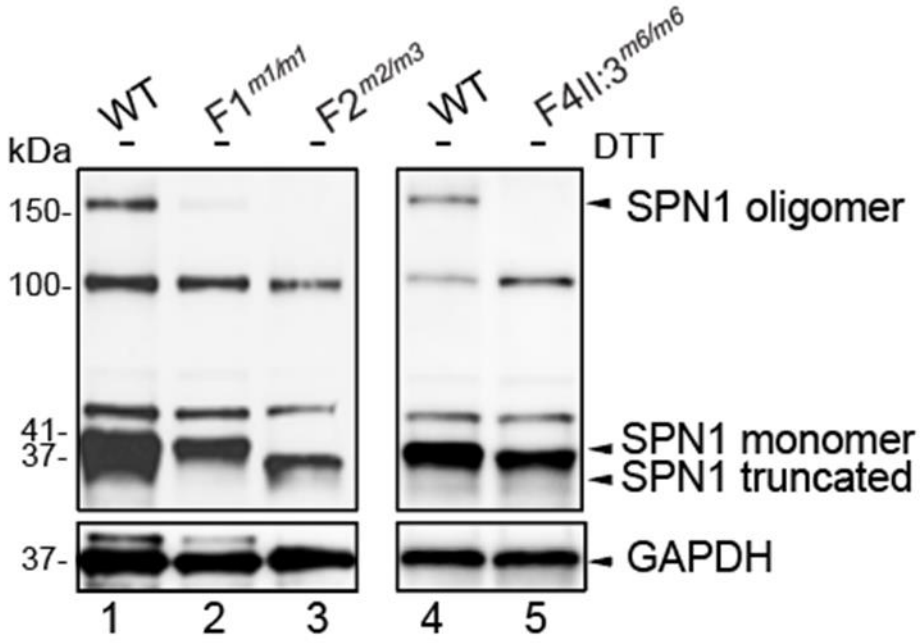

Figure 3.6: Western blot analysis of endogenous SPN1 protein in fibroblast cytoplasmic fraction. SPN1 bands appear at 41 kDa, 100 kDa and 150 kDa. SPN1 oligomer band (150 kDa) was faint in the missense mutant (F1$^{m1/m1}$) and entirely vanished in compound heterozygous (F2$^{m2/m3}$) and frameshift (F4II-2$^{m6/m6}$) mutant samples. GAPDH was used as a loading control.

3.1.5 Hypothesis

According to our structural predictions, the C-terminus of SPN1^{WT} harbors an alpha-helix spanning from Gly303 to Glu322. This region is predominantly composed of non-polar amino acid residues, known to facilitate hydrophobic interactions. Integrating our structural prediction and experimental data, our hypothesis has evolved to suggest that SPN1 may form homo-dimeric and/or homo-tetrameric complexes through hydrophobic intramolecular interactions facilitated by the alpha helix structure at the C-terminal (Figure 3.7). Furthermore, we propose that structural alterations in the alpha helix, induced by mutations, could potentially impede SPN1`s cellular function by disrupting the self-interaction of monomers for the formation of a higher-order complex.

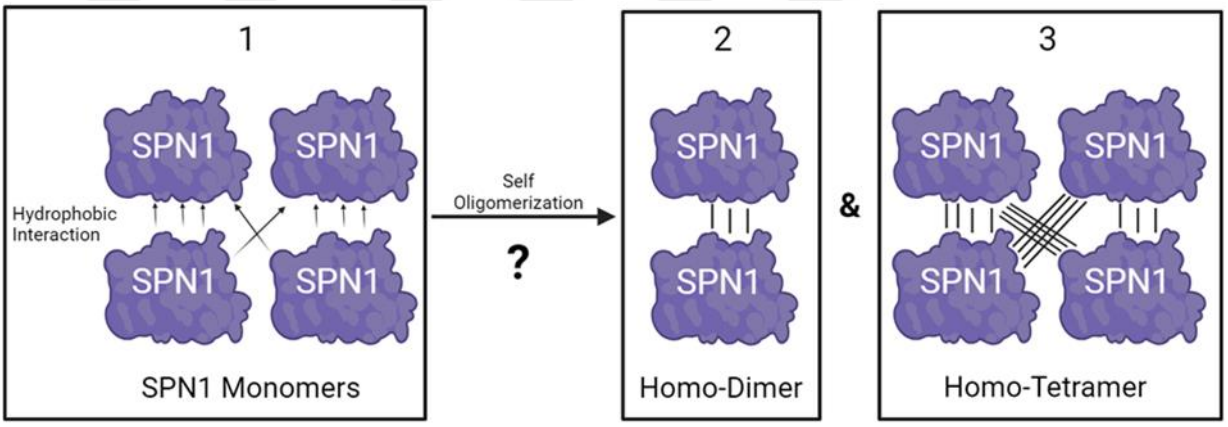

Figure 3.7: Illustration of the hypothesis of SPN1 Self-Oligomerization. (1) The diagram shows the self-oligomerization hypothesis of SPN1 monomers (in purple) through hydrophobic interactions. Arrows (right) indicate the possible interaction routes and lines (left) indicate the formation of probable hydrophobic interactions. Schematic also includes the hypothesis about (2) SPN1 homo-dimer formation, suggested by the band at 100 kDa, and (3) homo-tetramer formation, suggested by the band around 150 kDa.

3.2 Generation of the Tool

Normally, exploring the endogenous protein interactions within patient cells offers the most reliable insights for the functional characterization of protein domains. However, low endogenous protein levels present methodological challenges for biochemistry analyses (Prelich, 2012). Thus, to deepen our understanding of the SPN1 C-terminal domain, we chose to transfect *SNUPN* WT or mutant plasmids in HEK293T/HeLa, well-established cell lines used for transient gene overexpression.

3.2.1 Construct Design

Transient gene overexpression is a useful technique for functional studies that enables the bypass of methodological challenges and aids the investigation of the role of proteins and their domains. One of the most common approaches for establishing a transient overexpression tool is the integration of the gene of interest into an expression plasmid vector. For the initial step of generating overexpression tools, we designed primers that specifically target the coding sequence (CDS) of the human *SNUPN* gene (NM_005701.4) to integrate it into mammalian expression vector pCS2+. In the design of the forward primer, we included additional 3 bases at 5` to increase restriction digestion efficiency, followed by a BamHI restriction enzyme recognition site for directional cloning. To enhance the initiation of the translation we included a leader sequence of 6 nucleotides before the first ATG. After the start codon, either a Flag or Myc-epitope tag sequence was incorporated at the N-terminal for specific detection and isolation of overexpressed proteins. The design is concluded with the rest of the CDS to ensure specific primer initiation (Figure 3.8A). Moreover, the design of the reverse primer contains a 3`-CDS, stop codon to terminate the translation, XhoI restriction enzyme recognition site, and leader sequence (Figure 3.8B).

In our primer design strategy, we determined the length of the additional bases according to manufacturer instructions (Cleavage Close to the End of DNA Fragments, New England Biolabs) to increase restriction digestion efficiency. Furthermore, BamHI and XhoI enzyme sequences were specifically chosen for the restriction digestion since these enzymes do not exhibit any cleavage activity neither within the insert nor outside of the multiple coding site of the pCS2+ plasmid. Flag or Myc-epitope tags were integrated into forward primer specifically to tag the proteins from the N-terminal for a couple of reasons: (1) to preserve the native conformation of the C-terminal for our investigation, and (2) generation of the frameshift patient mutations will disrupt the C-terminal along with the epitope tag.

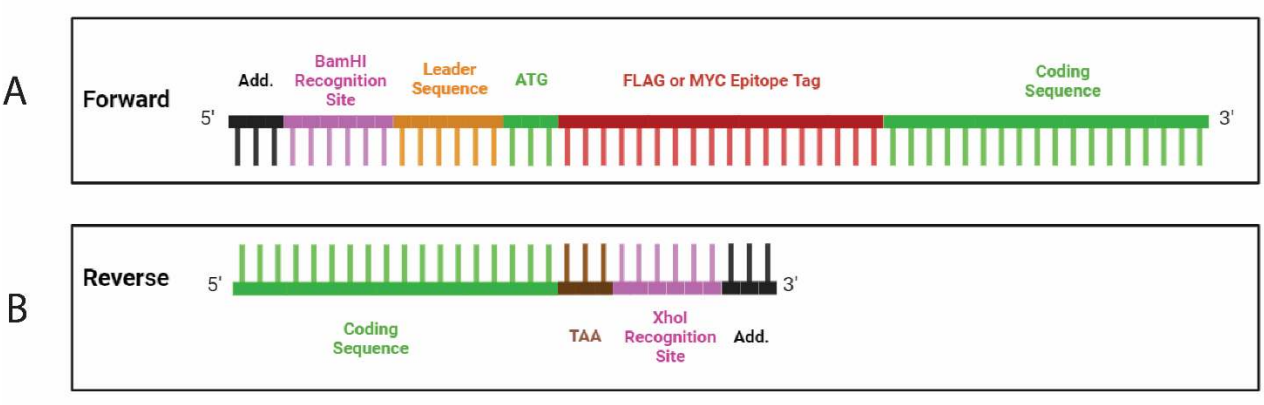

Figure 3.8: : Illustration of the primer designs used in molecular cloning of SNUPN. A) Sequence blueprint of the forward primer. Sequence order is composed of additional nucleotides (Add.) (black), BamHI restriction enzyme recognition site (pink), Leader sequence (orange), start codon (ATG) (green), Flag or Myc epitope tag sequence (red) and the coding sequence (green). B) Sequence blueprint of the reverse primer which consists of the coding sequence (green), stop codon, XhoI restriction enzyme recognition site (pink), and additional nucleotides (Add.).

3.2.2 Generation of the wild-type SNUPN constructs

At the initial stage, we optimized a PCR reaction with two different annealing temperatures to amplify the wild-type *SNUPN* gene with desired modifications for epitope tagging and molecular cloning. The aim behind the usage of a single PCR reaction with two different annealing temperatures was to accommodate the difference between the melting temperatures (Tm) of the primers. This difference is due to the inclusion of the 6-nucleotide leader sequence and epitope tag sequences in the forward primer.

Following the successful amplification of the *SNUPN* WT inserts (*SNUPN* N-Flag WT, *SNUPN* N-Myc WT) (Figure 3.9A), PCR reactions were polyadenylated and integrated into pGEM-T easy vector via TA cloning. This intermediate employment of the pGEM-T easy vector provides a couple of advantages in the process of molecular cloning: (1) integration of the gene of interest into a stable plasmid backbone for storage, (2) obtaining high amounts of the gene of interest to be used in subsequent restriction digestion reaction.

After integrating the gene of interest in the pGEM-T easy vector, a restriction digestion reaction was performed with the BamHI and XhoI enzymes. This reaction is carried out for the pGEM-T easy vectors that either contain *SNUPN* N-Flag WT or *SNUPN* N-Myc WT, as well as the empty pCS2+ mammalian expression vector.

Next, inserts were isolated by gel extraction and integrated into the pCS2+ vector via ligation reaction. Ligates of *SNUPN* N-Flag WT and *SNUPN* N-Myc WT were transformed into DH5-alpha bacteria for selection and amplification. Subsequently, the isolated plasmids were digested with BamHI and XhoI enzymes to validate the integration of the inserts into the pCS2+ vector. The results of the gel analysis showed that both *SNUPN* N-Flag WT and *SNUPN* N-Myc WT inserts were successfully integrated into the pCS2+ vector (Figure 3.9B).

Cycle sequencing of the constructs, with the usage of primers specific to the SP6 and T7 promoter regions located at the 5` and 3` ends of the multiple cloning site of the pCS2+ vector, further validated the successful integration of the inserts into the vector. Also, it confirmed the successful cloning of *SNUPN* N-Flag WT, and *SNUPN* N-Myc WT without any unspecific mutation (Figure 3.13)

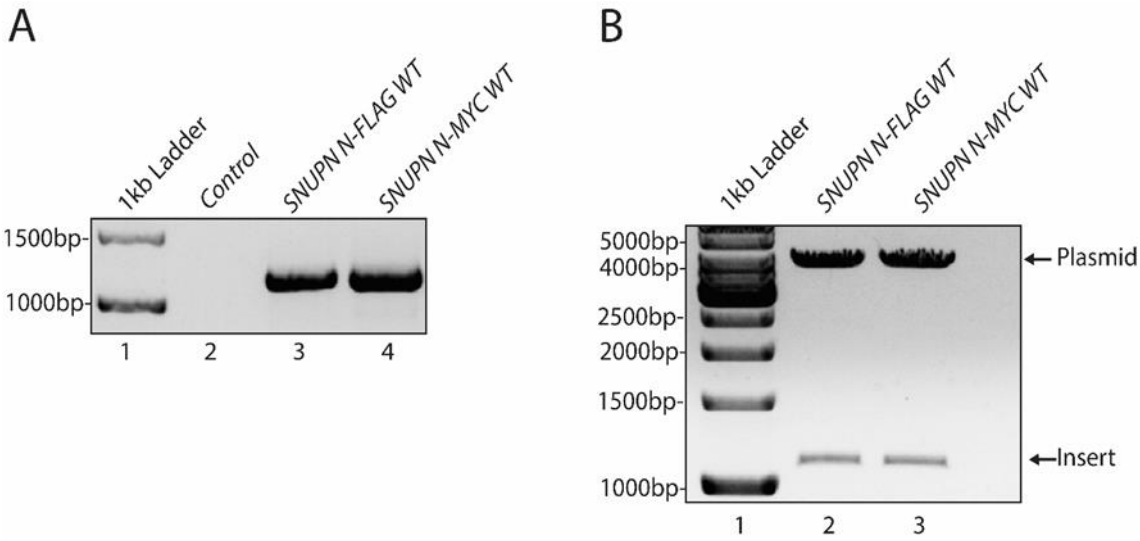

Figure 3.9: PCR amplification and integration of SNUPN constructs into a mammalian expression vector. A) Agarose gel image of the PCR amplification of the Flag (3), and Myc (4) epitope tagged WT *SNUPN* coding sequence. Bands are at the correct size for both Flag (1107 base pair (bp)) and Myc (1113 base pair (bp)) epitope-tagged inserts. Samples migrated under 120V for 25 minutes in 0.8% agarose gel.

B) Validation of the integration of the inserts into mammalian expression plasmid via restriction digestion. For both *SNUPN* N-Flag WT (2), and *SNUPN* N-Myc WT (3) constructs, upon digestion with the two restriction enzymes used for cloning and migration on agarose gel, two bands were observed. A band at 4000 base pair (bp) and one at 1000 base pair (bp) correspond to pCS2+, the mammalian expression plasmid and *SNUPN* WT inserts, respectively.

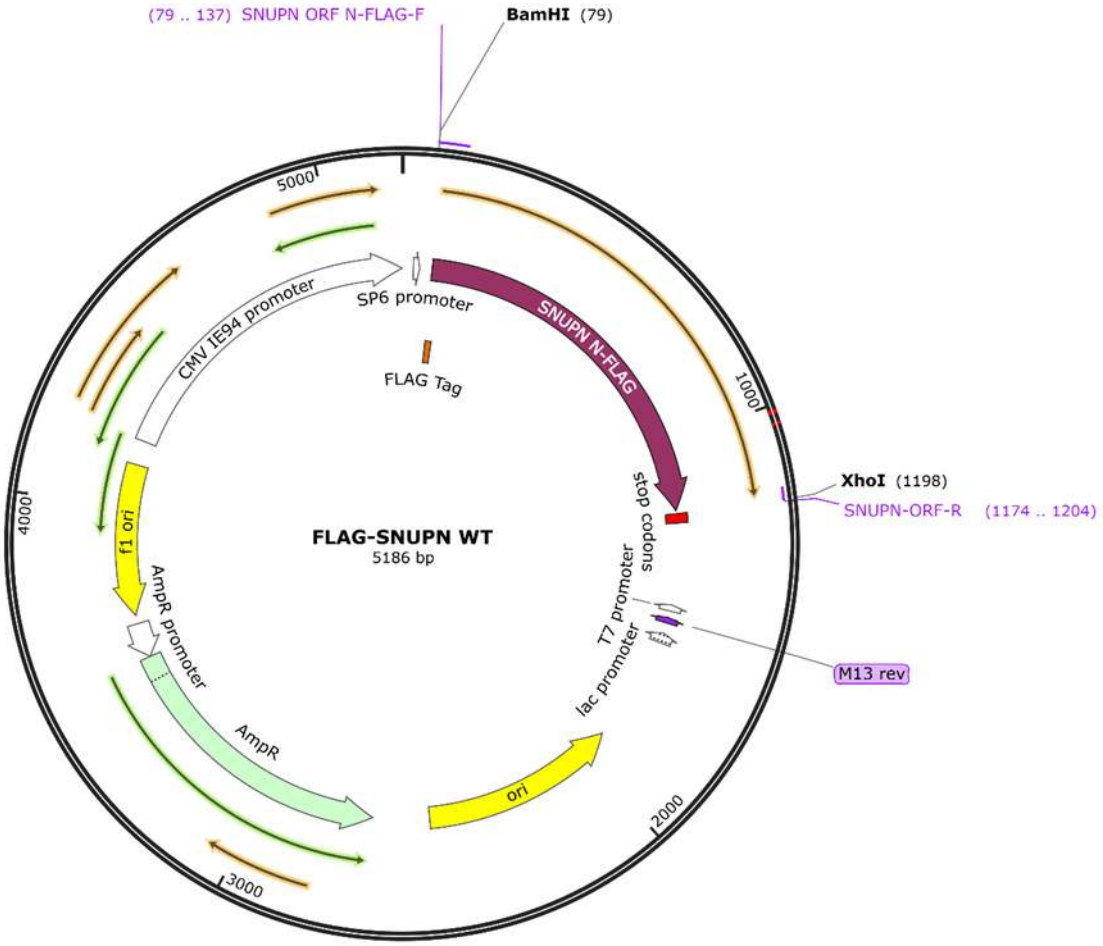

Figure 3.10: Circular plasmid map of the Flag-SNUPN wild-type construct. Genetic map of the 5186 bp pCS2+ plasmid that contains Flag-*SNUPN* WT (*SNUPN* N-Flag, in purple). The map is generated by using SnapGene software (from Insightful Science).

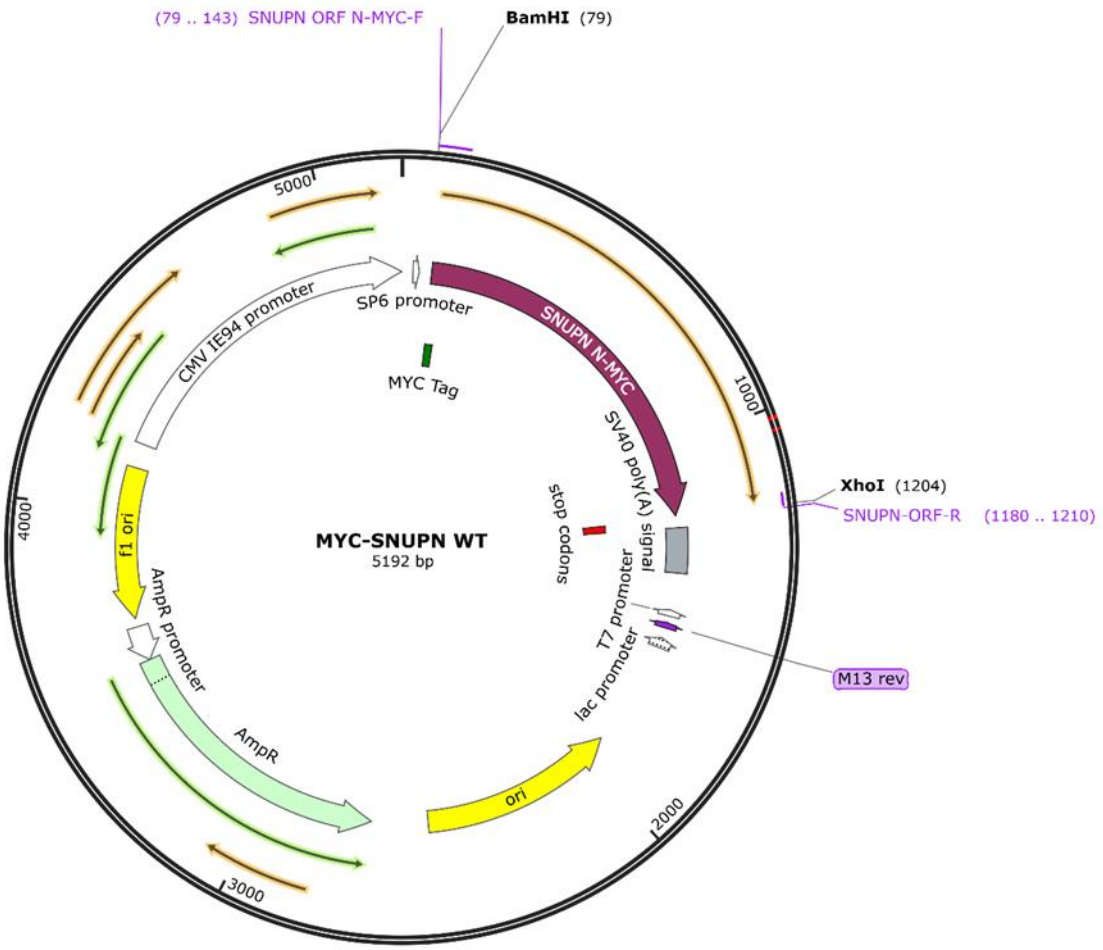

Figure 3.11: Circular plasmid map of the Myc-SNUPN wild-type construct. Genetic map of the 5192 bp pCS2+ plasmid that contains Myc-*SNUPN* WT (*SNUPN* N-Myc, in purple). The map is generated by using SnapGene software (from Insightful Science).

3.2.3 Generation of the Homozygous Pathogenic Mutations

Pathogenic mutations in the *SNUPN* gene are localized between SPN1 residues 254 and 320 in a region identified as a mutational hotspot. They are anticipated to alter the TMG binding domain (aa. 162 to 301) or the uncharacterized C-terminus domain (aa. 301 to 320). To further evaluate the impact of these mutations on the structure and function of SPN1 protein, we next generated homozygous pathogenic patient mutation constructs by using wild-type (WT) *SNUPN* N-Flag and wild-type (WT) *SNUPN* N-Myc constructs as templates.

For this purpose, three mutations were selected: a missense mutation from our index patient (F1$^{m1/m1}$) and two frameshift mutations from family 4 and family 5 (F4-II:2$^{m6/m6}$, F5-II:2$^{m7/m7}$). The missense mutation changes only a single amino acid in the C-terminal region, leading to a mild phenotype in the patient, reduced levels of SPN1, and impaired oligomer formation. Conversely, frameshift mutations result in premature termination codon and the loss of the SPN1 C-terminal region. Notably, these frameshift mutations lead to a much stronger patient phenotype, accompanied by a reduction in SPN1 levels and complete loss of oligomer formation.

These constructs were generated by using the "Agilent-QuikChange Primer Design" tool that design complementary primer sequences containing the desired modification (single nucleotide substitution for *m*1: c.926T>G; p.lle309Ser, two nucleotide deletion for *m*6: c.902_903delAT; p.Tyr301Cysfs*29 and *m*7: c.899_900delAC; p.Asp300Valfs*30) at the center of the primers. Following the amplification of the mutant strand by PCR reaction, the plasmid mix was subjected to DpnI enzyme for the cleavage of the parental DNA. DpnI specifically recognizes and cleaves methylated DNA, thus removing only the parental DNA, which is methylated, to avoid contamination of the WT construct. Finally, the mutant construct samples were transformed into bacteria for the nick repair, selection, and amplification (Figure 3.12). Plasmids were isolated from the bacteria, and finally, mutant constructs without any undesired mutation were validated by Cycle sequencing.

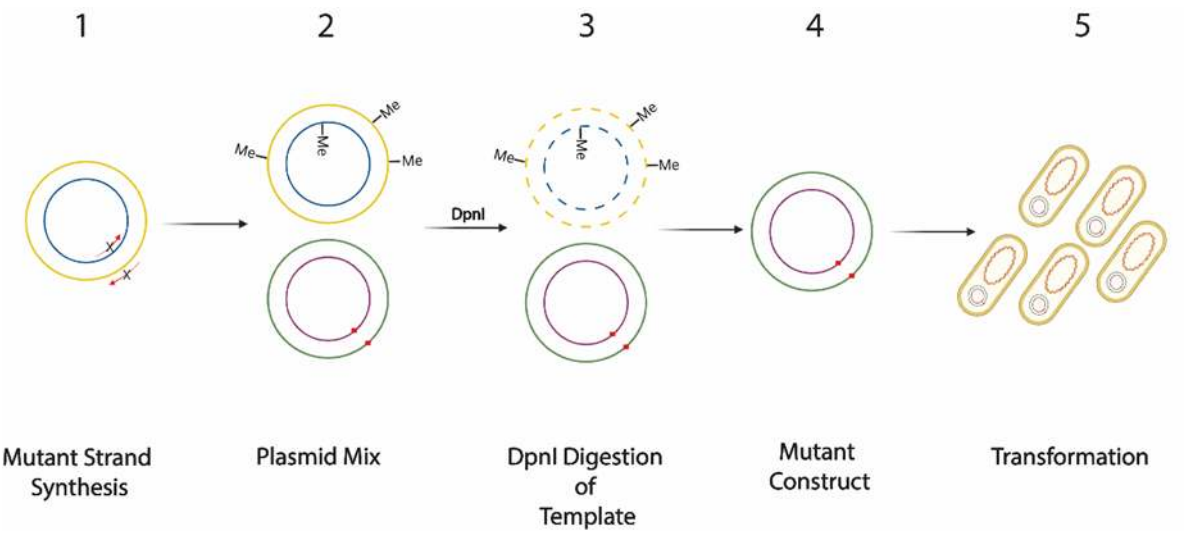

Figure 3.12: Site-directed mutagenesis for the creation of homozygous pathogenic SNUPN mutants. The schematic diagram shows the experimental flow of the site-directed mutagenesis to create homozygous pathogenic *SNUPN* mutants located at the uncharacterized C-terminal region of the gene. (1) Mutant strand synthesis via PCR reaction, (2-3) Treatment of the heterogeneous plasmid mixture with DpnI enzyme, which is composed of methylated parental DNA (yellow and blue), and unmethylated mutant construct (green and purple). (4-5) Obtaining and transformation of the mutant construct to XL10-Gold Ultracompetent Bacteria Cells for nick repair, amplification, and selection.

3.2.4 Validation of the Tools

As previously mentioned, the final step of each construct generation involved the verification of the *SNUPN* sequence through cycle sequencing, a modified version of Sanger sequencing for the determination of DNA sequence (Blakesley, 1993). The main difference between the two methods is the incorporation of a heat-resistant DNA polymerase in Cycle sequencing, giving the advantage of repeating the sequencing reaction by using DNA over and over again in the same tube as a PCR reaction (Blakesley, 1993; Platt et al., 2007).

For this purpose, primers specific to the SP6 and T7 promoter regions located at the 5` and 3` ends of the multiple cloning site of the pCS2+ vector, were used. The analysis confirmed the successful integration of the Flag and Myc epitope tags at the N-terminus of the *SNUPN* WT constructs (*SNUPN* N-Flag WT, and *SNUPN* N-Myc WT respectively) (Figure 3.13). Additionally, cycle sequencing also further validated their integration into the pCS2+ vector. mutant constructs *m*1: c.926T>G; p.lle309Ser, *m*6: c.902_903delAT; p.Tyr301Cysfs*29 and *m*7: c.899_900delAC; p.Asp300Valfs*30 were validated as well.

The final step of validation for all the constructs was to assess their ability to express SPN1 in easy transfectable cellular system such as HEK293T. Protein lysates were prepared from HEK293T cells transfected with wild-type or mutant, Flag- or Myc-Tagged *SNUPN* constructs. Flag or Myc epitope tag-specific antibodies revealed the presence of SPN1 monomer at the correct size (around 41 kDa) in all samples, except for the untransfected control. This specific detection indicated the successful generation, transfection, and translation of the constructs (Figure 3.14).

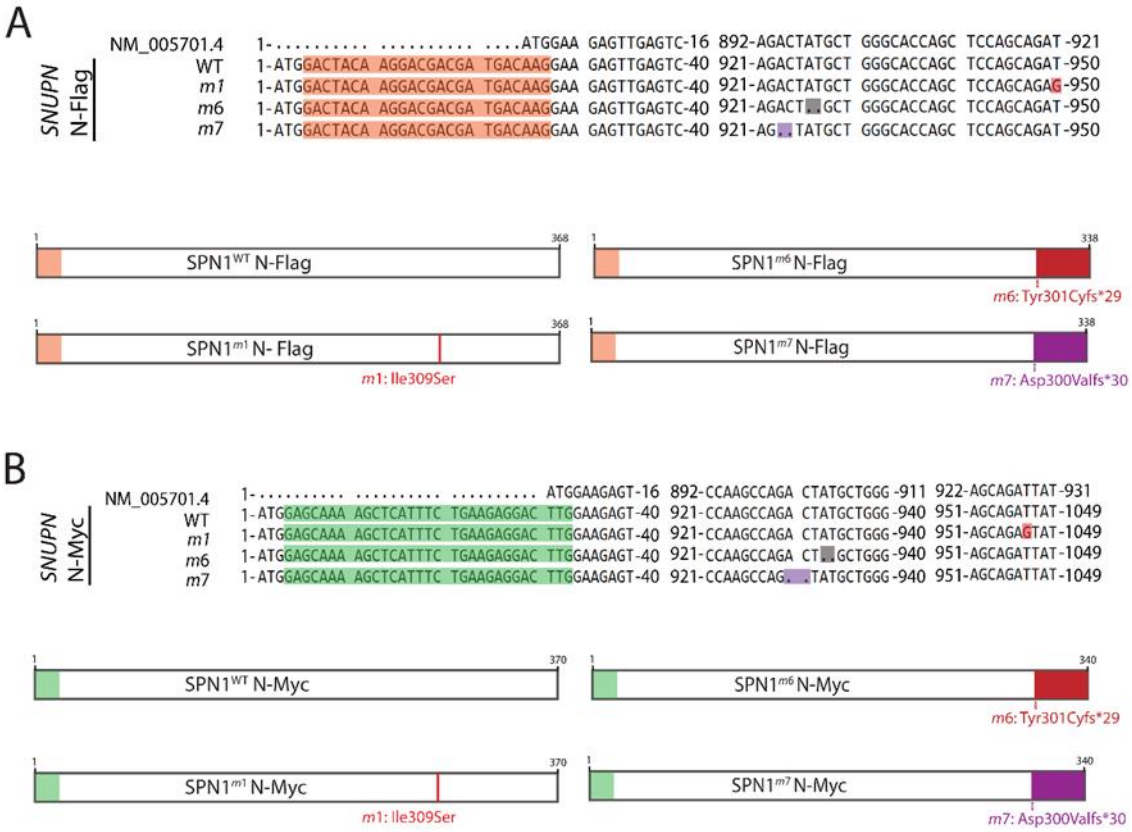

Figure 3.13: Validation of the generated tools by Cycle sequencing. A) Top: Multiple sequence alignment of sequencing results of the *SNUPN* N-Terminal Flag (highlighted in orange) tagged constructs showing wild-type (WT) and variants m1 (orange), m6 (red), and m7 (purple). Bottom: schematics of the corresponding SPN1 proteins. B) Top: Multiple sequence alignment of sequencing results of the *SNUPN* N-Terminal Myc (highlighted in green) tagged constructs showing wild-type (WT) and variants m1 (orange), m6 (red), and m7 (purple). Bottom: schematics of the corresponding SPN1 proteins.

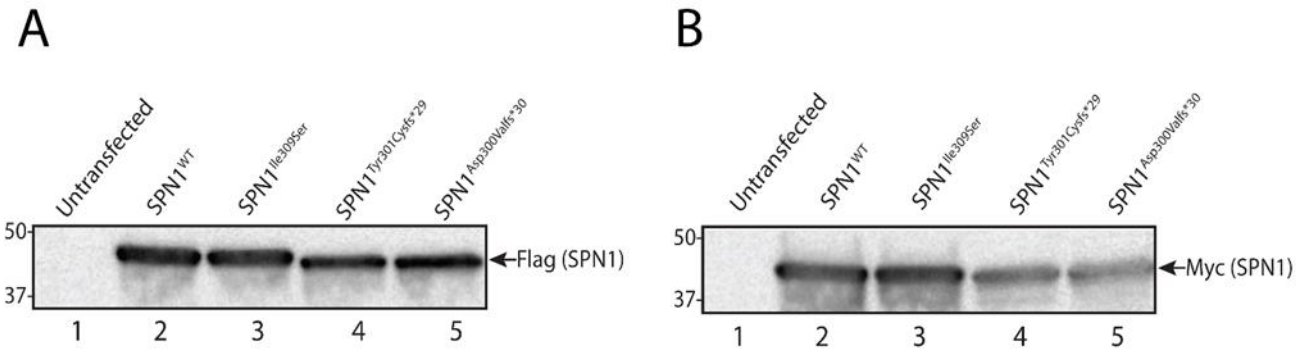

Figure 3.14: Validation of the constructs by western blot. Western blot assay was performed with the protein samples from HEK293T cells transfected with wild-type or mutant, Flag or Myc-Tagged SPN1 constructs. A) Membrane blotted with anti-Flag antibody reveals the expression of SPN1 for wild-type (WT) (2) and mutant (3-5) samples. B) Membrane blotted with anti-Myc antibody reveals the expression of SPN1 for wild-type (WT) (2) and mutant (3-5) samples. Non transfected HEK293T proteins were used as a negative control to check the specificity of the SPN1 expression.

3.3 Wild-Type SPN1 is Capable of Self-Oligomerization

To further investigate the self-oligomerization capability of SPN1 protein we overexpressed Myc-epitope tagged $SPN1^{WT}$ construct in HeLa cells. Cytoplasmic fractions from the SPN1 overexpressed sample, along with the untransfected control were analyzed by western blot experiment in the absence of DTT, a reducing agent. A Myc-tag-specific antibody was used to enable the detection of the SPN1 protein in the entire membrane. We observed in the sample $SPN1^{WT}$ overexpressed, the presence of two robust bands at 41 kDa and 150 kDa and a faint band around 75 kDa likely corresponding to the SPN1 dimer (Figure 3.15, lane 2). This data corroborated our findings obtained in primary fibroblasts demonstrating the presence of SPN1 oligomer band at 150 kDa whose size appears to correspond to a tetramer complex (Figure 3.6, Lane 1).

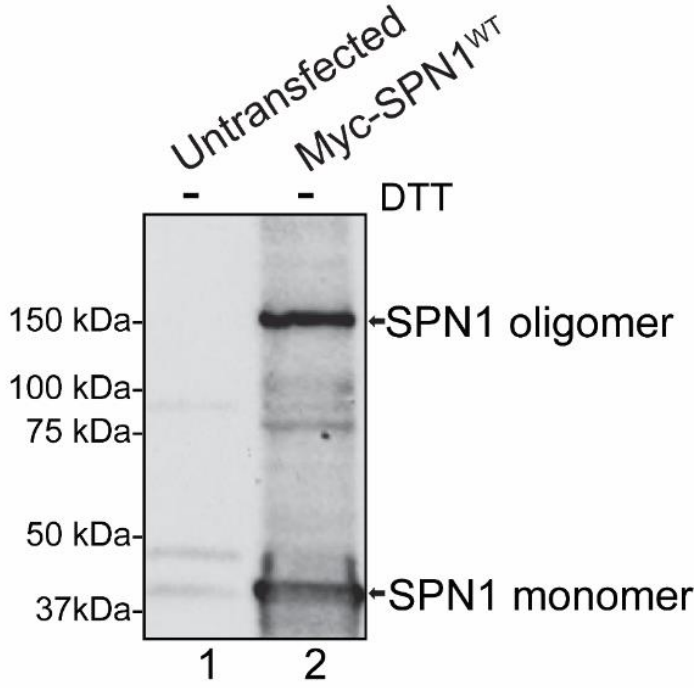

Figure 3.15: : Wild-type SPN1 is capable of self-oligomerization. Western blot analysis of the HeLa cells cytoplasmic fractions without DTT treatment. In the Myc-SPN1^{WT} sample, two major specific bands at 41 and 150 kDa appear in Myc-SPN1^{WT} sample compared to untransfected sample. Note the presence of a faint band above 100 kDa.

3.4 SPN1 Mutants Affect Self Oligomerization

Next, to further confirm the role of the C-terminus of SPN1 in its self-oligomerization and validate our findings, we compared the overexpression of Myc-epitope tagged SPN1^{WT} with the two frameshift mutants SPN1^{Tyr301Cysfs*29} and SPN1^{Asp300Valfs*30} constructs in HeLa cells. We specifically chose these mutant constructs based on the severe predicted impairment of their alpha helix in the C-terminus and the complete loss of oligomerization in mutant fibroblast (Figure 3.6, Lanes 2, 3, and 5).

A western blot experiment was performed without the use of DTT, and the entire membrane was probed with e Myc epitope tag-specific antibody. This analysis revealed the presence of the SPN1 monomer around 41 kDa in all samples. As expected, a higher band at 150 kDa band was present in SPN1^{WT} (Lane 1) whereas in frameshift mutant samples (SPN1^{Tyr301Cysfs*29} and SPN1^{Asp300Valfs*30}), it was completely absent (Lanes 2 and 3). Results were difficult to interpret again for the band at 100 kDa (Figure 3.16A).

In the following experiment, cytoplasmic extracts of the Myc-tagged SPN1^{WT}, SPN1^{Ile309Ser}, SPN1^{Tyr301Cysfs*29}, and SPN1Asp^{300Valfs*30} constructs were treated with and without DTT allowing us to perform a comparative analysis. Our findings revealed the presence of a 150 kDa band in SPN1^{WT} lysate without DTT treatment (Figure 3.16B, Lanes 1 and 2). Notably, the SPN1 oligomer band at 150 kDa appeared moderately faint in SPN1^{Ile309Ser} and completely disappeared in lysates from SPN1^{Tyr301Cysfs*29} and SPN1^{Asp300Valfs*30} without treatment (Figure 3.16B, Lanes 2, 3 and 4). Interestingly, DTT treatment of overexpressed SPN1^{WT} and SPN1^{Ile309Ser} samples resulted in the total disappearance of oligomeric complexes (Figure 3.16B, Lanes 5 and 6). Moreover, DTT treatment of the SPN1^{Tyr301Cysfs*29} and SPN1^{Asp300Valfs*30} did not show any significant differences compared to their untreated counterparts (Figure 3.16B, Lanes 7 and 8).

Thus, we can conclude that SPN1 oligomers are resistant to dissociation via SDS. However, these structures were sensitive to dissociation through DTT treatment, indicating the covalently linked nature of the interactions. Moreover, the SPN1 self-oligomerization process seems to be significantly hindered by mutations that induce alteration of the 3D conformation of the protein's C-terminal region.

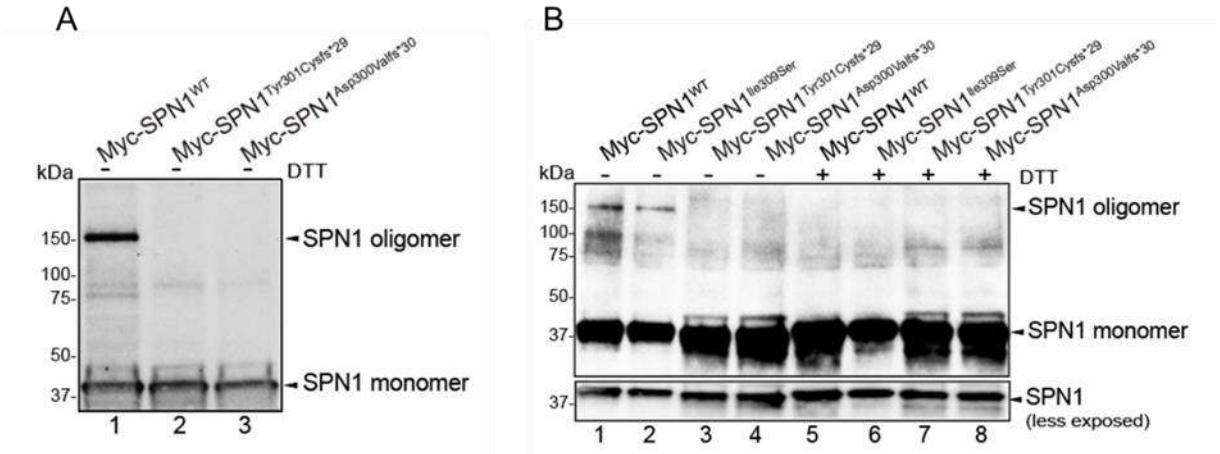

Figure 3.16: SPN1 mutants affect self-oligomerization. A) Western blot analysis of HeLa cells cytoplasmic fraction transfected either with Myc-SPN1^{WT} (Lane 1) Myc SPN1^{Tyr301Cysfs*29} (Lane 2) and Myc-SPN1^{Asp300Valfs*30} (Lane 3). SPN1 oligomer band at 150 kDa entirely disappeared in the Myc-SPN1^{Tyr301Cysfs*29} and Myc-SPN1^{Asp300Valfs*30} sample in absence of DTT. B) Analysis of cytoplasmic fraction of HeLa cells transfected with wild-type or mutant Myc-SPN1 constructs. two bands at 41 and 150 kDa (SPN1 oligomer) were observed in the samples transfected with Myc-SPN1^{WT} (Lane 1), and Myc-SPN1^{Ile309Ser} (Lane 2). SPN1 oligomer band completely disappeared in lysates transfected with Myc-

SPN1^{Tyr301Cysfs*29} or Myc-SPN1^{Asp300Valfs*30}. Band at 150 kDa was not seen in the WT or mutant samples after they were treated with DTT (Lanes 5 and 6).

3.4.1 SPN1 Self-Oligomerization Confirmed by Co-Immunoprecipitation

The next step of our experiments aimed to validate SPN1 self-oligomerization through co-immunoprecipitation. In this experiment, we co-overexpressed SPN1 constructs Myc- and Flag-epitope-tagged in HEK293T cells. Subsequently, their protein lysates were subjected to Flag immunoprecipitation (Figure 3.17)

First, we confirmed that Flag-SPN1^{WT} was capable of specifically pulling down Myc-SPN1^{WT} (Figure 3.17, Lane 2), thereby validating for the first time a self-interaction between SPN1 monomers. Interestingly, as anticipated in the earlier DTT treatment experiment, Flag-SPN1^{Ile309Ser} pulled down Myc-SPN1^{Ile309Ser} in a less effective manner compared to SPN1^{WT} (Figure 3.17, Lane 3). Furthermore, in line with our hypothesis and earlier results, SPN1^{Tyr301Cysfs*29} and SPN1^{Asp300Valfs*30} mutants, predicted to have much more severe structural defects at the C-terminal domain were incapable of forming self-oligomers, unlike the SPN1^{WT} and SPN1^{Ile309Ser} (Figure 3.17, Lanes 4 and 5).

Thus, our overall data strongly suggests that SPN1 can form oligomeric structures, and the C-terminal region of the protein is essential for oligomerization.

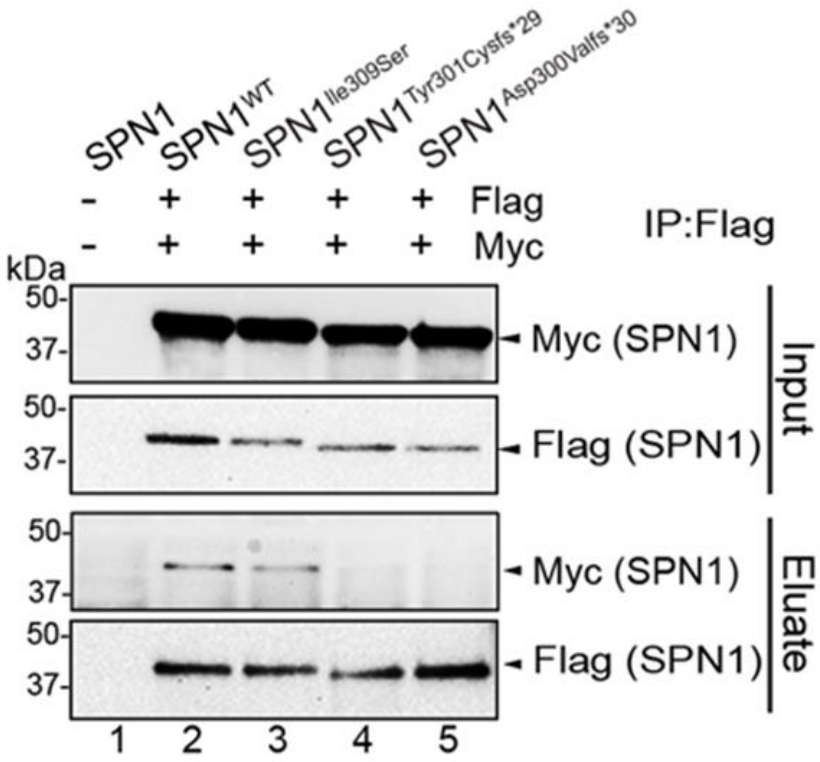

Figure 3.17: Co-immunoprecipitation (co-IP) assay validated SPN1 self-interaction. Western blot analysis on samples co-immunoprecipitated by Flag-specific beads. HEK293T cells co-transfected with Flag- and Myc- epitope-tagged SPN1. Anti-Myc (top) and anti-Flag (bottom) antibodies used to blot input and eluate samples, revealed self-interaction in $SPN1^{WT}$ (Lane 2) and $SPN1^{Ile309Ser}$ (Lane 3) whereas this interaction disappeared in $SPN1^{Tyr301Cysfs*29}$ and Flag-$SPN1^{Asp300Valfs*30}$ (Lane 4 and 5 respectively).

3.5 C-Terminal Domain of the SPN1 is Sufficient for Self-Oligomerization

Overall, our findings suggest that SPN1 protein functions within a higher complex in the cytoplasm, firmly demonstrating the protein's ability to form homodimers or homo-tetramers. This is evident from the observed sizes of the oligomeric bands which are around 100 and 150 kDa. However, the C-terminal region of the protein appears to be more significant for the formation of the tetramer rather than the dimer, as portrayed by the systematic decrease of the 150 kDa band.

3.5.1 In silico Prediction of the C-Terminal Interaction

Alpha helices play a pivotal role in the formation of coiled-coil structures, contributing to the stability and structural integrity of the protein (Lupas et al., 2017). Our structural model pointed out a well-defined alpha helix in the C-terminal domain between Gly303 and Glu322 in WT SPN1 protein which mainly comprises hydrophobic residues. Our experimental results strongly suggest that mutations at the C-terminal domain, which harbors the alpha-helix, disrupt the self-oligomerization capability of SPN1. This strongly suggests that the structural motif of the SPN1 tetramer potentially relies on hydrophobic interactions at the C-terminus in accordance with the coiled-coil concept. To gain a better understanding of the nature of this self-oligomerization and the role of the C-terminal region in the interaction Marwan Ahmad Nashabat Ph.D. student in the laboratory, conducted a prediction of coiled-coil SPN1 oligomeric states using LOGICOIL. The results delineated a region in the C-terminus between amino acids Gly303 to Leu339 that is strongly predicted to form tetrameric structures. Next, this particular region was used to generate a prediction of tetramer using Alphafold2-Colab and Socket2 for the coiled-coil interaction. The best Predicted Aligned error (PAE) resulted in a model in which two antiparallel helices interact with each other in a parallel orientation. Notably, this model suggested that Ile309 residue was directly involved in these hydrophobic interactions (Figure 3.18).

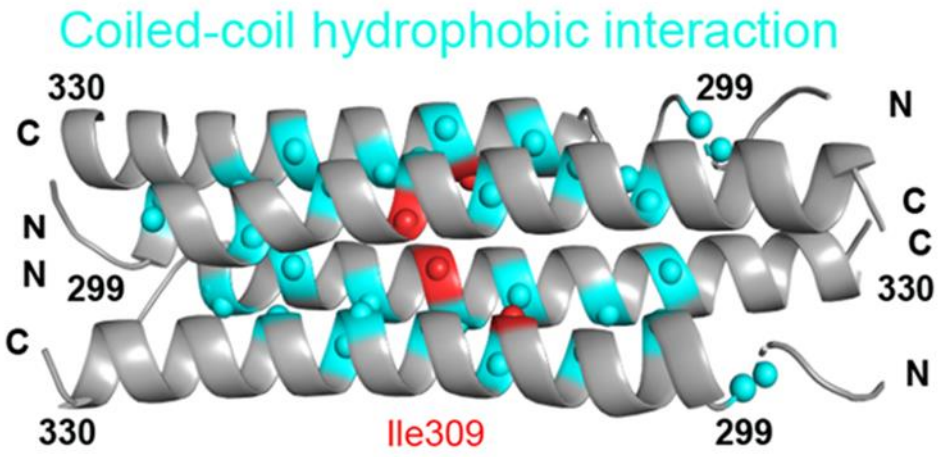

Figure 3.18: In silico prediction of tetramer formation with C-terminal region. Alphafold2 colab and Socket2 coiled-coil prediction of coiled-coil interactions which forms tetramer with C-terminal region (residues 299-330). Predicted hydrophobic interactions are highlighted in blue and Ile309 residue is highlighted in red (Nashabat et al., 2024).

3.5.2 Role of the SPN1 C-Terminus in Self-Oligomerization

In light of our experimental findings and prediction models regarding the self-oligomerization of the SPN1 and the essential role of the C-terminal region, we aimed to further investigate the interaction between WT and mutant C-terminal region. For this purpose, we generated several truncated constructs containing exclusively the C-terminal region of the protein. We selected the region between amino acid residues 254-360 to encompass all our pathogenic variants and tagged the constructs with either Flag or Myc (Flag-SPN1$^{254\text{-}360}$ and Myc-SPN1$^{254\text{-}360}$) (Figure 3.19).

First, Flag-SPN1$^{WT(254\text{-}360)}$ C-terminal fragment construct along with its Myc epitope-tagged counterpart was co-overexpressed in HEK293T cells. Subsequently, we subjected their protein lysates to Flag immunoprecipitation (Figure 3.20B). Our observation indicated that Flag-SPN1$^{WT(254\text{-}360)}$ could interact with its Myc-tagged counterpart (Lane 1). Moreover, Myc-SPN1$^{WT(254\text{-}360)}$ was also able to interact with SPN1 full-length (Flag-SPN1$^{WT(1\text{-}360)}$) confirming that the SPN1 C-terminal region alone is sufficient for the formation of the SPN1 multimeric complex (Figure 3.20B, Lane 2).

After establishing the essential role of the C-terminal part in the formation of the oligomer, our subsequent aim was to investigate the interacting capabilities of the SPN1 mutants using the same experimental setup. For this purpose, we co-overexpressed either wild-type Flag epitope-tagged SPN1 full-length or fragment (SPN1$^{WT(1\text{-}360)}$ SPN1$^{WT(254\text{-}360)}$) constructs along with Myc epitope-tagged SPN1 mutant fragment (Myc-SPN1$^{(254\text{-}360)}$) constructs in HEK293T cells (Figure 3.21A). Additionally, we overexpressed Flag SPN1$^{m1(254\text{-}360)}$ along with its Myc epitope-tagged counterpart. Missense mutant (Myc-SPN1$^{254\text{-}360(Ile309Ser)}$) was capable of interacting with SPN1$^{WT(254\text{-}360)}$ fragments (Figure 3.21B, Lane 1) whereas frameshift mutants (Myc-SPN1$^{254\text{-}328(Tyr301Cysfs*29)}$ and Myc-SPN1$^{254\text{-}328(Asp300Valfs*30)}$) completely lost the ability to interact with SPN1$^{WT(254\text{-}360)}$ fragments (Figure 3.21B, Lane 2 and 3) which validated our previous findings. Interestingly, the binding efficiency of the Myc-SPN1$^{254\text{-}360(Ile309Ser)}$ fragment to its Flag counterpart was significantly lower compared to its binding to the Flag-SPN1$^{WT(254\text{-}360)}$ (Figure 3.21B, Lane 4), confirming our predictions regarding the specific involvement of Ile309Ser residue in the self-oligomerization.

In conclusion, our data demonstrates that SPN1 molecules are capable of forming oligomeric structures through the interaction of their C-terminal region. Furthermore, pathogenic variants affecting the SPN1 C-terminal region can disrupt the self-interaction underscoring the significance of the C-terminal region for stability and functioning of SPN1 (Nashabat et al., 2024).

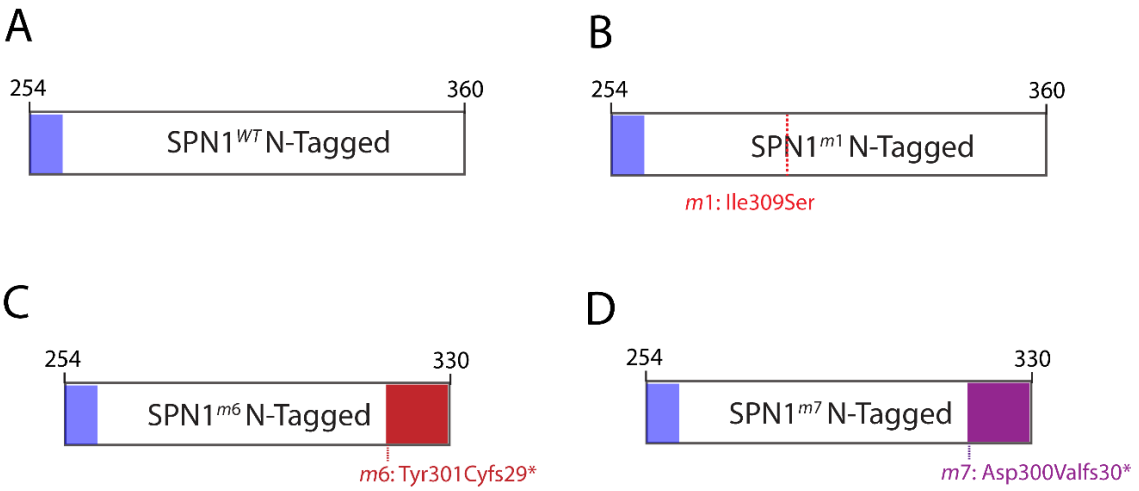

Figure 3.19: Schematic representation of the SPN1 C-terminal truncated constructs region. A) Wild-type SPN1 $^{(254\text{-}360)}$. B) *m*1 mutant SPN1 $^{(254\text{-}360)}$. C) *m*6 mutant SPN1 $^{(254\text{-}360)}$ showing last 29 amino acids modified in red. D) *m*7 mutant SPN1 $^{(254\text{-}360)}$ showing last 30 amino acids modified in red. N-terminal tags represented in (purple).

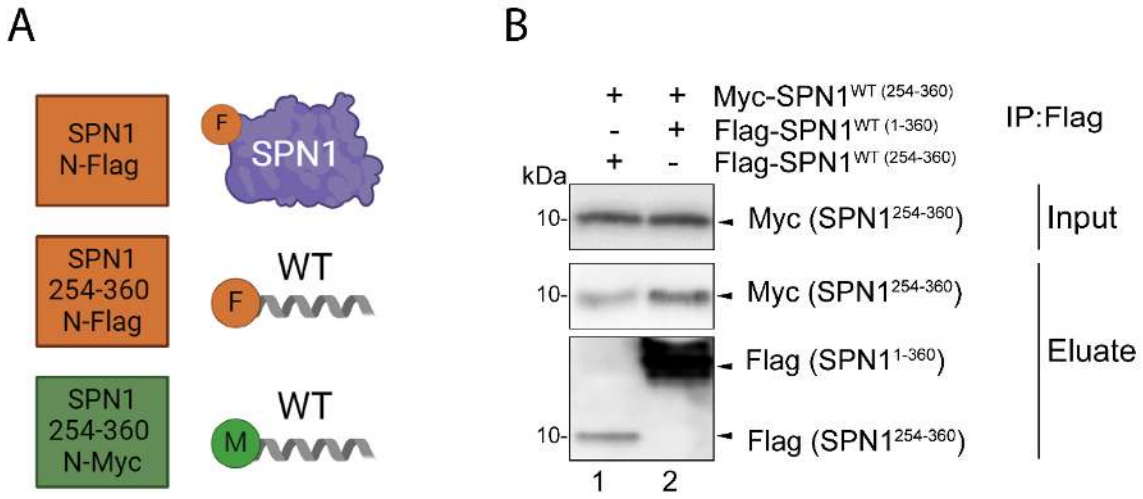

Figure 3.20: Wild-type SPN1 C-Terminus domain is capable of self-interaction. A) Schematic illustration of the structure of the N-terminally Flag (shown in orange) wild-type (WT) full SPN1(1-360) (shown in purple), and wild-type (WT) SPN1 254-360 (C-terminus) fragments tagged with Flag- (orange) or Myc-epitope tag (green) in which they keep their alpha-helix structure (in grey). B) Co-immunoprecipitation (co-IP) assay performed in HEK293T cells co-transfected with Flag- and Myc-epitope-tagged SPN1 C-terminus fragments by the usage of Flag-specific beads. Anti-Myc (top) and anti-Flag (bottom) antibodies used to blot input and eluate samples, revealed self-interaction between Myc-SPN1$^{WT\ (254-360)}$ and Flag-SPN1$^{WT\ (254-360)}$ (Lane 1). Furthermore, it also showed interaction between Myc-SPN1$^{WT\ (254-360)}$ and Flag-SPN1$^{WT\ (1-360)}$ (Lane 2).

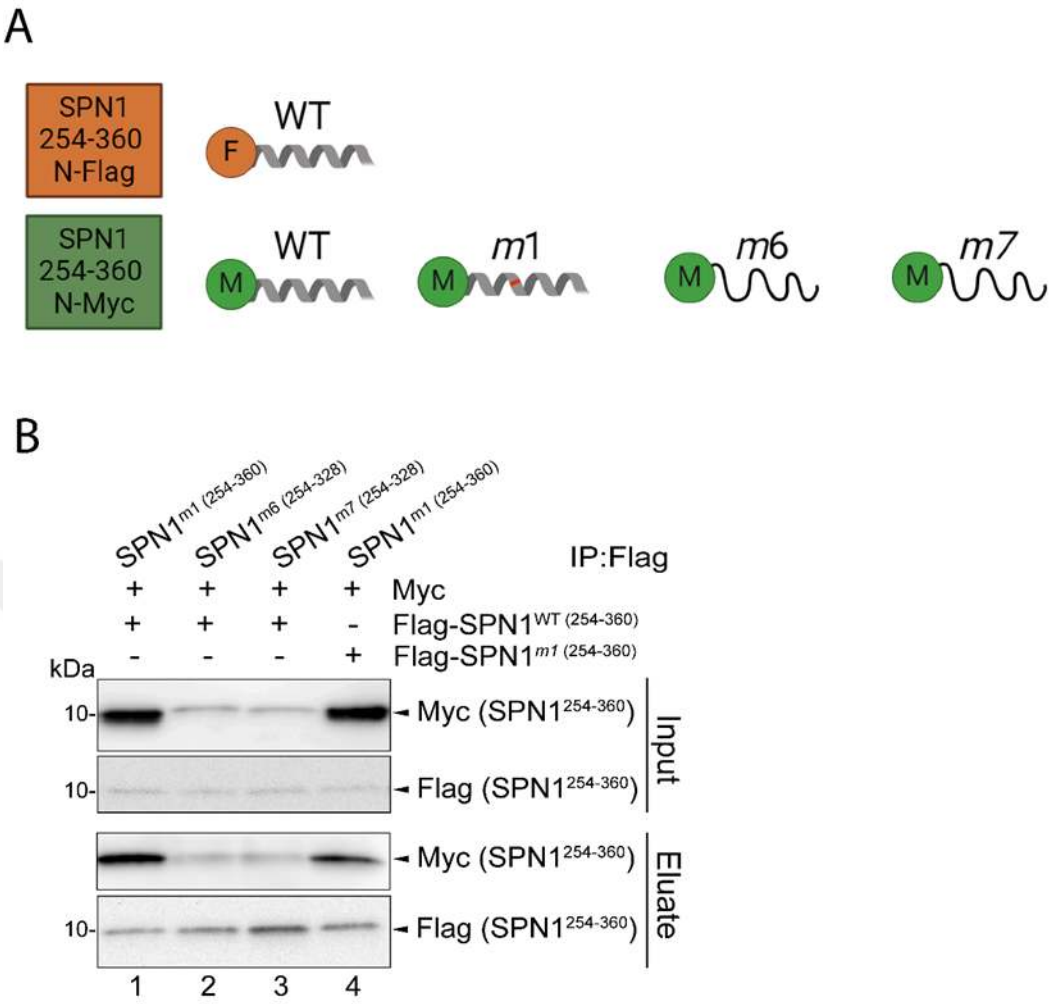

Figure 3.21: SPN1 mutants affect the interaction between the C-terminal domain of the protein. A) Schematic illustration of the structure of the mutant SPN1 254-360 (C-terminus) fragments tagged with Flag- (orange) or Myc (green)-epitope tag. SPN1$^{m1\ (254\text{-}360)}$ fragments keep their alpha-helix structure, yet they have a change (in red) in the structure. On the other hand, SPN1$^{m6\ (254\text{-}328)}$ and SPN1$^{m7\ (254\text{-}328)}$ constructs completely lost the alpha-helix structure. B) HEK293T cells co-transfected with Flag- and Myc- epitope-tagged SPN1 C-terminus fragments. Anti-Myc (top) and anti-Flag (bottom) antibodies used to blot input and eluate samples, revealed the interaction between Myc-SPN1$^{WT(254\text{-}360)}$ and missense mutant (Flag-SPN1$^{m1(254\text{-}360)}$) (Lane 1).this interaction disappeared in fragments with frameshift mutation (Flag-SPN1$^{m6(254\text{-}328)}$ and Flag-SPN1$^{m7(254\text{-}328)}$) (Lane 2 and 3). Also, self-interaction appears to be reduced in missense co-transfected lysate (SPN1$^{m1\ (254\text{-}360)}$)) (Lane 4).

Chapter 4:
CONCLUSION

SPN1 protein plays a crucial role in the biogenesis of spliceosomal UsnRNPs by facilitating their nuclear transport for maturation. However, despite its pivotal role and interactions with proteins necessary for spliceosome formation, this protein has not yet been fully characterized. In the course of our investigation, we identified pathogenic mutations in SPN1 protein in eighteen individuals exhibiting muscular dystrophy features. In total, nine variants were identified, eight of which are located in the uncharacterized C-terminus region of SPN1, indicating a mutational hotspot. Furthermore, the severity of muscular dystrophy in patients, ranging from mild to severe, aligns with the severity of the mutational consequence of the SPN1 variants at the C-terminus region. Individuals with homozygous missense pathogenic variants at the C-terminal show milder phenotypes, whereas patients with mutations altering completely the C-terminus present severe muscle atrophy. Thus, investigating the role of SPN1's C-terminus in modulating the protein's conformation and subsequent interactions was essential to understanding its role in the disease.

This study, for the first time in the literature, demonstrates the essential role of the uncharacterized C-terminus region of SPN1 in its self-oligomerization. Moreover, we show that the C-terminus of the wild-type SPN1 alone is sufficient for the formation of a higher homo-tetrameric complex, a process impaired or abolished in the SPN1 mutants.

Interestingly our findings are in line with the literature. Indeed, a large proportion of proteins in eukaryotic cells are part of oligomeric protein complexes, constituting approximately 30% of all proteins in the cell (Kumari & Yadav, 2019). Protein self-oligomerization is a significant event that provides proteins with evolutionary advantages, including the development of new structural and functional properties, the emergence of new regulatory mechanisms, and the establishment of higher-order complexity for diverse functions (Ali & Imperiali, 2005; Gotte & Menegazzi, 2023). Moreover, the most prevalent oligomeric form of proteins is homo-oligomers, more specifically, homo-tetramers (Goodsell & Olson, 2000; Gotte & Menegazzi, 2023; Kumari & Yadav, 2019). In our western blot analysis, the consistent detection of SPN1 higher bands around 100

kDa and 150 kDa in control samples led us to speculate about the presence of SPN1 homo-dimer and -tetramer. Moreover, the decrease or absence of the band at 150 kDa in patients with a missense mutation (F1$^{m1/m1}$) or compound heterozygous (F2$^{m2/m3}$) and frameshift (F4II-2$^{m6/m6}$) mutations, respectively, confirmed the specificity of this band. These data, combined with the previously described intermolecular interaction between SPN1 C- and N-terminus (Ospina et al., 2005) and the *in silico* prediction of the C-terminal alpha helix, provided strong evidence that this protein exists in a higher-order complex.

The specificity of the 150 kDa homo-tetrameric band was later confirmed by in vitro analysis of transient overexpression of WT and mutant SPN1 constructs, aligning precisely with the data observed in patient samples. Strikingly, our study also demonstrates that the alpha helix at the C-terminus region is essential and sufficient for the formation of SPN1 homo-tetramer (Figure 4.1A). The frameshift variants (*m*6: SPN1^{Tyr301Cysfs*29} and *m*7 SPN1^{Asp300Valfs*30}) which led to the most severe deficiencies in oligomerization, were identified in patients displaying the most extreme muscular dystrophy phenotypes. Conversely, patients with milder symptoms carried missense mutation (*m*1: SPN1^{Ile309Ser}) characterized by relatively minor self-oligomerization disruptions. Hence, there appears to be a correlation between the extent of alteration in the C-terminus of SPN1, its biological function, and the clinical features observed in patients.

On the other hand, the 100 kDa band, which we speculate might correspond to a homo-dimer, did not exhibit any systematic decrease and provided inconclusive results throughout our analyses. Combining our experimental data, which does not demonstrate a significant systematic decrease of the homo-dimeric band in mutant samples, with Matera's findings, we hypothesize that the 100 kDa band may result from the interaction of the N- and C-termini to form SPN1 homo-dimer. Meanwhile, the interaction of C-termini is likely to play a role in the formation of the homo-tetramer (Ospina et al., 2005).

Notably, similar observations have been reported by Niba et al. (2022), in which they associate the degree of disruption of the oligomerization of mutated SMN protein with the severity of spinal muscular atrophy (SMA) (Niba et al., 2022). We next hypothesize that the oligomerization defect in SPN1 mutants may lead to impaired nuclear transport of the UsnRNPs. Indeed, other members of our laboratory have demonstrated retention of Sm proteins in the cytoplasm of several patient's fibroblasts lines (Nashabat et al., 2024). Lastly, we have also shown improper spliceosome maturation in the patient's cells. Taken together, these defects support the muscular dystrophy phenotype observed in patients.

Interactions between SMN and SPN1 proteins are well-established. These proteins known to function in the UsnRNP import complex have been reported to indirectly bind to each other through snRNPs (Narayanan et al., 2002). Given the close association between SMN and SPN1, along with their ability to form higher complexes, future studies focusing on the hetero-oligomerization of these proteins could offer crucial insights into the development and progression of muscular dystrophies.

It is well established that the spliceosome complex plays a crucial role in cellular processes. Interestingly, it many other proteins within this complex are also implicated in Mendelian diseases. For instance, UsnRNPs have been linked to various conditions such as retinitis pigmentosa, neurodegeneration, and cancer (Griffin & Saint-Jeannet, 2020). Additionally, proteins within the Sm core domain have been associated with cerebrocostomandibular syndrome, and notable proteins like SMN and Gemin 5 are implicated in neurodegenerative disorder (Vazquez-Arango & O'Reilly, 2018). Thus, delving deeper into the role of proteins within the snRNP complex such as Importinβ-1 (Impβ-1), which have not yet been linked to hereditary disorders could offer new invaluable insights.

While our study strongly demonstrates the involvement of the C-terminal in homo-tetramer formation, it still lacks a clear explanation for its homo-dimerization. Further studies delving deeper into the interaction between the N- and C-terminus of SPN1 may help address this question and provide additional insights into the biological function of the C-terminus. Interestingly, in the database PhosphoSitePlus, which contains

experimentally verified post-translational modifications (PTMs) of proteins, it has been shown that the SPN1 C-terminus is highly susceptible to phosphorylation (Hornbeck et al., 2019). Indeed, 8 residues have been found phosphorylated between residues 300 to 350. Considering the dynamic nature of the C-terminus alpha helix, we speculate that changes in the phosphorylation state of these residues may alter its structural conformation or modify its function. Thus, in the future, it will be interesting to assess the impact of PTMs in the C-terminus region on the interaction and cellular functions of SPN1.

In conclusion, this thesis marks the first establishment of the self-oligomerization capability of SPN1 monomers through coiled-coil interactions of the C-terminus alpha helices. Our findings strongly suggest that a higher complex of SPN1 is necessary for its cellular role. Further investigation of the C-terminus holds promise for fully uncovering the role of SPN1, thereby providing a better understanding of the development and progression of muscular dystrophies.

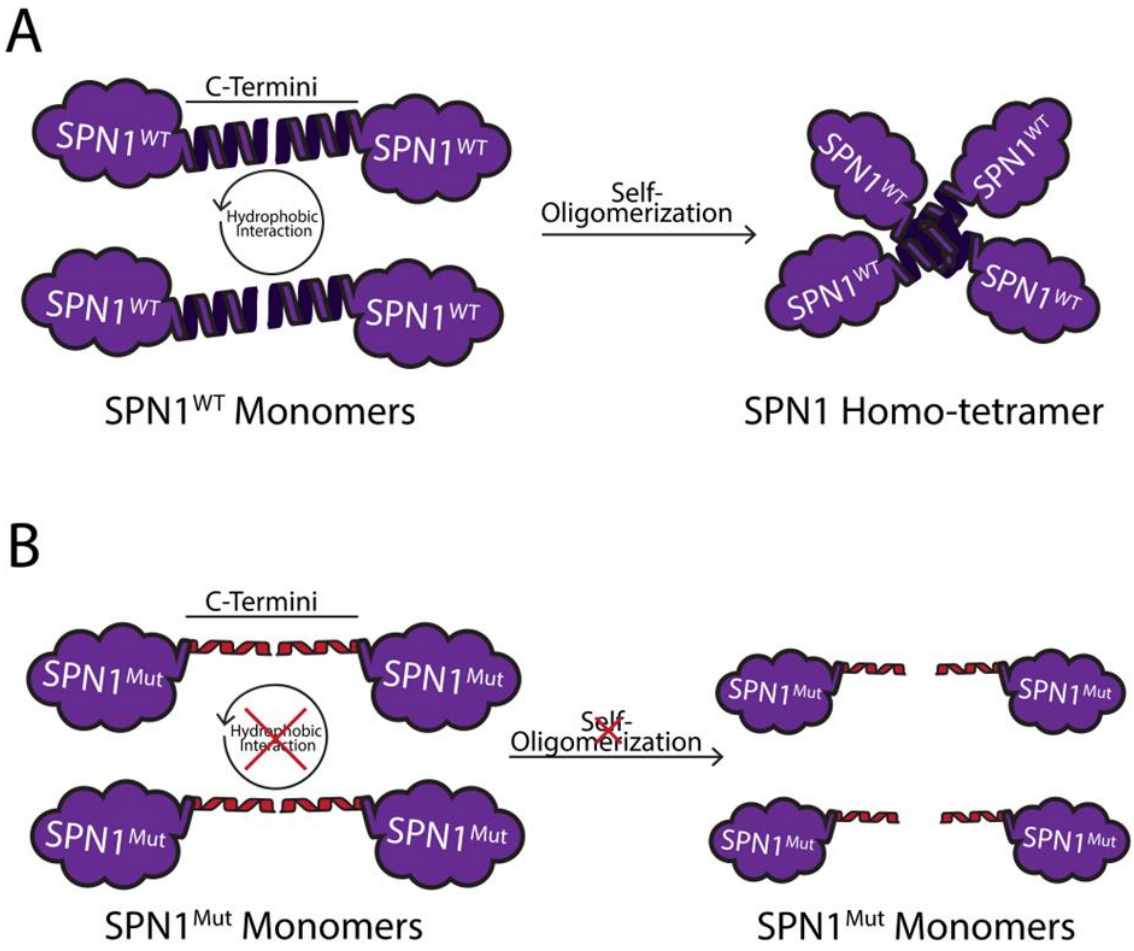

Figure 4.1: Illustration of the SPN1 homo-tetramerization. A) The diagram illustrates the self-oligomerization hypothesis of wild-type (WT) SPN1 monomers (in purple) through hydrophobic interactions of the C-terminal alpha helices. B) The diagram illustrates the impairment of the self-oligomerization of the mutant SPN1 proteins, due to disruption of C-terminus region (in red).

BIBLIOGRAPHY

Ali, M. H., & Imperiali, B. (2005). Protein oligomerization: how and why. *Bioorganic & Medicinal Chemistry*, *13*(17), 5013–5020. https://doi.org/10.1016/J.BMC.2005.05.037

Blakesley, R. W. (1993). Cycle sequencing. *Methods in Molecular Biology (Clifton, N.J.)*, *23*, 209–217. https://doi.org/10.1385/0-89603-248-5:209

Cappi, C., Brentani, H., Lima, L., Sanders, S. J., Zai, G., Diniz, B. J., Reis, V. N. S., Hounie, A. G., Conceição Do Rosário, M., Mariani, D., Requena, G. L., Puga, R., Souza-Duran, F. L., Shavitt, R. G., Pauls, D. L., Miguel, E. C., & Fernandez, T. V. (2016). Whole-exome sequencing in obsessive-compulsive disorder identifies rare mutations in immunological and neurodevelopmental pathways. *Translational Psychiatry*, *6*(3). https://doi.org/10.1038/TP.2016.30

Creamer, T. P., & Rose, G. D. (1995). Interactions between hydrophobic side chains within α-helices. *Protein Science*, *4*(7), 1305–1314. https://doi.org/10.1002/pro.5560040706

Ding, B., & Sepehrimanesh, M. (2021). Nucleocytoplasmic transport: Regulatory mechanisms and the implications in neurodegeneration. In *International Journal of Molecular Sciences* (Vol. 22, Issue 8). Int J Mol Sci. https://doi.org/10.3390/ijms22084165

Emery, A. E. H. (2002). The muscular dystrophies. *Lancet*, *359*(9307), 687–695. https://doi.org/10.1016/S0140-6736(02)07815-7

Faustino, N. A., & Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes & Development*, *17*(4), 419–437. https://doi.org/10.1101/GAD.1048803

Fischer, U., Englbrecht, C., & Chari, A. (2011). Biogenesis of spliceosomal small nuclear ribonucleoproteins. In *Wiley Interdisciplinary Reviews: RNA* (Vol. 2, Issue 5, pp. 718–731). Wiley Interdiscip Rev RNA. https://doi.org/10.1002/wrna.87

Goodsell, D. S., & Olson, A. J. (2000). Structural symmetry and protein function.

Annual Review of Biophysics and Biomolecular Structure, *29*, 105–153. https://doi.org/10.1146/ANNUREV.BIOPHYS.29.1.105

Görlich, D., & Mattaj, I. W. (1996). Nucleocytoplasmic transport. *Science*, *271*(5255), 1513–1518. https://doi.org/10.1126/science.271.5255.1513

Gotte, G., & Menegazzi, M. (2023). Protein Oligomerization. *International Journal of Molecular Sciences*, *24*(13). https://doi.org/10.3390/IJMS241310648

Griffin, C., & Saint-Jeannet, J. P. (2020). Spliceosomopathies: Diseases and mechanisms. *Developmental Dynamics*, *249*(9), 1038–1046. https://doi.org/10.1002/DVDY.214

Gruss, O. J., Meduri, R., Schilling, M., & Fischer, U. (2017). UsnRNP biogenesis: mechanisms and regulation. In *Chromosoma* (Vol. 126, Issue 5, pp. 577–593). Springer. https://doi.org/10.1007/s00412-017-0637-6

Haimov, B., & Srebnik, S. (2016). A closer look into the α-helix basin. *Scientific Reports*, *6*. https://doi.org/10.1038/srep38341

Hornbeck, P. V., Kornhauser, J. M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B., & Gnad, F. (2019). 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Research*, *47*(D1), D433–D441. https://doi.org/10.1093/NAR/GKY1159

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., & Lührmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO Journal*, *17*(14), 4114–4126. https://doi.org/10.1093/emboj/17.14.4114

Karijolich, J., & Yu, Y. T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biology*, *7*(2), 192–204. https://doi.org/10.4161/RNA.7.2.11207

Kiss, T. (2004). Biogenesis of small nuclear RNPs. *Journal of Cell Science*, *117*(Pt 25), 5949–5951. https://doi.org/10.1242/JCS.01487

Kumari, N., & Yadav, S. (2019). Modulation of protein oligomerization: An overview. In *Progress in Biophysics and Molecular Biology* (Vol. 149, pp. 99–113). Prog Biophys Mol Biol. https://doi.org/10.1016/j.pbiomolbio.2019.03.003

Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., & Lu, M. (2006). A seven-helix coiled coil. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(42), 15457–15462. https://doi.org/10.1073/PNAS.0604871103

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, *579*(9), 1900–1903. https://doi.org/10.1016/J.FEBSLET.2005.02.047

Lott, K., & Cingolani, G. (2011). The importin β binding domain as a master regulator of nucleocytoplasmic transport. *Biochimica et Biophysica Acta*, *1813*(9), 1578–1592. https://doi.org/10.1016/J.BBAMCR.2010.10.012

Love, S. L., Emerson, J. D., Koide, K., & Hoskins, A. A. (2023). Pre-mRNA splicing-associated diseases and therapies. *RNA Biology*, *20*(1), 525–538. https://doi.org/10.1080/15476286.2023.2239601

Luo, Y., Wang, Z., Tian, L., & Li, X. (2013). The function of importin β1 is conserved in eukaryotes but the substrates may vary in organisms. *Plant Signaling and Behavior*, *8*(8). https://doi.org/10.4161/psb.25106

Lupas, A. N., Bassler, J., & Dunin-Horkawicz, S. (2017). The Structure and Topology of α-Helical Coiled Coils. *Fibrous Proteins: Structures and Mechanisms*, *82*, 95. https://doi.org/10.1007/978-3-319-49674-0_4

Machyna, M., Heyn, P., & Neugebauer, K. M. (2013). Cajal bodies: where form meets function. *Wiley Interdisciplinary Reviews. RNA*, *4*(1), 17–34. https://doi.org/10.1002/WRNA.1139

Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews. Molecular Cell Biology*, *8*(3), 209–220. https://doi.org/10.1038/NRM2124

Mercuri, E., & Muntoni, F. (2013). Muscular dystrophies. *Lancet (London, England)*, *381*(9869), 845–860. https://doi.org/10.1016/S0140-6736(12)61897-2

Mitrousis, G., Olia, A. S., Walker-Kopp, N., & Cingolani, G. (2008). Molecular basis for the recognition of snurportin 1 by importin. *Journal of Biological Chemistry*, *283*(12), 7877–7884. https://doi.org/10.1074/jbc.M709093200

Narayanan, U., Ospina, J. K., Frey, M. R., Hebert, M. D., & Gregory Matera, A. (2002). SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Human Molecular Genetics*, *11*(15), 1785–1795. https://doi.org/10.1093/HMG/11.15.1785

Nashabat, M., Nabavizadeh, N., Saraçoğlu, H. P., Sarıbaş, B., Avcı, Ş., Börklü, E., Beillard, E., Yılmaz, E., Uygur, S. E., Kayhan, C. K., Bosco, L., Eren, Z. B., Steindl, K., Richter, M. F., Bademci, G., Rauch, A., Fattahi, Z., Valentino, M. L., Connolly, A. M., … Escande-Beillard, N. (2024). SNUPN deficiency causes a recessive muscular dystrophy due to RNA mis-splicing and ECM dysregulation. *Nature Communications*, *15*(1), 1758. https://doi.org/10.1038/s41467-024-45933-5

Niba, E. T. E., Nishio, H., Wijaya, Y. O. S., Ar Rochmah, M., Takarada, T., Takeuchi, A., Kimizu, T., Okamoto, K., Saito, T., Awano, H., Takeshima, Y., & Shinohara, M. (2022). Stability and Oligomerization of Mutated SMN Protein Determine Clinical Severity of Spinal Muscular Atrophy. *Genes*, *13*(2). https://doi.org/10.3390/GENES13020205

Ospina, J. K., Gonsalvez, G. B., Bednenko, J., Darzynkiewicz, E., Gerace, L., & Matera, A. G. (2005). Cross-talk between snurportin1 subdomains. *Molecular Biology of the Cell*, *16*(10), 4660–4671. https://doi.org/10.1091/mbc.E05-04-0316

Piecyk, K., Niedzwiecka, A., Ferenc-Mrozek, A., Lukaszewicz, M., Darzynkiewicz, E., & Jankowska-Anyszka, M. (2015). How to find the optimal partner--studies of snurportin 1 interactions with U snRNA 5’ TMG-cap analogues containing modified 2-amino group of 7-methylguanosine. *Bioorganic & Medicinal Chemistry*, *23*(15), 4660–4668. https://doi.org/10.1016/J.BMC.2015.05.054

Pistoni, M., Ghigna, C., & Gabellini, D. (2010). Alternative splicing and muscular

dystrophy. *RNA Biology*, *7*(4), 441–452. https://doi.org/10.4161/RNA.7.4.12258

Platt, A. R., Woodhall, R. W., & George, A. L. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques*, *43*(1), 58–62. https://doi.org/10.2144/000112499

Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. *Genetics*, *190*(3), 841–854. https://doi.org/10.1534/GENETICS.111.136911

Raimer, A. C., Gray, K. M., & Matera, A. G. (2017). SMN - A chaperone for nuclear RNP social occasions? In *RNA Biology* (Vol. 14, Issue 6, pp. 701–711). RNA Biol. https://doi.org/10.1080/15476286.2016.1236168

Sawyer, I. A., Sturgill, D., Sung, M. H., Hager, G. L., & Dundr, M. (2016). Cajal body function in genome organization and transcriptome diversity. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *38*(12), 1197–1208. https://doi.org/10.1002/BIES.201600144

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews. Genetics*, *17*(1), 19–32. https://doi.org/10.1038/NRG.2015.3

Sefidkar, N., Fathizadeh, S., Nemati, F., & Simserides, C. (2022). Energy Transport along α-Helix Protein Chains: External Drives and Multifractal Analysis. *Materials*, *15*(8). https://doi.org/10.3390/ma15082779

Sleeman, J. E., & Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Current Biology : CB*, *9*(19), 1065–1074. https://doi.org/10.1016/S0960-9822(99)80475-8

Sobreira, N., Schiettecatte, F., Valle, D., & Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human Mutation*, *36*(10), 928–930. https://doi.org/10.1002/HUMU.22844

Strasser, A., Dickmanns, A., Lührmann, R., & Ficner, R. (2005). Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *The EMBO Journal*, *24*(13), 2235–2243. https://doi.org/10.1038/SJ.EMBOJ.7600701

Strasser, A., Dickmanns, A., Schmidt, U., Penka, E., Urlaub, H., Sekine, M., Lührmann, R., & Ficner, R. (2004). Purification, crystallization and preliminary crystallographic data of the m3G cap-binding domain of human snRNP import factor snurportin 1. *Acta Crystallographica Section D: Biological Crystallography*, *60*(9), 1628–1631. https://doi.org/10.1107/S0907444904015380

Urlaub, H., Raker, V. A., Kostka, S., & Lührmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO Journal*, *20*(1–2), 187–196. https://doi.org/10.1093/emboj/20.1.187

Vazquez-Arango, P., & O'Reilly, D. (2018). Variant snRNPs: New players within the spliceosome system. *RNA Biology*, *15*(1), 17. https://doi.org/10.1080/15476286.2017.1373238

Wang, G. S., & Cooper, T. A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews. Genetics*, *8*(10), 749–761. https://doi.org/10.1038/NRG2164

Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*, *89*, 359–388. https://doi.org/10.1146/ANNUREV-BIOCHEM-091719-064225

Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, *3*(7), 1–2. https://doi.org/10.1101/CSHPERSPECT.A003707

Yakimov, A. P., Afanaseva, A. S., Khodorkovskiy, M. A., & Petukhov, M. G. (2016). Design of Stable α-Helical Peptides and Thermostable Proteins in Biotechnology and Biomedicine. *Acta Naturae*, *8*(4), 70. https://doi.org/10.32607/20758251-2016-8-4-70-81

Yang, Y., Guo, L., Chen, L., Gong, B., Jia, D., & Sun, Q. (2023). Nuclear transport proteins: structure, function, and disease relevance. In *Signal Transduction and Targeted Therapy* (Vol. 8, Issue 1, pp. 1–29). Nature Publishing Group. https://doi.org/10.1038/s41392-023-01649-4

Zhang, S. Q., Kulp, D. W., Schramm, C. A., Mravic, M., Samish, I., & Degrado, W. F. (2015). The membrane- and soluble-protein helix-helix interactome: Similar geometry via different interactions. *Structure*, *23*(3), 527–541. https://doi.org/10.1016/j.str.2015.01.009

Appendix A: …..