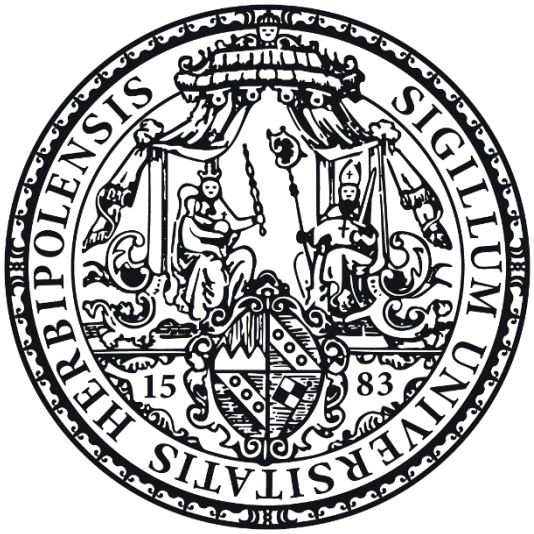


Biochemical and structural characterisation of modules within the SMN complex

Biochemische und strukturelle Charakterisierung von Modulen des SMN-Komplexes

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Biomedicine

submitted by

Aravindan Viswanathan
from Mumbai, India

Würzburg, 2019

Submitted on:

Members of the *Promotionskomitee*:
Chairperson:

Primary Supervisor
Prof. Dr. Utz Fischer
Department of Biochemistry, Biocenter
University of Würzburg,
Germany

Supervisor (Second)
Prof. Dr. Christoph Sotriffer
Institute for Pharmacy and Food Chemistry
University of Würzburg,
Germany

Supervisor (Third)
Prof. Dr. Alexander Buchberger
Department of Biochemistry, Biocenter
University of Würzburg,
Germany

Supervisor (Fourth)
Dr. Clemens Grimm
Department of Biochemistry, Biocenter
University of Würzburg,
Germany

Date of Public Defence: ……

Date of Receipt of Certificates: ……..

தாயிற் சிறந்த தயா ஆன தத்துவனே
மாசற்ற சோதி மலர்ந்த மலர்ச்சுடரே
தேசனே தேன் ஆர்அமுதே சிவபுரானே
பேராது நின்ற பெருங்கருணைப் போராறே
ஆரா அமுதே அளவிலாப் பெம்மானே
நீராய் உருக்கி என் ஆருயிராய் நின்றானே

Table of Contents

Summary……...i

Zusammenfassung……...iii

1 Introduction 1

1.1 Spliceosomal U snRNPs 1

1.1.1 Splicing 2

1.1.2 Molecular architecture of U snRNPs 4

1.2 Sm class U snRNP biogenesis: A segmented pathway 6

1.2.1 The cytosolic assembly phase of Sm class U snRNPs 7

1.2.2 Early cytoplasmic assembly phase 8

1.2.3 Late cytosolic assembly phase 11

1.3 Structural and functional facets of the SMN complex components 13

1.3.1 Survival Motor Neuron (SMN) protein 13

1.3.2 Gemins 2-8 and Unrip 16

2 Thesis Objectives 21

3 Materials and Methods 23

3.1 Materials 23

3.1.1 General Materials 23

3.1.2 Commercial Kits 23

3.1.3 Oligonucleotides 23

3.1.4 Enzymes 26

3.1.5 Bacterial Plasmid Vectors 27

3.1.6 Bacterial Expression Vectors 27

3.1.7 Bacterial Strains 31

3.1.8 Buffers and Solutions 31

3.1.9 Protein Chromatography 36

3.1.10 Protein Crystallization 37
3.1.11 Software and Servers 38
3.2 Methods 38
3.2.1 Molecular Biology Methods 38
3.2.2 Protein Biochemistry Methods 43
3.2.3 X-ray Crystallography 46
4 Results 49
4.1 Introductory notes 49
4.2 The Gemin7:6 sub-complex within the SMN complex 50
4.2.1 The *Homo sapiens* (*Hs*) Gemin7:6 heterodimer 50
4.2.2 Characterisation of the nematode *C. elegans* Gemin7:6 sub-complex 53
4.3 The Gemin8:7:6 sub-complex 55
4.3.1 The nematode Gemin8:7 interaction domains 55
4.3.2 The *H. sapiens* Gemin8:7:6 sub-complex 59
4.3.3 Structural insights into the minimal *Hs* Gemin8:7:6 sub-complex 64
4.3.4 The fission yeast *S. pombe* (*Sp*) Gemin8:7:6 sub-complex 65
4.4 A putative role for the Gemin7:6 sub-complex during the heptameric Sm core assembly 67
4.5 The Gemin8: SMN interaction interface 70
4.5.1 The MBP-YG-box biochemical handle 70
4.5.2 Gemin8's interaction with SMN YG-box module is affected in the presence of SMA patient mutations 83
4.5.3 The fission yeast *S. pombe* (*Sp*) Gemin8: SMN interaction interface 87
5 Discussion 97
5.1 Conservation of the Gemin7:6 module in nematodes and humans 97
5.2 The Gemin7:6 sub-complex tethers to the SMN complex via Gemin8's conserved C-terminal helix 100

5.3 The Gemin7:6 module does not behave as a surrogate Sm-dimer during U snRNP assembly 103
5.4 Conservation of SMN's YG-box domain in humans and nematodes 104
5.5 Influence of SMA patient mutations in Gemin8 incorporation into the SMN complex 105
6 Conclusions and Outlook 111
7 Appendix 115
7.1 Unrip engages with Gemin7's N-terminus 115
7.2 Sequence and structural alignment of SMN orthologues 116
7.3 Vectors pETM41 and pETM41* variations 116
7.4 PDB codes of atomic resolution structures used for scientific illustrations in this work 117
7.5 SMN orthologues UniProtKB accession numbers 117
7.6 Symbols, Units, Acronyms and Abbreviations 118
7.7 Nucleotide bases and Amino acid codes 121
7.8 Individual contributions of personnel towards the structural work presented in this dissertation 122
8 References 123
9 List of publications 131
10 Curriculum Vitae 133
11 Declaration 135

List of Figures

Figure 1.1 Mechanism of pre-mRNA splicing and assembly of spliceosomes on pre-mRNA.3
Figure 1.2 Components of spliceosomal U snRNPs 5
Figure 1.3 Sm proteins and their mode of interaction 6
Figure 1.4 Assembly of U snRNPs 8
Figure 1.5 Trans-acting factors of the early cytoplasmic U snRNP assembly pathway 9
Figure 1.6 Structure of the kinetically trapped 6S complex 10
Figure 1.7 Schematic representation of the multimeric SMN complex 11
Figure 1.8 An overview of the Sm-class U snRNP biogenesis pathway and associated key regulators 12
Figure 1.9 A consensus interaction map of the Survival Motor Neuron (SMN) complex 13
Figure 1.10 Domain organisation of Human Survival Motor Neuron (SMN) protein 14
Figure 1.11 Structural models of SMN complex components 18
Figure 3.1 Poly-cistronic cloning strategy 41
Figure 3.2 Generation of pGEX6P1 vector 43
Figure 4.1 Conservation of the multi-subunit SMN complex in *H. sapiens* and *C. elegans* . 50
Figure 4.2 Recapitulating the *H. sapiens* Gemin7:6 sub-complex..... 51
Figure 4.3 *H. sapiens Hs* Gemin7:6 sub-complex structure 53
Figure 4.4 Conserved regions of the putative *C. elegans* Gemin6 and 7 orthologues..... 54
Figure 4.5 Reconstitution of the *C. elegans* Gemin7:6 dimeric complex 55
Figure 4.6 *C. elegans* Gemin8:7:6 sub-complex..... 56
Figure 4.7 Mapping of the *C. elegans* Gemin8:7 interaction interface..... 58
Figure 4.8 The *H. sapiens* (*Hs*) Gemin8:7 interface..... 61
Figure 4.9 Delineation of Gemin7's Gemin8 binding domain 63
Figure 4.10 Molecular architecture of *Hs* Gemin8:7:6 sub-complex..... 64
Figure 4.11 The Gemin8:7 interaction interface..... 65
Figure 4.12 A minimal *S. pombe* Gemin8:7:6 trimeric complex 66
Figure 4.13 Gemin7:6 sub-complex does not engage with Sm protein hetero-oligomers.... 69
Figure 4.14 Structure of the *Hs* SMN YG-box domain 71
Figure 4.15 The *Hs* MBP-YG box biochemical handle..... 73
Figure 4.16 Structural conservation of SMN YG-box module in *C. elegans* (*Ce*) 75
Figure 4.17 Effects of SMA patient mutations on MBP-*Hs* $YG^{252-284}$ oligomerisation 79
Figure 4.18 Effects of SMA patient mutations on MBP-*Ce* $YG^{169-207}$ oligomerisation..... 82
Figure 4.19 Gemin8 binding to SMN is affected by SMA patient mutations 86

Figure 4.20 The minimal *S. pombe (Sp)* SMNΔL variant 88
Figure 4.21 N-terminal truncation variants of the *S. pombe (Sp)* Gemin8ΔL:7:6 sub-complex 90
Figure 4.22 Reconstitution of the tetrameric *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex 93
Figure 5.1 Sequence conservation of Gemin6 orthologues in diverse organisms 99
Figure 5.2 Sequence conservation of Gemin7 orthologues in diverse organisms 100
Figure 5.3 Conservation of Gemin8:7 interaction interface in diverse organisms 102
Figure 5.4 SMN Oligomerisation: A structural model 107
Figure 6.1 Gemin8 sculpts the modular architecture of the multimeric SMN complex 112
Figure 7.1 Unrip engages with the SMN complex via Gemin7 115
Figure 7.2 Conservation of domains in SMN orthologues 116
Figure 7.3 Alignment of pETM41 vector variants 116

List of Tables

Table 4.1: Oligomeric status of MBP-*Hs* $YG^{252-284}$ and MBP-*Ce* $YG^{169-207}$ harbouring SMA patient mutations 83
Table 4.2 Crystallographic data collection statistics for *Hs* $Gemin7^{31-131}$:6 (fl) (Section 4.2.1), *Hs* $Gemin8^{190-230}$:7^{46-131}:6^{1-92} (Section 4.3.3) and *Ce* MBP*-$YG^{182-207}$ (Section 4.5.1.2) .. 94
Table 4.3 Crystallographic refinement statistics for *Hs* $Gemin7^{31-131}$:6 (fl) (Section 4.2.1), *Hs* $Gemin8^{190-230}$:7^{46-131}:6^{1-92} (Section 4.3.3) and *Ce* MBP*-$YG^{182-207}$ (Section 4.5.1.2) 95
Table 5.1 SMN's YG-box SMA variants influencing Gemin8 binding 108

Summary

Cellular proteome profiling revealed that most biomolecules do not exist in isolation, but rather are incorporated into modular complexes. These assembled complexes are usually very large, consisting of 10 subunits on an average and include either proteins alone, or proteins and nucleic acids. Consequently, such macromolecular assemblies rather than individual biopolymers perform the vast majority of cellular activities. The faithful assembly of such molecular assemblies is often aided by trans-acting factors *in vivo*, to preclude aggregation of complex components and/or non-cognate interactions. A paradigm for an assisted assembly of a macromolecular machine is the formation of the common Sm/LSm core of spliceosomal and histone-mRNA processing U snRNPs. The key assembly factors united in the Protein Arginine Methyltransferase 5 (PRMT5) and the Survival Motor Neuron (SMN) complexes orchestrate the assembly of the Sm/LSm core on the U snRNAs. Assembly is initiated by the PRMT5-complex subunit pICln, which pre-arranges the Sm/LSm proteins into spatial positions occupied in the mature U snRNPs. The SMN complex subsequently binds these Sm/LSm units, displaces pICln and catalyses the Sm ring closure on the Sm-site of the U snRNA.

The SMN complex consists of the eponoymous SMN protein linked in a modular network of interactions with eight other proteins, termed Gemins 2-8 and Unrip. Despite functional and structural characterisation of individual protein components and/or sub-complexes of this assembly machinery, coherent understanding of the structural framework of the core SMN complex remained elusive. The current work, employing a combined approach of biochemical and structural studies, aimed to contribute to the understanding of how distinct modules within the SMN complex coalecse to form the macromolecular SMN complex.

A novel atomic resolution (1.5 Å) structure of the human Gemin8:7:6 sub-complex, illustrates how the peripheral Gemin7:6 module is tethered to the SMN complex via Gemin8's C-terminus. In this model, Gemin7 engages with both Gemin6 and Gemin8 via the N- and C-termini of its Sm-fold like domain. This highly conserved interaction mode is reflected in the pronounced sequence conservation and identical biochemical behaviour of similar sub-complexes from divergent species, namely *S. pombe* and *C. elegans*.

Despite lacking significant sequence similarity to the Sm proteins, the dimeric Gemin7:6 complex share structural resemblance to the Sm heteromers. The hypothesis that the dimeric Gemin7:6 functions as a Sm-surrogate during Sm core assembly could not be confirmed in this work. The functional relevance of the structural mimicry of the dimeric Gemin7:6 sub-complex with the Sm heterodimers therefore still remains unclear.

Reduced levels of functional SMN protein is the cause of the devastating neurodegenerative disease, Spinal Muscular Atrophy (SMA). The C-terminal YG-zipper motif of SMN is a major hot-spot for most SMA patient mutations. In this work, adding to the existing inventory of the human and fission yeast YG-box models, a novel 2.2 Å crystal structure of the nematode SMN's YG-box domain adopting the glycine zipper motif has been reported. Furthermore, it could be assessed that SMA patient mutations mapping to this YG-box domain greatly influences SMN's self-association competency, a property reflected in both the human and nematode YG-box biochemical handles. The shared molecular architecture and biochemical behaviour of the nematode SMN YG-box domain with its human and fission yeast counterparts, reiterates the pronounced conservation of this oligomerisation motif across divergent organisms.

Apart from serving as a multimerization domain, SMN's YG-box also acts as interaction platform for Gemin8. A systematic investigation of SMA causing missense mutations uncovered that Gemin8's incorporation into the SMN complex is influenced by the presence of certain SMA patient mutations, albeit independent of SMN's oligomerisation status. Consequently, loss of Gemin8 association in the presence of SMA patient mutations would also affect the incorporation of Gemin7:6 sub-complex. Gemin8, therefore sculpts the heteromeric SMN complex by bridging the Gemin7:6 and SMN:Gemin2 sub-units, a modular feature shared in both the human and nematode SMN complexes.

These findings provide an important foundation and a prospective structural framework for elucidating the core architecture of the SMN complex in the ongoing Cryo-EM studies.

Zusammenfassung

Systematische Untersuchungen von zellulären Bestandteilen haben gezeigt, dass viele Proteine nicht isoliert, sondern vielmehr in modularen Komplexen organisiert vorliegen. Mit durchschnittlich zehn Untereinheiten sind diese Komplexe sehr groß, wobei sie entweder ausschließlich aus Proteinen oder aber aus Proteinen und Nukleinsäuren bestehen können. Daher wird der Großteil zellulärer Aktivitäten nicht von einzelnen Biopolymeren, sondern von makromolekularen Komplexen verrichtet. Die Zusammenlagerung dieser Komplexe wird *in vivo* häufig von Hilfsfaktoren unterstützt, um die Aggregation der Einzelkomponenten und/oder unspezifische Wechselwirkungen zu verhindern. Ein Beispiel für eine derartige Zusammenlagerungshilfe ist die Bildung des Sm/LSm-Cores der mRNA-prozessierenden U snRNPs. Dabei wird die Anlagerung von Sm/LSm Proteinen an die U snRNAs durch eine Anzahl von Hilfsfaktoren orchestriert, die in Protein-Arginin-Methyltransferase 5 (PRMT5)- und dem Survival Motor Neuron (SMN)-Komplexen organisiert sind. Die Zusammenlagerung wird durch die PRMT5-Untereinheit pICln initiiert, die die räumliche Anordnung von Sm/LSm-Proteinen in höher-geordneten Komplexen stabilisiert. Diese werden anschließend auf den SMN-Komplex übertragen, wobei pICln verdrängt und die Verbindung mit der Sm-Seite der U snRNA sichergestellt wird.

Der SMN-Komplex besteht aus dem SMN-Protein, das in einem modularen Netzwerk mit acht weiteren Proteinen (Gemins 2-8 und Unrip) interagiert. Auch wenn funktionale und strukturelle Charakterisierungen einzelner Proteinkomponenten und Module dieser Zusammenlagerungs-Maschinerie vorliegen, steht ein tiefergehendes Verständnis des strukturellen Organisation des Gesamt-Komplexes noch aus. In der vorliegenden Arbeit sollte unter Anwendung biochemischer und struktureller Techniken ein Beitrag dazu geleistet werden, die Interaktionen der verschiedenen Komponenten innerhalb des SMN-Komplexes zu verstehen, die so die dreidimensionale Organisation des SMN-Komplexes zu verstehen.

Eine neuartige Kristallstruktur des humanen Gemin8:7:6-Subkomplexes bei einer Auflösung von 1.5 Å zeigt, wie der periphere Gemin7:6-Abschnitt durch den C-Terminus von Gemin8 zum SMN-Komplex dirigiert wird. In diesem Modell interagiert Gemin7 sowohl mit Gemin6 als auch Gemin8 über den N- und C-Terminus der Sm-ähnlichen Domäne.

Dieser hochkonservierte Interaktionsmodus wird in der erwähnten konservierten Sequenz und dem gleichen biochemischen Verhalten ähnlicher Subkomplexe in divergenten Spezies einschließlich *S. pombe* und *C. elegans* widergespiegelt. Obwohl es keine signifikante Übereinstimmung mit der Sequenz von Sm-Proteinen gibt, weist der dimere Gemin7:6-Komplex markante strukturelle Ähnlichkeit mit dem einem Sm-Heterodimer auf. Die Annahme, der dimere Gemin7:6-Subkomplex würde als Hilfsfaktor über die direkte Interaktion mit Sm-Proteinen fungieren konnte in der vorliegenden Arbeit nicht bestätigt werden. Folglich bleibt die Funktion des dimeren Gemin7:6-Subkomplexes im Kontext der SMN-Zusammenlagerungsmaschinerie unklar.

Verringerte Mengen des funktionellen SMN-Proteins sind die Ursache für die neurodegenerative Erkrankung *Spinale Muskelatrophie* (SMA). Das C-terminale YG-Zipper-Motiv von SMN stellt einen Hotspot für die meisten SMA-Mutationen dar. In dieser Arbeit wurde der bereits bekannten YG-Box aus *H. sapiens* und *S. pombe* eine neuartige Kristallstruktur der SMN YG-Box aus *C. elegans* mit einer Auflösung von 2.2 Å hinzugefügt. Zusätzlich wurde gezeigt, dass SMA-verursachende *Missense*-Mutationen in der YG-Box einen beträchtlichen Einfluss auf die Selbst-Interaktion von SMN haben, was aus biochemischen Versuchen mit der YG-Box aus *H. sapiens* und *C. elegans* ersichtlich wurde. Der molekulare Aufbau und das biochemische Verhalten der SMN YG-Box aus *C. elegans*, *S. pombe* und *H. sapiens* betont die Konservierung dieses Oligomerisierungsmotives über mehrere Organismen hinweg.

Neben der Funktion als Multimerisationsdomäne dient die YG-Box von SMN auch als Interaktionsplattform für Gemin8. Eine systematische Untersuchung von SMA-verursachenden *Missense*-Mutationen ergab, dass die Einbindung von Gemin8 in den SMN-Komplex durch definierte Substitutionen massiv beeinflusst wird. Interessanterweise ist dieser Bindungsdefekt unabhängig vom SMN-Oligomerisierungsstatus. Demzufolge würde diese Klasse von SMA-Mutationen spezifisch die Inkorporation des Gemin7:6-Subkomplexes beeinflussen.

Die Resultate dieser Arbeit bilden eine wichtige Grundlage für weitere strukturelle Untersuchungen des SMN-Komplexes über Kryo-Elektronenmikroskopie.

1 Introduction

Most biomolecules do not act in isolation within the cellular microenvironment but are rather assembled to form macromolecular complexes with several layers of modular architecture and function (Gavin et al. 2006; Hartwell et al. 1999). Prominent examples of such macromolecular assemblies include the spliceosomes, ribosomes and exosomes, which accomplish diverse and vital biochemical processes. The colligative properties of most biomolecules dictates the self-assembly of such large complexes via diffusion-driven random collision of subunits *in vitro* (Berg and von Hippel 1985). Conversely within the crowded cellular environment, the propensity for irreversible component aggregations and/or non-cognate interactions is more pronounced. As a consequence, the proper and faithful formation of macromolecular machines is achieved by several assembly strategies that have evolved to cope with macromolecular crowding (Ellis and Minton 2006; Zimmerman and Minton 1993). These cellular strategies are exemplified in the assembly of spliceosomal U snRNPs, wherein a sophisticated spatio-temporal regulation by trans-acting factors ensure an efficient and faithful assembly line (Chari and Fischer 2010).

1.1 Spliceosomal U snRNPs

Eukaryotic genes often are discontinuous with the non-coding intervening sequences (introns) amidst the coding sequences (exons) (Gilbert 1978). An essential step in the expression of these genes is the precise removal of the introns from the precursor-messenger RNAs (pre-mRNA) by splicing. The abundant class of RNA-protein complexes, the uridine-rich small nuclear ribonucleoproteins (U snRNPs), assemble on eukaryotic pre-mRNAs forming active splicing complexes (spliceosomes). Together with additional non-snRNP factors, the U snRNPs catalyse removal of introns and the ligation of exons (splicing) with high fidelity (Rino and Carmo-Fonseca 2009). Two distinct classes of spliceosomes orchestrate the nuclear splicing event in metazoans: while the highly abundant major spliceosomes process the vast majority of the canonical introns (U2-type), the low abundance minor spliceosomes assemble on the rare ATAC class introns (U12-type) mediating their excision. The U12-type minor spliceosome includes the U11, U12, U4atac and U6atac snRNPs which are distinct and yet

functionally analogous to the major spliceosome U1, U2, U4, and U6 snRNPs respectively. Only the U5 snRNP is unique in serving as a component to both the minor and major spliceosomes (Patel and Steitz 2003).

1.1.1 Splicing

The assembly of the spliceosomal U snRNPs on the pre-mRNA depends on the recognition of conserved sequence elements within the introns namely the 5' splice site (SS), branch point sequence (BPS), the poly-pyrimidine tract and 3' SS. The splicing event entails two sequential S_N2 type transesterification reactions (Figure 1.1 A). In the initial step (branching), the 5' exon-intron junction cleavage occurs when the 2' hydroxyl (OH) of a conserved adenosine (A) of the branch point (BP) sequence element attacks the phosphate at the 5' SS resulting in a 5' exon with a 3' OH group and an intron lariat−3' exon intermediate. In the subsequent step ('exon ligation'), the 3' intron-exon junction is cleaved when the new 3'-OH of the 5' exon attacks a phosphate at the 3' SS facilitating the ligation of the 5' and 3' exons with the concomitant release of the intron lariat (Wahl, Will, and Luhrmann 2009). The U12-dependent splicing by the minor spliceosomes is determined by the longer and more tightly constrained consensus sequences at the 5' SS and branch site of minor-class introns, as well as by the lack of a poly-pyrimidine tract upstream of the 3' SS (Sharp and Burge 1997).

The spliceosome is not a preformed enzyme but is rather assembled *de novo* on each intron of the pre-mRNA in a step-wise manner (Wahl, Will, and Luhrmann 2009). For the major class of U2-type introns, the spliceosomal assembly begins with the U1 and U2 snRNPs association with the 5' SS and the BPS respectively, forming the pre-spliceosomal complex. Successively, the U4/U6-U5 tri-snRNP is recruited to the assembling spliceosome and several RNA-RNA and RNA-protein rearrangements with the concomitant displacement of U1 and U4 snRNPs precede the first transesterification reaction. The now activated spliceosomal complex comprised of U2, U6 and U5 tri-snRNPs undergoes additional structural rearrangements prior to the second transesterification reaction. Following exon ligation and release of the intron lariat, the spliceosomal components dissociate and the released U snRNPs are recycled for the next splicing event (Figure 1.1 B) (Will and Luhrmann 2001).

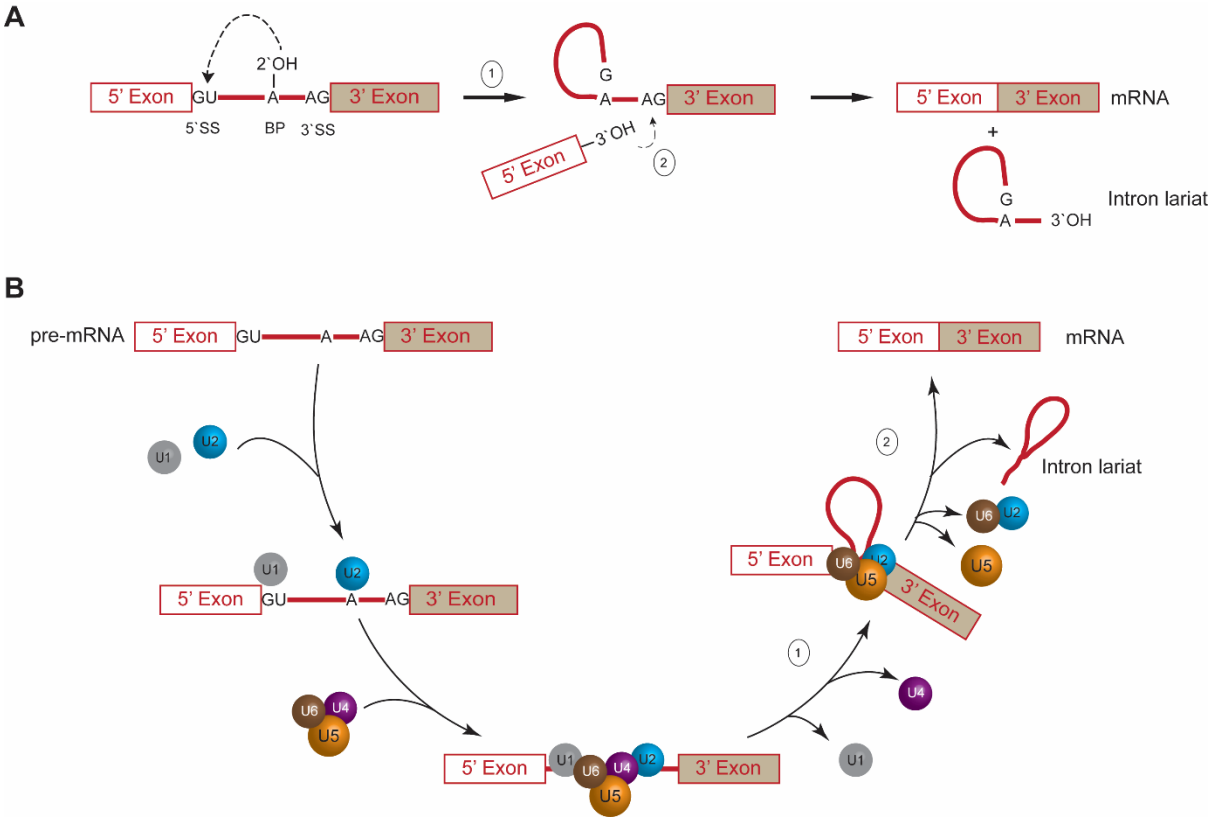


Figure 1.1 Mechanism of pre-mRNA splicing and assembly of spliceosomes on pre-mRNA
(A) Mechanism of U2 type intron splicing. The first transesterification step (1) during splicing involves the nucleophilic attack (dashed line) on the terminal phosphodiester bond of the 5' splice site (5' SS) by the 2' hydroxyl (OH) of the branch point (BP) adenosine (A). In the second splicing step (2), the 3' OH of the upstream exon (exon 1) attacks the 3' splice site (3' SS) culminating in the ligation of exon 1 and exon 2.
(B) Schematic representation of pre-mRNA splicing by the U2-dependent major spliceosomes. pre-mRNA splicing begins co-transcriptionally with the U1 and U2 snRNPs binding to the 5' SS and BP of the U2-type intronic sequences, respectively. Subsequently, the U4/U6-U5 tri-snRNP is recruited for the first splicing step (see A). With the release of U1 and U4 snRNPs, the second splicing step ensues releasing the ligated exons (mRNA) and the intron lariat. With the culmination of each splicing event, the released U snRNPs are recycled for the next round of splicing. For the sake of simplicity, the intermediate U snRNP spliceosomal assembly and/or re-arrangements as well the non-snRNP factors assisting the splicing event have not been depicted.
Image adapted and modified from Will and Lührmann, 2001.

The sequential structural and compositional rearrangements of the assembling spliceosomes on the pre-mRNA dictate the RNA-RNA and RNA-protein remodelling events for the spatial positioning of the pre-mRNA reactive groups for catalysis. The rearrangements of the intricate snRNA: snRNA interaction network configures the U2 and U6 snRNAs to form a RNA catalytic centre co-ordinated by magnesium (Mg^{2+}) ions, making the spliceosome a metallo-ribozyme (Fica et al. 2013; Papasaikas and Valcarcel 2016). Although the basic chemical reaction of splicing is isoenergetic (Figure 1.1 A), the splicing event demands cellular energy consumption to regulate the dynamics and fidelity of spliceosomal assembly, activation and disassembly. For example, several of the key structural transition events driving the interconversion of the spliceosomal complexes are orchestrated by ATP-dependent spliceosomal DExD/H-

type RNA helicases (Staley and Guthrie 1998; Cordin, Hahn, and Beggs 2012). While the major and minor spliceosomes share remarkable mechanistic similarities during their assembly on the pre-mRNA substrates highlighting their common ancestry, primary differences occur during the early steps of the assembly event (Turunen et al. 2013).

1.1.2 Molecular architecture of U snRNPs

Though functionally distinct, the various U snRNPs are compositionally similar. Each U snRNP particle consists of a tight complex between a single uridyl-rich small nuclear RNA (U snRNA) moiety, seven core proteins (Sm B/B', D1, D2, D3, E, F and G) and a variable number of U snRNP specific proteins (Figure 1.2) (Will and Luhrmann 2001). The core Sm proteins were initially identified as major autoantigens in patients with systemic lupus erythematosus (SLE), thus acquiring their name after one of the first patients, Stephanie Smith (Lerner and Steitz 1979). The U snRNAs possess the 'Sm site' conforming to a consensus 'PuAU$_{4-6}$GPu' sequence flanked between two stem loops, to which the Sm proteins bind as a toroidal ring forming the Sm core ((Figure 1.3 C) (Stark et al. 2001; Weber et al. 2010). In contrast, U6 and U6atac snRNAs do not possess the classical Sm site but rather 3' uridine rich tails where seven 'like Sm' (LSm) proteins 2-8 bind (Achsel et al. 1999; Pannone et al. 2001). The non-spliceosomal U7 snRNP comprises of the U7 snRNA bound by a LSm/Sm hetero-heptameric ring wherein the canonical Sm D1/D2 are replaced with LSm 10/11 to assemble into a core structurally similar to the spliceosomal U snRNPs. The specialized U7 snRNP facilitates the 3'-UTR stem loop processing of the intron lacking histone-mRNAs (Pillai et al. 2003; Mowry and Steitz 1987).

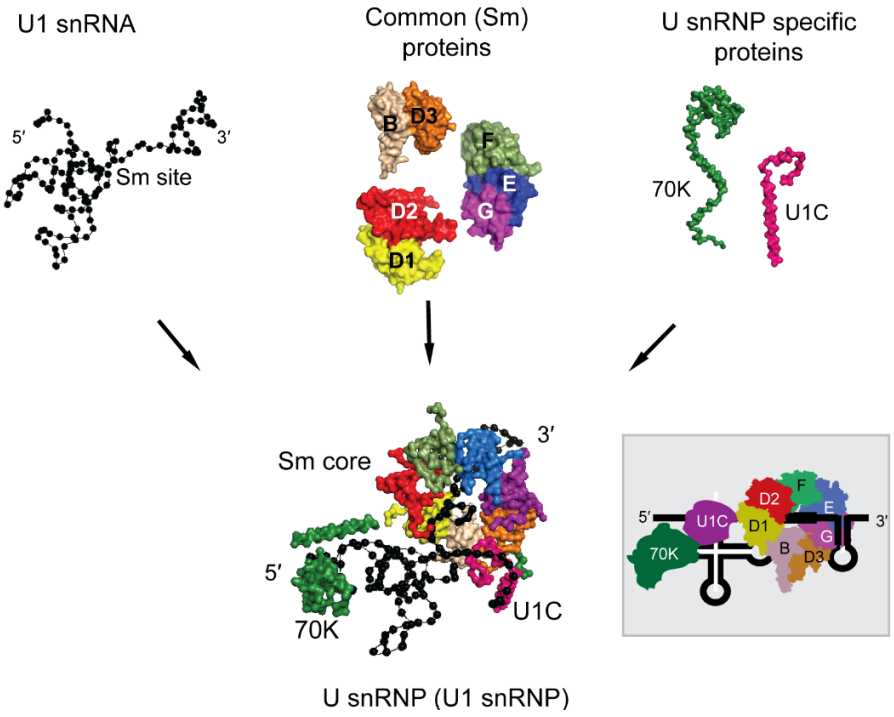


Figure 1.2 Components of spliceosomal U snRNPs
The uridine-rich small nuclear ribonucleotide proteins (U snRNPs) comprise of the U snRNA, the common set of Sm (or LSm) proteins and the snRNP specific proteins. The Sm proteins coalesce to form the toroidal Sm core around the Sm site of U snRNAs. Here, the Sm-class U1 snRNP components have been exemplified. The inset on the right provides a schematic representation of a mature U1 snRNP for better comprehension.
The surface representations for the RNA, Sm proteins and U1 snRNP have been adapted from published structures and rendered with PyMOL graphical interface. The references and the associated PDB codes are provided in Appendix 7.4.

The structural hallmark of the Sm and LSm family of proteins is the characteristic ‘Sm-fold’: a five stranded highly bent anti-parallel β sheet flanked by an N-terminal α helix (Figure 1.3 B) (Hermann et al. 1995; Seraphin 1995). Structural studies have elucidated the role of this Sm fold in establishing the heptameric Sm core around the Sm site of U snRNAs: the β4 strand of one Sm protein interfaces in an anti-parallel manner with the β5 strand of its neighbour forming a continuous intermolecular β sheet (Figure 1.3 B) (Kambach et al. 1999; Urlaub et al. 2001; Stark et al. 2001). The heptameric Sm ring, arranged in the order Sm E-G-D3-B-D1-D2-F (Figure 1.3 B) has an outer diameter of 70 Å and an inner hole of 20 Å (Pomeranz Krummel et al. 2009; Stark et al. 2001; Leung, Nagai, and Li 2011). Using their canonical Sm fold, each Sm protein contacts a distinct uridyl (U) nucleotide in the Sm site of the U snRNA, while the terminal extensions and extended internal loops of some of Sm proteins guide the U snRNA into and out of the toroidal Sm ring, stabilizing the Sm core RNP. The accumulation of positively charged amino acid residues within the cartwheel like arrangement of the Sm/LSm core most likely assists in overcoming the electrostatic repulsion

between the RNA components during splicing and in providing a structural basis for pre-mRNA-snRNA interactions (Kambach et al. 1999; Thore et al. 2003).

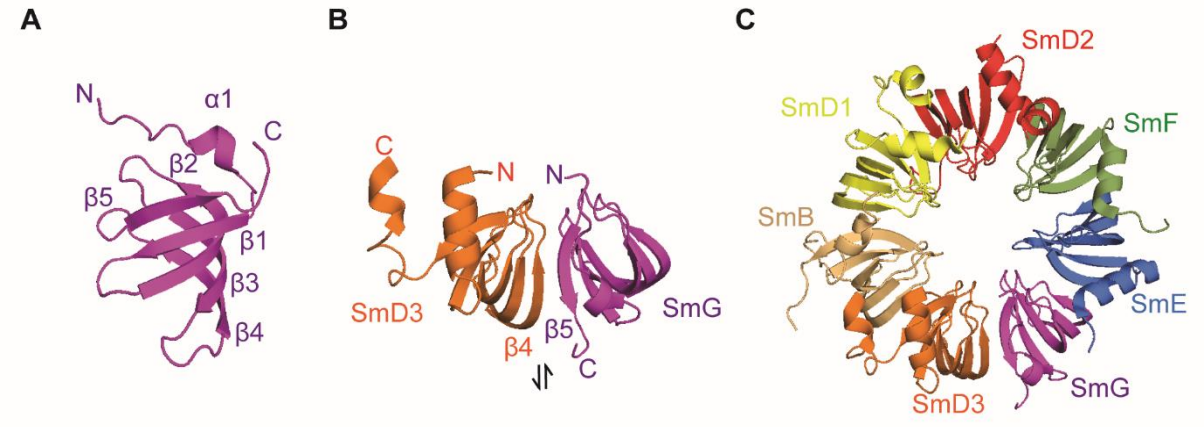


Figure 1.3 Sm proteins and their mode of interaction
(A) The Sm proteins share a common structural feature, the 'Sm-fold': a five stranded highly bent anti-parallel β sheet flanked by an N-terminal α helix. This has been exemplified here by the structural depiction of SmG (magenta).
(B) The β4 strand of one Sm protein (here, SmG) interfaces in an anti-parallel (indicated by the black arrow heads) manner with the β5 strand of its neighbour (here, SmD3) forming a continuous intermolecular β sheet. The respective N- and C-termini are labelled.
(C) In a similar manner as in (B), the seven Sm proteins engage with each other to establish the heptameric 'Sm-core' on the U-site of U snRNAs. The colour code employed for the individual proteins is the same throughout this dissertational work.
The above structural depictions have been adapted and modified from the U4 snRNP structural model (PDB code: 4WZJ) (Leung, Nagai, and Li 2011).

1.2 Sm class U snRNP biogenesis: A segmented pathway

A well-adjusted and synergistic relation between each individual step during U snRNP biogenesis guarantees faithful assembly of these macromolecular complexes (Prusty et al. 2017; Matera and Wang 2014). The sequential assembly of the spliceosomal U snRNPs (Section 1.1.1) on pre-mRNAs occurs exclusively within the nucleus co-transcriptionally, however the assembly of the individual U snRNP particles prevails over different cellular compartments (Carrillo Oesterreich, Bieberstein, and Neugebauer 2011). While the transcription of the U snRNAs occurs within the nucleus, a transient cytosolic phase is required to accomplish the assembly of the U snRNPs (Gruss et al. 2017).

The U snRNP biogenesis (Figure 1.8) begins with the transcription of the U snRNAs within the nucleus by RNA polymerase II (Pol II) (Mattaj and Zeller 1983; Egloff, O'Reilly, and Murphy 2008), with the exceptions of U6 and U6atac snRNAs being Pol III transcripts (Reddy et al. 1987). The Pol II snRNA transcripts acquire a 5' monomethyl guanosine (m^7G) co-transcriptionally (Baillat et al. 2005) and are then assembled into an export complex, with the 5' m^7G cap recruiting a distinct set of

factors: the cap binding complex (CBC) (Izaurralde et al. 1995), the phosphorylated adapter of snRNA export (PHAX) (Ohno et al. 2000), the arsenite resistance 2 factor (ARS2) (Hallais et al. 2013), the export receptor chromosome region maintenance 1 (CRM1) and GTP-bound Ran (Fornerod et al. 1997). Following the cytoplasmic translocation of the U snRNAs, the disassembly of the export complex ensues with the dephosphorylation of PHAX and hydrolysis of RanGTP. This marks the onset of the cytoplasmic assembly phase of the U snRNPs wherein the step-wise assembly of the Sm core on the Sm site of the U snRNAs is orchestrated by non-snRNP factors that are united in the PRMT5 (Protein Arginine Methyltransferase 5) and SMN (Survival Motor Neuron) complexes (Section 1.2.1) (Matera and Wang 2014).

Assembly of the Sm core on U snRNAs initiates the downstream snRNA processing events and its nuclear re-entry. The RNA methyltransferase trimethylguanosine synthase 1 (TGS1) is recruited for the hyper-methylation of the 5' m^7G-cap of the U snRNA to form the $m^{2,2,7}{}_3$G cap (2, 2, 7 trimethylguanosine/TMG) (Plessel, Fischer, and Luhrmann 1994). The assembled Sm core together with the TMG cap now provides a bipartite nuclear import signal for the SMN complex loaded with the assembled U snRNPs (Fischer et al. 1994; Hamm et al. 1990). This bipartite nuclear localisation signal triggers the assembly of the nuclear import complex comprised of the TMG binding factor, snurportin (SPN) and the Sm core binding import receptor, importin β (Huber et al. 1998; Palacios et al. 1997). Upon nuclear import, the U snRNPs are then transiently trafficked to the Cajal bodies for further maturation steps involving the binding of snRNP-specific proteins and covalent modifications (pseudouridylation and 2'-O-methylation) of the U snRNAs (Karijolich and Yu 2010). Concurrently, the SMN complex dissociates and the mature U snRNPs then concentrate to the nucleoplasmic splicing speckles awaiting spliceosomal assembly on pre-mRNA substrates (Sleeman and Lamond 1999).

In contrast, the nascent Pol III U6 and U6atac snRNA transcripts, are bestowed with a 5'-γ-methylphosphate cap (Singh and Reddy 1989) and remain confined to the nucleus for their entire biogenesis and subsequent function (Wolin and Cedervall 2002).

1.2.1 The cytosolic assembly phase of Sm class U snRNPs

Nuclear export of the nascent U snRNAs delivers them to the same compartment where the Sm proteins are synthesised and stockpiled, thereby allowing the Sm core

assembly on U snRNAs to proceed (Section 1.2). The seven Sm proteins can form stable RNA-free heteromers composed of Sm D1/D2, Sm E/F/G and Sm B/D3 (Raker, Plessel, and Luhrmann 1996; Hermann et al. 1995) that readily coalesce into the Sm core domain *in vitro* (Figure 1.3 B and C) (Raker et al. 1999). These isolated Sm protein oligomers have therefore the propensity to spontaneously self-assemble onto the U snRNAs (Raker, Plessel, and Luhrmann 1996; Raker et al. 1999). On the contrary, the crowded cytoplasmic environment necessitates the assistance of two macromolecular units *in vivo*, namely the PRMT5 (Protein Arginine Methyltransferase 5) complex and the SMN (Survival Motor Neuron) complex to orchestrate the faithful assembly of the snRNP core particle (Figure 1.4) (Meister and Fischer 2002; Paushkin et al. 2002). The step-wise cytoplasmic assembly of U snRNPs thus includes distinct temporal phases: the early assembly phase directed by the PRMT5 complex (Section 1.2.2) and the late SMN complex (Section1.2.3) driven assembly phase of the cytoplasm (Gruss et al. 2017).

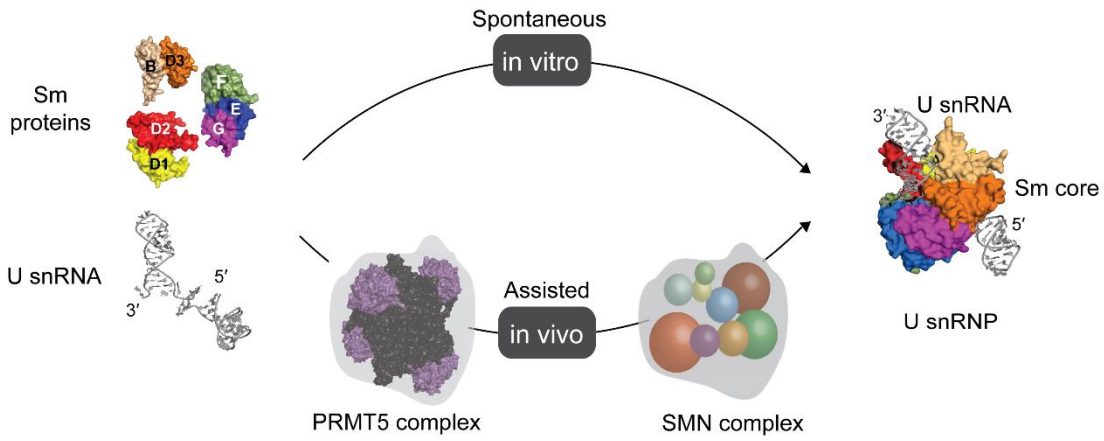


Figure 1.4 Assembly of U snRNPs
The Sm hetero-oligomers can spontaneously assemble on U snRNAs *in vitro* forming U snRNPs. Within the heterogeneous cellular cytoplasm, however, the Sm proteins are assembled on the U snRNAs in a staggered assembly line regulated by the multimeric PRMT5 (PRMT5/MEP50/pIcln) and SMN (SMN/Gemins 2-8/Unrip) complexes (Sections 1.2.2 and 1.2.3) (Chari and Fischer 2010).

1.2.2 Early cytoplasmic assembly phase

Three trans-acting factors, namely the protein arginine methyltransferase (PRMT) 5, the WD repeat protein WD45 (also termed as MEP50) (Figure 1.5 A) and pICln (chloride conductance regulatory protein) (Figure 1.5 B) unite in the PRMT5 complex to direct the early cytosolic U snRNP assembly pathway.

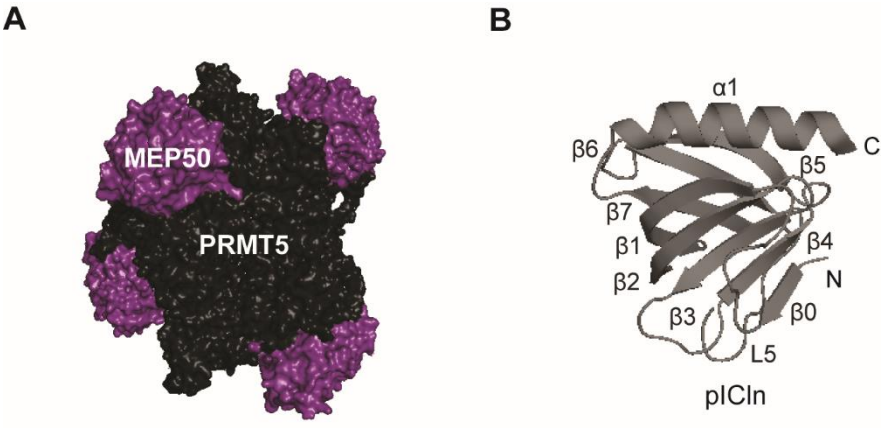


Figure 1.5 Trans-acting factors of the early cytoplasmic U snRNP assembly pathway
The PRMT5 methyltransferase enzyme, the MEP50 adaptor protein and pIcln, the assembly chaperone, pIcln unite in the PRMT5 complex to orchestrate the early cytoplasmic U snRNP assembly pathway.
(A) Surface representation of the core hetero-octameric PRMT5 (black):MEP50 (purple) complex (PDB ID: 4GQB).
(B) pICln (PDB ID: 4F7U) displays an extended pleckstrin homology topology (eight anti-parallel β strands, β0-β7) that mimics the canonical Sm fold (Figure 1.3 B), thus serving as topological organizer of the Sm core assembly intermediates. The L5 loop ('cover loop') of pICln is a highly mobile entity that occludes central cavity of the pICln/Sm ring within the 6S complex (see Figure 1.6 for details).

As a member of the type II methyltransferase family, PRMT5 catalyses the symmetrical di-methylation of the two-terminal ω-guanidino nitrogens of arginine residues (sDMA) on Sm proteins D1, D3 and B/B' (Friesen, Paushkin, et al. 2001; Brahms et al. 2001). The WD45 protein has been suggested to serve as an adaptor protein to recruit substrates and partner proteins to the PRMT5 methyltransferase enzyme (Friesen et al. 2002). Association of the WD45 protein with the PRMT5 enzyme favours the latter's oligomerisation in turn enhancing its stability and methyltransferase activity. The PRMT5:WD45 complex thus forms a core hetero-octameric module [$(PRMT5)_4$: $(WD45)_4$] that interacts with the pICln subunit during snRNP biogenesis (Figure 1.5 A) (Antonysamy et al. 2012; Meister et al. 2001). pICln functions as an assembly chaperone guiding the formation of Sm core assembly intermediates that progress to the SMN complex regulated snRNP late assembly phase (Chari et al. 2008). The newly translated Sm proteins D1/D2 stalled at the polypeptide exit tunnel of the ribosomes are sequestered by pICln for delivery to the PRMT5: WD45 complex for sDMA of SmD1 (Paknia et al. 2016). Simultaneously, the Sm hetero-trimer Sm E/F/G engages with the PRMT5 complex resulting in the formation of an assembly incompetent intermediate, the 6S complex, comprised of pICln, Sm D1, D2, E, F and G (Neuenkirchen et al. 2015). The N-terminus of pICln (Figure 1.5 B) adopts an extended pleckstrin homology (PH) topology with a conventional seven anti-parallel β strands (β1-β7) appended by an N-terminal β strand (β0). This structural feature of

pICln mimicking the canonical Sm fold (Figure 1.3 B) affords its interaction with the Sm proteins to generate the ring-shaped hexameric 6S complex (Figure 1.6). Within the 6S complex, pICln's β5 strand pairs in an anti-parallel orientation to the β4 strand of SmD1 similar to the canonical Sm-Sm contacts, thus pre-organizing the Sm proteins (Figure 1.6). However, the pICln: SmG interface is distinctly formed by a parallel rather than an anti-parallel β strand pairing, thus revealing a pre-determined breaking point for the displacement of the pICln subunit from the 6S complex during the later stages of the Sm core domain assembly (Section 1.2.3). pICln, thus occupying the width of 1.5 Sm proteins organises the five Sm proteins to their designated spatial positions comparable to an assembled heptameric Sm core (Figure 1.3 C). This generates a torus conferred with a constricted central pore preventing U snRNAs or non-cognate RNAs access (Grimm et al. 2013).

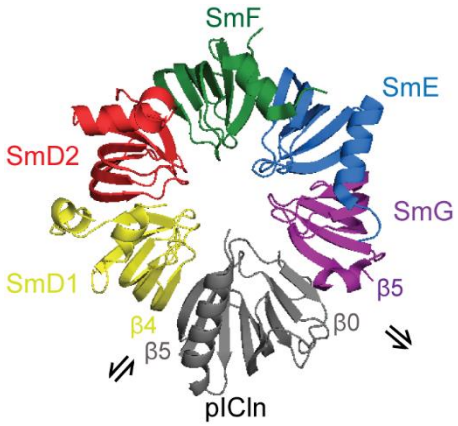


Figure 1.6 Structure of the kinetically trapped 6S complex
The molecular mimicry by the assembly chaperone, pICln (grey), allows the imitation of Sm-Sm interaction to establish the five Sm proteins (D1/D2/F/E/G) at their designated spatial positions later adopted in the assembled Sm-core (Figure 1.3 B). Within the 6S complex assembly intermediate (PDB ID: 4F7U), the Sm proteins engage via the anti-parallel pairing of β strands of their Sm folds. While SmD1 (β4 strand): pICln (β5 strand) interface resembles the canonical Sm:Sm interaction mode, the SmG (β5 strand):pICln (β0 strand) interface demonstrates a parallel pairing of the β strands albeit still establishing a continuous intermolecular β sheet within the 6S complex. The arrow heads adjacent to the interacting β strands of the pIcln:Sm protein interface depict the orientation.

Independently, another assembly incompetent intermediate, the heteromeric pIcln-Sm B/D3 is likely to engage with the PRMT5:WD45 complex for sDMA of Sm B and D3. Owing to the kinetic trap imposed by pICln, the Sm proteins fail to proceed spontaneously in the assembly pathway (Chari et al. 2008).

This PRMT5 complex directed early assembly (Figure 1.8) phase thus achieves two tasks: (1) methylation of the C-terminus arginine/glycine (RG) rich tails of Sm D1, D3, B/B' and (2) topological organization of Sm core intermediates by pICln, averting non-cognate RNA interactions and Sm protein aggregation (Gruss et al. 2017).

1.2.3 Late cytosolic assembly phase

The Sm core assembly intermediates chaperoned by pICln and the newly synthesized U snRNAs independently encounter the cytosolic SMN complex governing the late U snRNP assembly phase (Massenet et al. 2002). The step-wise assembly of the Sm core onto the U snRNAs is facilitated by the multimeric SMN (Survival Motor Neuron) complex which includes the eponymous SMN protein and eight other additional proteins, termed Gemins 2-8 and unrip (Figure 1.7) (Otter et al. 2007; Ogawa et al. 2009).

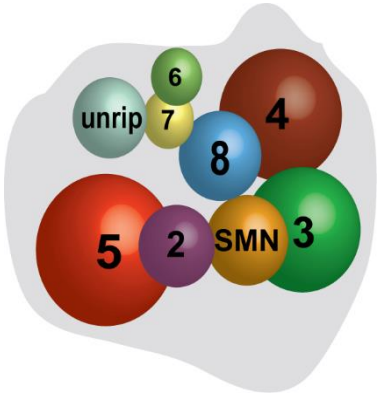


Figure 1.7 Schematic representation of the multimeric SMN complex
The human SMN (Survival Motor Neuron) complex includes the eponymous SMN protein engaging with eight additional proteins, Gemins 2-8 and unrip in a modular fashion (Section 1.3) to orchestrate the late cytoplasmic U snRNP assembly pathway.

Initially, the SMN complex accepts the pre-assembled yet kinetically trapped Sm core intermediates, the pICln bound 6S complex and pICln-SmB/D3 from the PRMT5 complex (Chari et al. 2008). While pICln-SmB/D3 is likely bound by SMN and/or Gemin8, the recognition of the 6S assembly intermediate extensively involves the Gemin2 component of the SMN complex (Zhang et al. 2011). The SMN complex then triggers the displacement of the assembly chaperone pICln, relieving the Sm protein complexes from their kinetic trap. This transforms the closed 6S complex to an open ring of the 5 pre-assembled Sm proteins D1-D2-E-F-G, held together by the Gemin2 subunit representing an assembly competent intermediate, the 7S complex (Zhang et al. 2011). Subsequently, the U snRNA substrate associates with the open 5-membered Sm protein ring on the SMN complex. Simultaneously, the activated SmB/D3 heteromer participates in the Sm protein ring closure around the Sm site of the U snRNA. (Chari et al. 2008). Recently, it was shown that the Gemin5 component of the SMN complex mediates a stringent recognition mechanism for the U snRNA substrates during the assembly reaction to preclude Sm core assembly on non-target

RNAs. The role of the other Gemins and unrip during the late snRNP assembly phase remains currently unclear due to the lack of structural information of the SMN complex as a whole entity (see Section 1.3).

The Sm core assembly stabilizes the U snRNA and initiates downstream processing events that culminate in the nuclear import of the pre-assembled U snRNPs for further maturation in the Cajal bodies (Matera and Wang 2014).

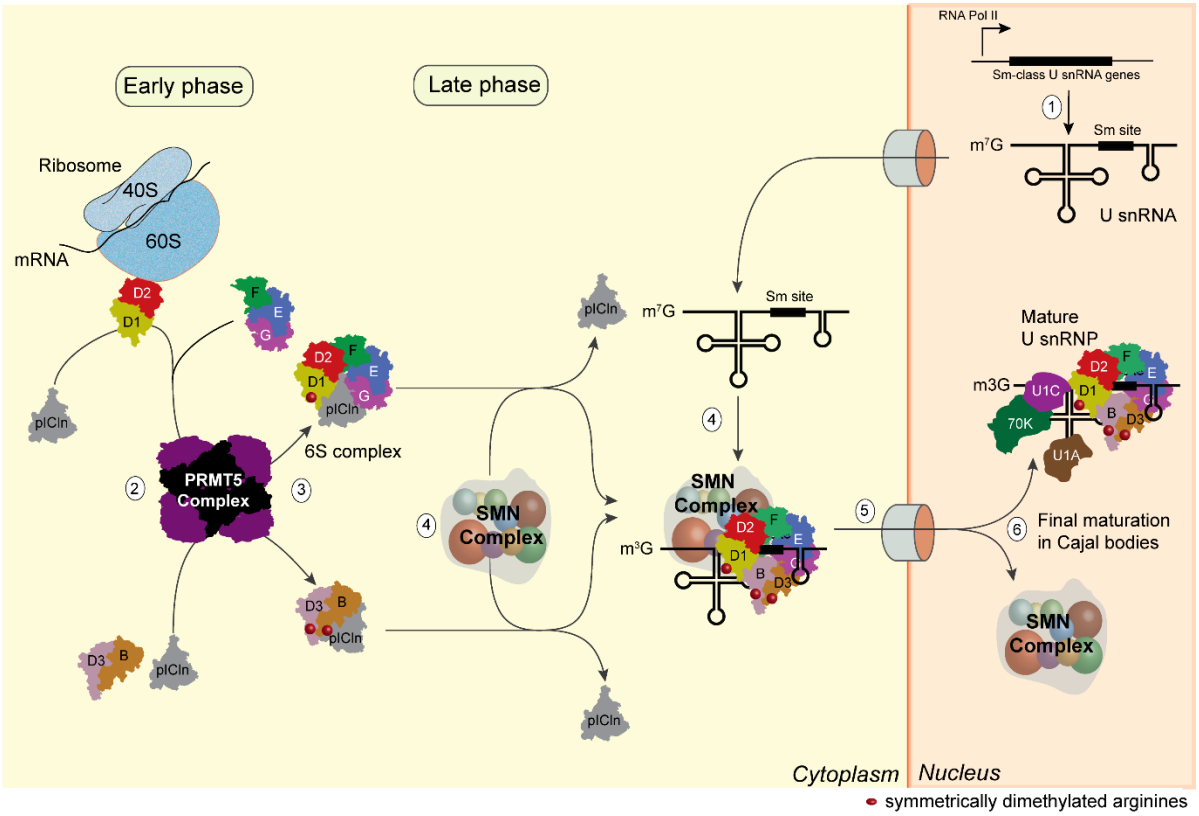


Figure 1.8 An overview of the Sm-class U snRNP biogenesis pathway and associated key regulators
The Sm-class U snRNP biogenesis pathway prevails over different cellular compartments: **(1)** In the nucleus, RNA polymerase II (RNA Pol II) transcribes the U snRNA genes (U1, U2, U4 and U5) followed by 5' capping and subsequent cytoplasmic export. The assembly of the Sm core and U snRNA loading occurs in the cytoplasm in a segmented pathway governed by the PRMT5 (early phase) and SMN (late phase) complexes. **(2)** The assembly chaperone pICln encounters the stalled Sm D1/D2 proteins at the ribosome and delivers them to the PRMT5:MEP50 complex and, in another assembly line, it sequesters Sm B/D3 and Sm E/F/G proteins to the latter complex. During the early assembly phase, the PRMT5 methylosome methylates the Arg (R) on the RG-rich tails of Sm B, D1 and D3, while pICln mimics the Sm fold architecture of Sm proteins and topologically organizes the five Sm proteins- Sm D1/D2/F/E/G- to generate a kinetically trapped intermediate, the 6S complex. **(3)** The late assembly phase begins with the delivery of the Sm core assembly intermediates, the pICln-bound 6S and Sm B/D3 complexes to the multimeric SMN complex. **(4)** The SMN complex relives the pIcln imposed kinetic trap and assembles the Sm core on the Sm site of U snRNAs. **(5)** This prepares the assembled U snRNPs for down-stream processing and re-entry to the nucleus. **(6)** Briefly localised to the Cajal bodies within the nucleus, the U snRNPs undergo further maturation with the addition of non-Sm proteins with simultaneous dissociation from the SMN complex. The mature U snRNPs eventually inhabit nuclear speckles awaiting splicing. Figure adapted and modified from Gruss et al., 2018.

1.3 Structural and functional facets of the SMN complex components

Along the evolutionary timeline, the ancestral assembly system of SMN-Gemin2 in fission yeast (*S. pombe*) and in plants has evolved into the multimeric human SMN complex with the addition of other Gemins (Gemins 3-8) and unrip, reflecting the emergence of complexity and refined cellular regulation in divergent organisms (Kroiss et al. 2008). Distinct subunits of this complex comprised of SMN-Gemin2, Gemins 3-5, Gemin5 and Gemin6-7/ unrip are formed *in vivo*, revealing a modular architecture (Battle et al. 2007). Gemin6 and unrip are recruited to the SMN complex via Gemin7's association with Gemin8, while Gemins 3, 4 and 5 engage directly with SMN, Gemin8 and Gemin2, respectively (Figure 1.9). Thus within this modular composition of the SMN complex, the SMN protein together with Gemin2 and Gemin8 forms the core scaffold providing a binding platform for the association Gemins 3-8 and unrip via multiple interactions (Otter et al. 2007; Ogawa et al. 2009).

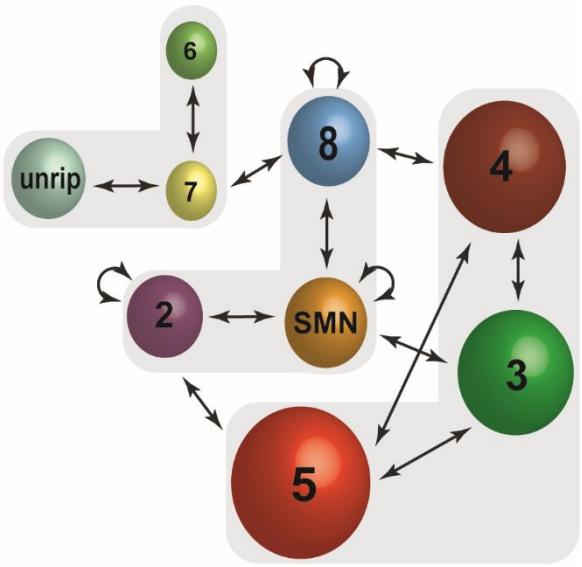


Figure 1.9 A consensus interaction map of the Survival Motor Neuron (SMN) complex
A schematic representation of all interactions within the SMN complex as described by Otter et al., 2007 and Ogawa et al., 2009. SMN together with Gemins 6-8 and unrip engage in reciprocal and oligomeric (depicted by circular arrows) interactions forming distinct modules that coalesce to form the highly oligomeric SMN complex. The colour code used for the individual proteins is maintained throughout this dissertation work.

1.3.1 Survival Motor Neuron (SMN) protein

The human SMN protein harbours multiple domains (Figure 1.10): (1) the N-terminal domain, (2) a central Tudor domain, (3) C-terminal proline-rich and (4) YG-box domain. All SMN orthologues show high sequence homology at their N- and C-termini, with only metazoans sharing the central canonical Tudor domain. Additionally, the

poly-proline region is sustainably shorter in non-mammalian vertebrates, in comparison to the mammalian SMN (Singh et al. 2017).

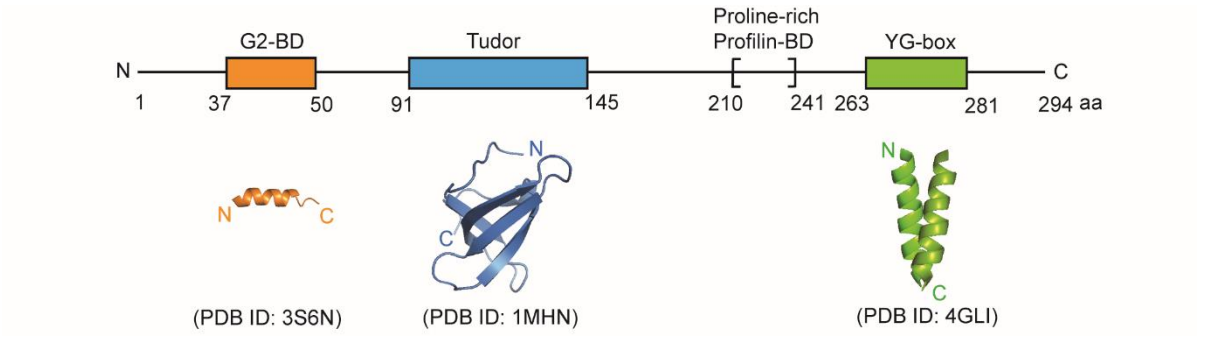


Figure 1.10 Domain organisation of Human Survival Motor Neuron (SMN) protein
Diagrammatic representation of the SMN protein depicting the N-terminus Gemin2 (G2) binding domain (BD), the central Tudor domain and the C-terminus profilin binding poly-proline stretch followed by the oligomerisation inducing YG-box domain. The domains and their boundaries (amino acid ranges) are indicated above and below the diagram respectively. The structural models for the helical Gemin2 binding domain (PDB code: 5XJL), the β-barrel Tudor domain (PDB code: 1MHN) and the glycine zipper motif bearing YG-box domain from a dimeric SMN moiety (PDB code: 4GLI) are depicted in the lower panel with the respective N- and C-termini labelled for reference.

A mounting body of structural knowledge of these individual domains have provided interesting insights on SMN's role during U snRNP biogenesis, albeit the protein in its entirety has been elusive to structural probing (Figure 1.10). The N-terminus of SMN, residues 36-51, adopts a helical structure that contacts Gemin2 within the SMN complex (Sarachan et al. 2012; Zhang et al. 2011). The SMN helix is strongly amphipathic with its hydrophobic face interacting with the C-terminus of Gemin2 within the SMN complex. The N-terminus nucleic acid binding domain of SMN partially overlapping with the Gemin2 binding domain, likely mediates U snRNA association and subsequent nuclear translocation (Lorson and Androphy 1998). The central Tudor domain appears to assert exclusivity to SMN's role in U snRNP biogenesis and has been structurally characterized by both NMR spectroscopy and X-ray crystallography. Similar to canonical Tudor domains, SMN's Tudor (residues 91-145) forms a highly bent β-sheet consisting of five strands assuming a barrel-like fold (Sprangers et al. 2003). Several studies highlight its preferential interactions with the symmetrically di-methylated arginine-rich C-terminal tails of Sm D1, D3 and B (Selenko et al. 2001; Buhler et al. 1999). This central Tudor domain also engages with coilin, a major component of the nuclear Cajal bodies where final maturation of U snRNPs occurs (Section 1.2.3, Figure 1.8) (Tapia et al. 2014). The poly-proline stretch of SMN interacts with profiling IIa which plays important roles in neuronal axons and growth cones (Sharma et al.

2005). The highly oligomeric SMN forms hetero-disperse complexes on sucrose gradients (40S-80S) with *in vivo* experiments speculating molecular dimensions of over 1 MDa. The C-terminus YG-box (residues 263-281) of SMN forming glycine zipper oligomers (Figure 1.10) is primarily responsible for SMN self-association (Martin et al. 2012; Lorson et al. 1998). The highly conserved tyrosine (Y) and glycine (G) repeats within this domain pack to stabilize inter-helical interactions, providing a structural explanation for being the most conserved motif across divergent organisms. Recent work on the *S. pombe* SMN inferred that SMN self-association via the YG-box domain would be heteromeric: the YG-box helical bundles of different SMN moieties interact in a non-symmetrical manner forming the oligomeric structural scaffold within the SMN complex (Gupta et al. 2015).

1.3.1.1 *Role of SMN in Spinal Muscular Atrophy (SMA) pathogenesis*

Complete loss or decreased levels of the SMN protein causes the autosomal recessive neurodegenerative disorder, Spinal Muscular Atrophy (SMA). Clinically, SMA is characterized by degeneration of α-motor neurons in the anterior horn of the spinal cord, progressive skeletal muscle atrophy and wasting and, in severe cases, paralysis and death (Burghes and Beattie 2009). The SMA causative *SMN* gene on chromosome 5 (5q13.2) exists in two copies: telomeric *SMN1* and centromeric *SMN2* (Lefebvre et al. 1995). While the telomeric *SMN1* copy generates functional SMN protein, the centromeric *SMN2* primarily produces the truncated isoform SMNΔ7, due to predominant exon7 exclusion arising from a translationally silent C → T transition within this exon (Kashima and Manley 2003). As a consequence, with only 10 % of functional SMN generated from the *SMN2* gene locus, rendering *SMN2* incapable of compensating for the loss of *SMN1*. SMA is therefore the result of low levels of functional SMN protein derived from *SMN2* and not its absence (Lefebvre et al. 1997; Coovert et al. 1997). The copy number of *SMN2* correlates with the phenotypic disease severity manifesting as four clinical forms (Type I-IV) of SMA. Type I SMA being the most severe is typically diagnosed in the first months of life, Types II and III SMA are less severe forms of the disease with later childhood onset and Type IV SMA is diagnosed in patients older than 30 (Coady and Lorson 2011).

Mutations in all the domains of SMN have been linked to SMA pathogenesis implying the structural importance for SMN's function (Figure 1.10). A small number of mutations such as D30N (Type II SMA) and D44V (Type III SMA) appear in the N-

terminus of SMN. Notably, the D44V mutation abrogates SMN's ability to interact with Gemin2 within the SMN complex leading to reduced U snRNP assembly activity (Sun et al. 2005). SMA missense mutations localized in the SMN Tudor domain (E134K, Q136E) have been reported to disrupt Sm protein-SMN interactions (Selenko et al. 2001; Cusco et al. 2004; Sun et al. 2005). Most of the severe SMA patient mutations cluster in the exon6 and exon7 coded YG-box domain (S266P, Y272C etc.) bestowing defective SMN self-association *in vitro* (Lorson et al. 1998; Martin et al. 2012; Pellizzoni, Charroux, and Dreyfuss 1999). For instance, the Y272C SMA patient mutation severely disrupting SMN oligomerisation, affects Gemin3 and Gemin8 binding within the SMN complex as well as abrogates Sm protein binding (Charroux et al. 1999; Otter et al. 2007).

1.3.2 Gemins 2-8 and Unrip

Gemin2 (Figure 1.11 A) within the SMN complex is a phylogenetically conserved subunit across divergent organisms along with the core SMN protein (Kroiss et al. 2008; Liu et al. 1997; Noble and Guthrie 1996). Consistent with this, Gemin2 is required for the viability of all eukaryotic organisms (Jablonka et al. 2002; Borg and Cauchi 2013; Ogawa et al. 2007). Structural studies of the Sm core assembly intermediates, the 8S and 7S complexes, demonstrated Gemin2 to be a key factor in U snRNP biogenesis linking the pre-assembled Sm proteins to the SMN complex. The U snRNP assembly incompetent 8S complex comprises of the kinetically trapped 6S complex (pICln-SmD1/D2/F/E/G) (Figure 1.6) bound by Gemin2 and a minimal SMN (N-terminus and Tudor domain) unit (Figure 1.11 F). In the stalled 8S complex, Gemin2 mediates direct interaction with the preformed Sm pentamer, however neither Gemin2 nor the minimal SMN unit can facilitate pICln expulsion to drive forward the U snRNP assembly pathway. Molecular dynamics simulations and EM reconstructions of the 8S intermediate revealed that the mobile cover loop (L5) of pICln constricts the central cavity of the 6S ring, precluding non-cognate RNA access (Grimm et al. 2013). The 7S complex, in contrast, represents an assembly competent intermediate following the release of the assembly chaperone, pICln. Notably, the 7S assembly intermediate shares certain similar structural features with the 8S complex: the C-terminal domain of Gemin2 (residues 100-280) contacts the N-terminus of SMN (Section 1.3.1, Figure 1.10), while the contact between Gemin2 and the Sm pentamer is remarkably extensive with interactions with both Sm D1/D2 and Sm E/F/G via distinct N- (residues 1-69, α1 chain) and

C- (residues 100-280, chains α2, α5-8) termini structural domains respectively (Figure 1.11 E). Here in the 7S complex, Gemin2 holds the pentameric Sm ring in an open configuration in preparation for U snRNA loading and Sm ring closure (Section 1.2) (Zhang et al. 2011; Grimm et al. 2013).

Gemin3 of the SMN complex is DEAD box (residues 211-214 in helicase motif II) containing RNA helicase (Figure 1.11 B) that directly engages with SMN via its C-terminus (residues 456-547) (Schutz et al. 2010). Gemin3 has been proposed to provide the enzymatic activity of SMN complex to render structural transitions during U snRNP biogenesis. Notably, the SMA patient mutation Y272C in the YG-box domain (Figure 1.10) or the exon7 deletion that severely affect SMN oligomerisation reduce Gemin3 association, impairing U snRNP assembly (Charroux et al. 1999).

Gemin4 is associated with the SMN complex via its interaction with both Gemin3 and Gemin5 (Ogawa et al. 2009; Otter et al. 2007). Gemin4 is presumed to act as a co-factor for the Gemin3 RNA helicase. The domain architecture of Gemin4 reveals a leucine rich C-terminus Gemin3 binding domain and a nuclear localisation signal (NLS) bearing N-terminus speculated to mediate nuclear import of SMN complex or its subunits (Charroux et al. 2000; Meier, Walker, and Matera 2018).

Gemin5 is recruited to the SMN complex via its C-terminal interaction (residues 721-1,508) with Gemin2 (Otter et al. 2007). Gemin5 contains 14 repeats of canonical WD40-motifs at its N-terminus that form two connected seven-bladed β-propellers (Figure 1.11 D). Gemin5 allows the faithful assembly of Sm core only on cognate U snRNAs by demonstrating a bimodal recognition mechanism: the Sm site on U snRNAs is recognized by Gemin5's first β propeller domain while its second β propeller binds the 5' m^7G cap on nascent U snRNAs to preclude internal ribosome entry site-dependent translation (Xu et al. 2016; Jin et al. 2016; Tang et al. 2016).

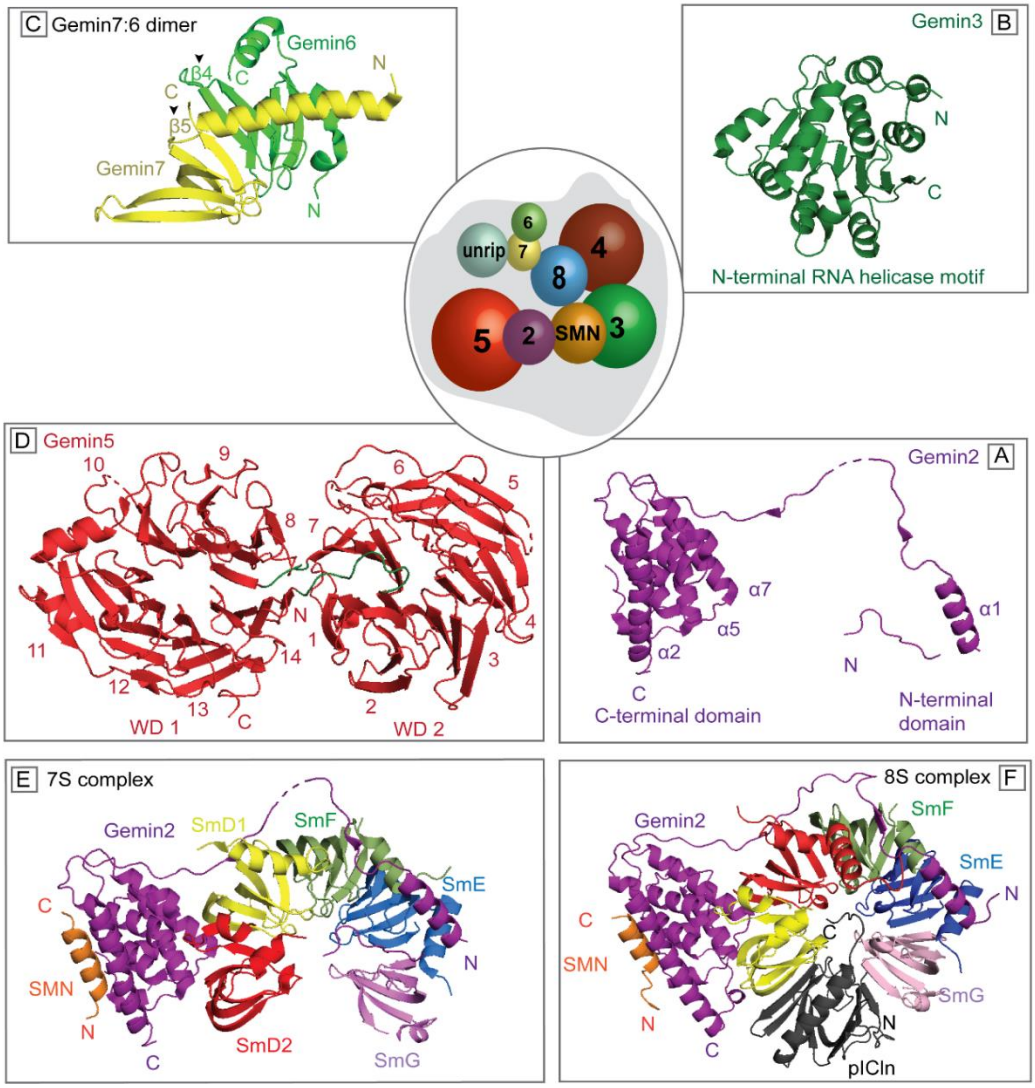


Figure 1.11 Structural models of SMN complex components
The multimeric SMN complex includes the eponymous SMN protein engaged in multiple interactions with the Gemins 2-8 and unrip (Section 1.3.1). In the centre, a schematic representation of the SMN complex is provided with the respective proteins occupying spatial orientations based on the interaction map shown in Figure 1.9.
(A) The Gemin2 protein (in purple, PDB ID: 36SN) has an extended conformation with distinct N- and C-terminal domains.
(B) The RNA helicase, Gemin3 (in green, PDB ID: 3B7G) bears the DEAD-box helicase motif at its N-terminus characterised by conserved and intertwined alpha helices and beta strands. Gemin3 interacts with SMN via its C-terminal domain.
(C) The Gemin7:6 sub-complex (in yellow and light green, respectively, PDB ID: 1Y96) engage with the SMN complex via Gemin8. Both Gemin7 and Gemin6 have the characteristic 'Sm-fold' and interact in a head-to-tail manner (indicated by arrow heads), very reminiscent of Sm hetero-oligomers Sm B/D3, Sm D1/D2 and Sm E/F/G.
(D) Gemin5's N-terminus harbours two seven bladed beta-propeller domains (WD1 and WD2) (in red, PDB ID: 5GXI). The loops linking these domains are coloured green. The WD1 domain recognizes the Sm-site on U snRNAs, while the WD2 domain interacts with 5' m7G cap on nascent U snRNAs, thus ensuring cognate RNA-Sm protein associations.
(E) Overall structure of the 7S complex comprising of Gemin2 bound to the Sm sub-core and N-terminus of SMN (PDB ID: 3S6N): Gemin2's (purple) C-terminal domain interacts with both Sm D1/D2 and with the N-terminus of SMN (in orange, residues 37-51) on opposite distal surfaces, while its N-terminal domain interacts with Sm F/E/G.
(F) Overall structure of the 8S complex (PDB ID: 4V98) comprising of Gemin2 bound to the 6S complex: Gemin2's C-terminal domain interacts with both Sm D1/D2 and with the N-terminus of SMN (orange), while its N-terminal domain interacts with Sm F/E/G. This minimal Gemin2:SMN unit is incapable of pICln release, rendering the 8S complex an assembly incompetent intermediate is contrast to the assembly competent 7S complex. Notably, the C-terminal cover loop of pICln localises within the Sm/pICln ring constricting its central cavity.

Although lacking any sequence similarity to Sm proteins, Gemins 6 and 7 possess the canonical Sm fold architecture (Figure 1.3 A), enabling their head-to-tail association, reminiscent of the Sm hetero-oligomers: Sm B/D3, Sm D1/D2 and Sm E/F/G (Figure 1.11 C). This architectural similarity has been speculated to aid during the step-wise Sm core assembly on the SMN complex (Ma et al. 2005).

Gemin8 being one of the Gemins to be evolutionarily conserved among metazoans demonstrates multiple direct interactions within the SMN complex (Kroiss et al. 2008). Gemin8 recruits Gemin7:6 via its C-terminus whilst Gemin8's N-terminus engages with SMN's C-terminus and Gemin4 (Otter et al. 2007). Either Gemin8 and/or the C-terminus of SMN could mediate pIcln release from the 6S complex with simultaneous Sm core formation with the addition of Sm B/D3 to the pre-assembled subcore (Chari et al. 2008).

Unrip (upstream of N-ras [unr]-interacting protein) is a WD-repeat protein that is recruited to the cytoplasmic SMN complex via its C-terminus (residues 195-350) interacting with Gemin7. Interestingly, reduced unrip protein levels triggered enhanced SMN nuclear localisation, probably influencing the intracellular distribution of SMN (Grimmler et al. 2005; Otter et al. 2007). Unrelated to its engagement with the SMN complex, unrip forms a complex with unr, a factor implicated in cap-independent translation of mRNAs (Hunt et al. 1999; Mitchell et al. 2003).

2 Thesis Objectives

While the general principles of the SMN complex in orchestrating U snRNP biogenesis has been understood to a certain extent, key structural and mechanistic facets are yet to be defined. To date, only the atomic resolution structures of individual SMN complex components and/or sub-complexes thereof have been elucidated (Section 1.3), while the structural framework of the entire SMN complex remains elusive.

The goal of this work was to define how the distinct modular units of the multimeric SMN complex coalesce to form this macromolecular machine by biochemical and structural studies. Employing the previously reported interaction map of the SMN complex (Otter et al. 2007), biochemical reconstitutions of sub-complexes from bacterial co-expression systems were pursued. To this end, SMN complex orthologues from divergent organisms (*H. sapiens*, *C. elegans*, *S. pombe*) were engaged to ensure high crystallization propensity for X-ray crystallography studies. Complimenting these structural studies, *in vitro* protein interaction assays were used to investigate the influence of SMA patient mutations in SMN's YG zipper motif in moulding the architecture of the heteromeric SMN complex.

3 Materials and Methods

3.1 Materials

3.1.1 General Materials

All chemicals and substances were of analytical grade and procured from reputed vendors such as Carl Roth GmbH & Co. KG (Karlsruhe, Germany), GE Healthcare (Munich, Germany) and Sigma Aldrich Chemie GmbH (Munich, Germany) unless otherwise specified.

3.1.2 Commercial Kits

Kit	Vendor	Purpose
NucleoSpin® Plasmid QuickPure	Macherey-Nagel, Germany	Plasmid DNA isolation (Mini-prep)
NucleoBond® PC100	Macherey-Nagel, Germany	Plasmid DNA isolation (Midi-prep)
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel, Germany	PCR clean-up/ Agarose Gel extraction of DNA
TNT® T7 Quick coupled Transcription/ Translation system	Promega, Germany	*In vitro* [^{35}S]-Methionine labelling of proteins

3.1.3 Oligonucleotides

Desalted oligonucleotides were custom-ordered from Sigma Aldrich Chemie GmbH, (Munich, Germany). Prior to use, the lyophilized oligonucleotides were re-suspended in appropriate volumes of sterilised water to obtain a final concentration of 100 µM. All the reverse orientation oligonucleotides bear the *Xho*I-*Not*I-*Nhe*I restriction site sequences in the 5'-3' orientation to facilitate polycistronic assembly of desired genes for co-expression (see Section 3.2.1.3). Only the oligonucleotides corresponding to the plasmid constructs presented in this work are presented here, while oligonucleotides obtained from RG Fischer have not been included.

3.1.3.1 *Cloning primers for S. pombe orthologues*

No.	Target gene (Deletion variant)	RE site	DNA sequence (5'-3')*
ohAV89	Gemin8ΔN5	*Nco*I	CATG*CCATGGCG*ACAGAAGGGACTTACAGAAATTCCA
ohAV90	Gemin8ΔN10	*Nco*I	CATG*CCATGGCG*CAGAAATTCCATGATGAGCATTT

ohAV91	Gemin8ΔN15	*Nco*I	CATG*CCATGG*AGCATTTTAATGCAAAAGCAGT
ohAV92	Gemin8ΔN20	*Nco*I	CATG*CCATGGCG*AAAGCAGTAAACCTATGGAATGTTG
ohAV93	Gemin8ΔN25	*Nco*I	CATG*CCATGGCG*TGGAATGTTGCTTTCGCG
ohAV94	Gemin8ΔN30	*Nco*I	CATG*CCATGGCG*GCGCAAAACGATCGTATCTC

*****Respective restriction enyzme (RE) sites have been italicised and underlined while cloning overhangs essential for RE activity are italicised.

3.1.3.2 Cloning primers for *C. elegans* orthologues

No.	Target gene (Deletion/ Mutation)	RE site	DNA sequence (5'-3')
oAV1	Gemin6ΔC74	*Xho*I	*GAATTCTCCGAG*GCGGCCGCGCTAGCTTATAATTCTATTCCGTCATTCTCG
oAV2	Gemin8 ΔN117	*Nde*I	*GAATTCATATG*TCAAGTTCAAAAAGTCAAGAGGAATACG
oAV3	Gemin8 ΔN145	*Nde*I	*GAATTCCATATG*CACCGAAACGCCAGCGCCGAATTCATAGC
oAV9	SMNΔN168	*Nco*I	*CATGCCATGG*CAATGGCCCCTGTTAATC
oAV11	SMNΔN181	*Nco*I	*CATGCCATGG*ACATGCTTATGAGCTGGTACATGAGC
Mutagenesis primers			
oAV24	SMN M182R	-	GTACCAGCTCATAAGCCTACTGTTCATGGCTTCTTTCT
oAV25			AGAAAGAAGCCATGAACAGTAGGCTTATGAGCTGGTAC
oAV26_F	SMN M182T	-	GTACCAGCTCATAAGCGTACTGTTCATGGCTTCTTTCT
oAV26_R			AGAAAGAAGCCATGAACAGTACGCTTATGAGCTGGTAC
oAV27	SMN S185P	-	GATATCCGCTCATGTACCAGGGCATAAGCAT-ACTGTTCATGG
oAV28			CCATGAACAGTATGCTTATGCCCTGGTACATGAGCG-GATATC
oAV29	SMN H192R	-	GCTTGATAGTATCCTGTACGATATCCGCTCATGTACC
oAV30			GGTACATGAGCGGATATCGTACAGGATACTATCAAGC

*****Respective restriction enzyme (RE) sites have been italicised and underlined while cloning overhangs for RE activity are italicised.

3.1.3.3 Cloning primers for *H. sapiens* orthologues

No.	Target gene (Deletion/ Mutation)	RE site	DNA sequence (5'-3')
ohAV1	Gemin6	*Nco*I	*CATGCCATGGCG*ATGAGTGAATGGATGAAGAAAGG
ohAV2		*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGCTCATTGGGAA-GCTGTAAGATGTC
ohAV3	Gemin6ΔC75	*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGCTCATTTGCAG-TCTCCAGACGT
ohAV4	Gemin7	*Nde*I	*GGAATTCCATATG*CAAACTCCAGTGAACATTC
ohAV7		*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGCTTATGGCTT-GAAGGTATATGAAATAAT
ohAV5	Gemin7ΔN30	*Nde*I	*GGAATTCCATATG*AGAGCCCCCTTGAGG

ohAV6	Gemin7ΔN45	*Nde*I	*GGAATTCATATG*ATAGCTCAAGAATCCCTG
ohAV8	Gemin8	*Nde*I	*GGAATTCCATATG*GCAGCGGTAAAGG
ohAV11		*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGC TCAGAACTTCAGGGGGATGA
ohAV10	Gemin8ΔN103	*Nde*I	*GGAATTCCATATG*AGTAGGATCCAGGCATC
ohAV13	Gemin8ΔN190	*Nde*I	*GGGAATTCCATATG*GAGAGGCCTGGTGAG
ohAV12	Gemin8ΔC12	*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGCTCAC-TTTCGGTCACAGTGCT
ohAV26	SMNΔN251	*Nco*I	*CATGCCATGG*ATTCTCTTGATGATGCTGATGC
ohAV28	SMNΔN262	*Nco*I	*CATGCCATGG*ACATGTTAATTTCATGGTACATGAGTGG
ohAV22	SMN	*Xho*I	*GAATTCCTCGAG*GCGGCCGCGCTAGCTTAATTTAAGG AATGTGAGCAC CT
ohAV25	SMNΔC10	*Xho*I	*GAATTCCTCGAG*GCTAGCGCGGCCGCTTATTGTCTGA AACCCATATAATAGCC
ohAV29	Unrip	*Nco*I	*CATGCCATGG*CAATGAGACAGACGC
ohAV30		*Xho*I	*GAATTCCTCGAGGCGGCCGCGCTAGCTCAGGCCTTAA CATCAGGAGC*
Mutagenesis primers			
ohAV66	SMN M263R	-	CTCATGTACCATGAAATTAACCTACTTCCCAAA-GCATCAGCAT
ohAV67			ATGCTGATGCTTTGGGAAGTAGGTTAATTTCATGG-TACATGAG
ohAV68	SMN M263T	-	CTCATGTACCATGAAATTAACGTACTTCCCAAA-GCATCAGCAT
ohAV69			ATGCTGATGCTTTGGGAAGTACGTTAATTTCATGG-TACATGAG
ohAV53	SMN S266P	-	CCACTCATGTACCATGGAATTAACATACTTCCCAAAGC ATCAGC
ohAV54			GCTGATGCTTTGGGAAGTATGTTAATTCCATGGTACAT GAGTGG
ohAV49	SMN 270E	-	ACCCATATAATAGCCAGTATGATAGCCCTCCATGTACC ATGAAATTAACATACTTCC
ohAV50			GGAAGTATGTTAATTTCATGGTACATGGAGGGCTATCA TACTGGCTATTATATGGGT
ohAV80	SMN Y272C	-	ATAATAGCCAGTATGACAGCCACTCATGTAC-CATGAAATTAACATA
ohAV81			TATGTTAATTTCATGGTACATGAGTGGCTGTCAT-ACTGGCTATTAT
ohAV70	SMN H273R	-	CCCATATAATAGCCAGTACGATAGCCACTCATGTACC
ohAV71		-	GGTACATGAGTGGCTATCGTACTGGCTATTATATGGG
ohAV72	SMN T274I	-	CCCATATAATAGCCAATATGATAGCCACTCATGTAC-CATGAAATT
ohAV73		-	AATTTCATGGTACATGAGTGGCTATCATATTGGC-TATTATATGGG

*****Respective restriction enzyme (RE) sites have been italicised and underlined while cloning overhangs for RE activity are italicised. Bases highlighted in grey include the *Not*I-*Nhe*I (or *Nhe*I-*Not*I for ohAV25 alone) sites for iterative cloning (See Section 3.2.1.3).

3.1.3.4 Mutagenesis primers for pETM13_AV and pETM30_AV vectors

The pETM13 and pETM30 vectors were modified by the introduction of a *Nco*I restriction enzyme site in place of the existing *Nde*I site by site-directed mutagenesis to generate the pETM13_AV and pETM30_AV vectors (See Section 3.1.5):

Mutagenesis primers	DNA sequence (5'-3')
pETM13_F	CACGCCGGGCTCTATCTGGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATT
pETM13_R	AATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCCAGATAGAGCCCGGCGTG
pETM30_F	GAAATCTTTATTTTCAGGGCCATATGGCTGGCAAGGCACACAGG
pETM30_R	CCTGTGTGCCTTGCCAGCATATGGCCCTGAAAATAAAGATTCTC

3.1.4 Enzymes

3.1.4.1 *DNA Modifying Enzymes*

Enzyme	Vendor
KAPA HiFi DNA polymerase	KAPA Biosystems, Germany
T4 DNA Ligase	Thermo Fisher Scientific, Germany
Restriction Enzymes	Thermo Fisher Scientific, Germany

3.1.4.2 *Proteases for Affinity tag removal*

The following proteases were obtained from in-house preps as detailed below:

Protease	Molecular weight* (kDa)	Protease recognition site**	Affinity tag for purification	Vector/ Resistance/ Expression strain
Tobacco Etch Virus (TEV) protease	26	ENLYFQ↓G	N-terminal His_6	pET28a/ Kanamycin/ *E. coli* BL21 (DE3)
GST-3C (PreScission™) Protease (PSc)	22	LEVLFQ↓GP	N-terminal GST	pGEX3X/ Ampicillin/ *E. coli* BL21 (DE3) pLyS pRARE)

*Molecular weight without affinity tag; **↓: Position of proteolytic cleavage

3.1.5 Bacterial Plasmid Vectors

Plasmid	Affinity tag	Protease Cleavage site	Resistance	Reference/Supplier/ Origin
pET21a	N/A	N/A	Ampicillin	Novagen, USA
pET28a	N- and C-terminal His_6	Thrombin	Kanamycin	Novagen, USA
pETM11	N-terminal His_6	TEV	Kanamycin	EMBL, Germany
pETM13	C-terminal His_6	N/A	Kanamycin	EMBL, Germany
pETM13_AV	C-terminal His_6	N/A	Kanamycin	In this study
pETM30	N-terminal His_6 –GST	TEV	Kanamycin	EMBL, Germany
pETM30_AV	N-terminal His_6 –GST	TEV	Kanamycin	In this study
pETM41	N-terminal His_6 –MBP	TEV	Kanamycin	EMBL, Germany
pETM41*[a]	N-terminal His_6 –MBP	N/A	Kanamycin	RG Fischer
pGEX6P1	N-Terminal GST	PreScission™ (PSc)	Ampicillin	GE Healthcare, UK
pGEX6P1_AV	N-Terminal GST	PreScission™ (PSc)	Ampicillin	In this study

[a] Refer Figure 7.3 for vectors pETM41 and pETM41* variations

3.1.6 Bacterial Expression Vectors

For 5’ and 3’ cloning sites and poly-cistronic cloning strategy, refer to Sections 3.2.1.1 and 3.2.1.2. The intermediate mono-cistronic cloning cassette containing plasmids have not been included in the following lists:

S. pombe

Plasmid	Protein (Amino acids, aa)	Source
pETM11: His_6(TEV)-G7:G6	Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	RG Fischer
pETM13: G8ΔL:His_6(TEV)-G7:G6	Gemin8 (1-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	RG Fischer

pETM13: G8 ΔN5ΔL:His_6(TEV)-G7:G6	Gemin8 (6-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	This work
pETM13: G8 ΔN10ΔL:His_6(TEV)-G7:G6	Gemin8 (11-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	This work
pETM13: G8 ΔN15ΔL:His_6(TEV)-G7:G6	Gemin8 (16-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	This work
pETM13: G8 ΔN25ΔL:His_6(TEV)-G7:G6	Gemin8 (22-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	This work
pETM13: G8 ΔN30ΔL:His_6(TEV)-:G6	Gemin8 (31-35/115-166)* Gemin7 (1-91) (fl) Gemin6 (1-93) (fl)	This work
pETM11: His_6(TEV)-SMNΔL	SMN (1-35/120-152)	RG Fischer

*ΔL: Loop deletion variant (lacking 36-114 residues)

C. elegans

Plasmid	**Protein (aa)/Mutations**	**Source**
pET21a: His_6G7:G6	Gemin7 (1-92) (fl) Gemin6 (1-168) (fl)	RG Fischer
pET21a: G6ΔC74	Gemin6 (1-94)	This work
pET28a: His_6G7:G6ΔC74	Gemin7 (1-92) (fl) Gemin6 (1-94)	This work
pET28a: His_6G8	Gemin8 (1-199) (fl)	RG Fischer
pET28a: His_6G8ΔN56	Gemin8 (57-199)	This work
pGEX6P1_AV: GST(PSc)-G8ΔN117	Gemin8 (118-199)	This work
pGEX6P1_AV: GST(PSc)-G8ΔN144	Gemin8 (145-199)	This work
pGEX6P1_AV: GST(PSc)-G8ΔN144ΔC12	Gemin8 (145-187)	This work
pETM41*: His_6–MBP-SMN_$YG^{182-207}$	SMN (182-207)	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$	SMN (169-207)	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_M182R	SMN (169-207)/ M182R	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_M182T	SMN (169-207)/ M182T	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_S185P	SMN (169-207)/ S185P	This work

pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_Y191C	SMN (169-207)/ Y191C	RG Fischer
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_H192R	SMN (169-207)/ H192R	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{169-207}$_T193I	SMN (169-207)/ T193I	RG Fischer

D. melanogaster

Plasmid	Protein (aa)	Reference
pETM30:G2	Gemin2 (1-245) (fl)	Kroiss et al., 2008
pETM13: dpIcln (ΔN_H144A_ΔC)	pIcln (Δ90-125 H144A Δ160-125)	Chari et al., 2008 Grimm et al., 2013

H. sapiens

Plasmid	Protein (aa) / Mutations	Source/ Reference
pETM13: G6	Gemin6 (1-167) (fl)	This work
pETM13: G6 ΔC75	Gemin6 (1-92)	This work
pETM30: His_6-GST(TEV)-G7:G6	Gemin7 (1-131) (fl) Gemin6 (1-167) (fl)	This work
pETM30: His_6-GST(TEV)-G7ΔN30:G6	Gemin7 (31-131) Gemin6 (1-167) (fl)	This work
pETM30: His_6-GST(TEV)-G7ΔN45:G6	Gemin7 (46-131) Gemin6 (1-167) (fl)	This work
pETM30: His_6-GST(TEV)-G7:G6ΔC75	Gemin7 (1-131) (fl) Gemin6 (1-92)	This work
pETM30: His_6-GST(TEV)-G7N30:G6ΔC75	Gemin7 (31-131) Gemin6 (1-92)	This work
pETM30: His_6-GST(TEV)-G7N45:G6ΔC75	Gemin7 (46-131) Gemin6 (1-92)	This work
pET21a: G8	Gemin8 (1-242) (fl)	This work
pET28a: G8	Gemin8 (1-242) (fl)	This work
pET30: His_6-GST(TEV)-G8	Gemin8 (1-242) (fl)	This work
pET41: His_6-MBP(TEV)-G8	Gemin8 (1-242) (fl)	This work
pET21a: G8ΔN102	Gemin8 (103-242)	This work
pET28a: G8ΔN102	Gemin8 (103-242)	This work

pETM30: His_6-GST(TEV)-G8ΔN102 G7:G6	Gemin8 (103-242) Gemin7 (1-131) (fl) Gemin6 (1-167) (fl)	This work
pETM30: His_6-GST(TEV)-G8ΔN189ΔC12:G7ΔN30:G6ΔC75	Gemin8 (190-230) Gemin7 (31-131) Gemin6 (1-92)	This work
pETM30: His_6-GST(TEV)-G8ΔN189ΔC12:G7ΔN45:G6ΔC75	Gemin8 (190-230) Gemin7 (46-131) Gemin6 (1-92)	This work
pETM41*: His_6–MBP_SMN_$YG^{252-284}$	SMN (252-284)	This work
pETM41*: His_6–MBP_SMN_$YG^{263-284}$	SMN (263-284)	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$	SMN (252-284)	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{263-284}$	SMN (263-284)	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_M263R	SMN (252-284)/ M263R	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_M263T	SMN (252-284)/ M263T	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_S266P	SMN (252-284)/ S266P	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_S270E	SMN (252-284)/ S270E	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_Y272C	SMN (252-284)/ Y272C	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_H273R	SMN (252-284)/ H273R	This work
pETM41: His_6–MBP(TEV)-SMN_$YG^{252-284}$_T274I	SMN (252-284)/ T274I	This work
pET28a: His_6-Unrip	Unrip (1-350) (fl)	This work
pRK172: Sm D1D2	Sm D1 (1-119) (fl) Sm D2 (1-118) (fl)	Kambach et al., 1999
pET15b: Sm EFG	Sm E (1-92) (fl) Sm F (1-86) (fl) Sm G (1-76) (fl)	Dr. Christian Kambach, Germany
pQE30:His_6 (TEV)-Sm D3B(174)	Sm D3 (1-126) (fl) Sm B (1-174)	Kambach et al., 1999

3.1.7 Bacterial Strains

Strain	Genotype	Resistance	NEB** Catalog No.
E. coli DH5α	*fhuA2Δ(argF-lacZ)U169 phoA glnV44Φ80Δ(lacZ)M15gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	N/A	C2987I
E. coli BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3=λsBamHIoΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	N/A	C2527I
E. coli BL21 (DE3) pLysS pRARE*	pRARE (Cam^R)/*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λDE3=λsBamHIoΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pRARE (Cam^R)	Chloramphenicol	N/A

*The BL21 (DE3) pLysS pRARE strain was generated in-house, by transforming the phage resistant BL21 (DE3) cells with the pLysS pRARE (Cam^R) plasmid from Rosetta™ (DE3) cells (Novagen, USA).
**NEB: New England Biolabs, Germany.

3.1.8 Buffers and Solutions

3.1.8.1 *Antibiotics*

Antibiotic	Solvent	Stock concentration	Working Concentration
Ampicillin	ddH_2O	100 mg/mL	100 µg/mL
Kanamycin	ddH_2O	50 mg/mL	25 µg/mL
Chloramphenicol	Ethanol	50 mg/mL	50 µg/mL

3.1.8.2 *Bacterial cell culture*

Medium/Buffer/Solution	Composition
LB (Lysogeny Broth)-Luria medium	1.0 % (w/v) Tryptone 0.5 % (w/v) Yeast extract 1.0 % (w/v) Sodium chloride

LB-Agar plates	LB medium 1.5 % (w/v) Agar Desired antibiotics (Section 3.1.8.1)
Terrific Broth (TB) medium	1.2 % (w/v) Tryptone 2.4 % (w/v) Yeast extract 0.5 % (v/v) Glycerol
10X TB salts	0.17 M Monopotassium phosphate 0.72 M Dipotassium phosphate
TB medium additive	2 mM Magnesium chloride

3.1.8.3 *Agarose gel electrophoresis*

Buffer/Solution	**Composition**
Agarose gel	0.5 %-2.0 % (w/v) Agarose 1X TBE or TAE 5 µg/mL Ethidium bromide
5X TBE (Tris/Borate/EDTA)	445 mM Tris-HCl, pH 8.3 445 mM Boric acid 10 mM Na_2EDTA
50X TAE (Tris/Acetate/EDTA)	2 M Tris base 1 M Acetate 50 mM Na_2EDTA
5X DNA sample loading dye	10 mM Tris-HCl, pH 7.5 60 % (w/v) EDTA, pH 8.0 0.03 % (w/v) Xylene cyanol 0.03 % (w/v) Bromophenol blue

3.1.8.4 *SDS-PAGE*

3.1.8.4.1 Sample/Loading buffer

Buffer	**Composition**
6X SDS-PAGE protein sample buffer	300 mM Tris-HCl, pH 6.8 12 % (w/v) SDS 30 % (v/v) Glycerol 600 mM DTT 0.04 % (w/v) Bromophenol blue

3.1.8.4.2 SDS-PAGE Protein ladders

Protein Ladder	Vendor
PageRuler™ Unstained Protein ladder (200-10 kDa) (Catalog No. 22614)	Thermo Fisher Scientific, USA
PageRuler™ Prestained Protein ladder (170-10 kDa) (Catalog No. 22616)	Thermo Fisher Scientific, USA

3.1.8.4.3 Tris-Tricine gel system

Buffer/Solution	Composition
3X Gel Buffer	3 M Tris, pH 8.45 0.3 % SDS
Stacking gel	5 % (w/v) Acrylamide-bisacrylamide (29:1) 1X Gel buffer 0.0005 % (w/v) APS 0.0025 % (v/v) TEMED
Resolving gel	15 % (w/v) Acrylamide-bisacrylamide (29:1) 1X Gel buffer 0.0005 % (w/v) APS 0.005 % (v/v) TEMED
10X Anode buffer	2 M Tris base, pH 8.9
10X Cathode buffer	1 M Tris base, pH 8.25 1 M Tricine 1 % SDS

3.1.8.4.4 Tris- Glycine gel system

Buffer/Solution	Composition
Stacking gel	5 % (w/v) Acrylamide-bisacrylamide (37.5:1) 0.125 M Tris-HCl, pH 6.8 0.001 % (w/v) SDS 0.0005 % (w/v) APS 0.0025 % (v/v) TEMED
Resolving gel	8 % (w/v) Acrylamide-bisacrylamide (37.5:1) 0.375 M Tris-HCl, pH 8.8 0.001 % (w/v) SDS 0.0005 % (w/v) APS 0.005 % (v/v) TEMED

10X Tris-Glycine running buffer	0.25 M Tris base 1.92 M Glycine 1 % (w/v) SDS

3.1.8.4.5 Bis-Tris gel system

Buffer/Solution	Composition
Stacking gel	5 % (w/v) Acrylamide-bisacrylamide (37.5:1) 0.357 M Bis-Tris, pH 6.8 0.0005 % (w/v) APS 0.0025 % (v/v) TEMED
Resolving gel	13.5 % (w/v) Acrylamide-bisacrylamide (37.5:1) 0.357 M Bis-Tris, pH 6.8 0.0005 % (w/v) APS 0.005 % (v/v) TEMED
10X MES running buffer	0.5 M MES 0.5 M Tris base 1 % (w/v) SDS 10 mM EDTA

3.1.8.4.6 Protein gel staining with Coomassie stain

Solution	Composition
Staining solution	0.15 % (w/v) SERVA Blue R (SERVA Electrohoresis GmbH, Germany) 25 % Isopropanol 10 % Acetic acid
De-staining solution stacking gel	7.5 % Ethanol 7.5 % Acetic acid

3.1.8.5 *Autoradiography*

Solution	Composition
Amersham Amplify fluorographic reagent (Catalog No. NAMP100)	GE Healthcare, Germany
Amersham Hyperfilm MP (Catalog No.28906843)	GE Healthcare, Germany

3.1.8.6 *Protease Inhibitors*

Protease Inhibitors	Protease family targeted	Inhibitor type	Sol-vent	Stock Conc.	Work-ing Conc.
PMSF Phenylmethylsulfonyl fluoride	Serine Proteases	Irreversible	Ethanol	200 mM	0.2 mM
AEBSF 4-(2-amninoethyl) benzenesulfonyl fluoride hydrchlorid (Pefabloc SC-Prote-ase Inhibitor)	Serine Proteases	Irreversible	ddH_2O	100 mM	0.1 mM
Aprotinin	Serine Proteases	Reversible	ddH_2O	1 mM	0.1 mM
Leupeptin / Pepstatin A	Serine and Cysteine Pro-teases/ Aspartic Pro-teases	Reversible	DMSO	1 mM	0.1 mM

3.1.8.7 *Recombinant protein purification buffers*

Buffer	Composition
IMAC A (Bacterial pellet resuspension buffer/Ni-NTA resin equilibration buffer/Ni-NTA resin wash buffer)	50 mM HEPES-NaOH, pH 7.0 200 mM NaCl 25 mM Imidazole 5 mM 2-mercaptoethanol Proteases inhibitors (Sec-tion 3.1.8.6)
IMAC B (Ni-NTA Elution buffer)	50 mM HEPES-NaOH, pH 7.0 200 mM NaCl 250 mM Imidazole 5 mM 2-mercaptoethanol
GSH resin elution buffer	50 mM HEPES-NaOH, pH 7.0 150 mM NaCl 20 mM Glutathione 2 mM DTT

TEV digestion and Dialysis buffer	10 mM HEPES, pH 7.0 150 mM Imidazole 5 mM 2-mercaptoethanol
SEC buffer	10 mM HEPES, pH 7.0 150 mM Imidazole 5 mM 2-mercaptoethanol

3.1.8.8 *Quantitative Protein Assay*

Solution	Composition
Bradford Assay Solution (Bradford 1976)	1X Protein Assay Dye Reagent Concentrate (Bio-Rad, Germany) ddH_2O

3.1.9 Protein Chromatography

3.1.9.1 *Affinity Resins*

Affinity Matrix (Resin)	Purpose	Vendor
Ni-NTA Superflow	For Hexahistidine (His_6) fusion proteins	QIAGEN, Germany
Glutathione Sepharose™ 4 Fast Flow	For Glutathione S-Transferase (GST) fusion proteins	GE Healthcare, Germany
Amylose	For Maltose binding protein (MBP) fusion proteins	NEB, Germany

3.1.9.2 *Size Exclusion Chromatography (SEC) Equipment*

SEC Column	Resolution (M_r)*	Bed Volume	Purpose	ÄKTA System
Superdex™ 75 10/300 GL	3-70	24 mL	Analytical	ÄKTA Explorer
Superdex™ 200 10/300 GL	10-600	24 mL	Analytical	ÄKTA Explorer
Superose™ 6 10/300 GL	5-5000	24 mL	Analytical	ÄKTA Purifier

*M_r: Relative Molecular Weight in kDa

3.1.9.3 *Size Exclusion Chromatography Calibration Standards*

Calibration Standard	Molecular Weight (M_r)* (kDa)	Concentration used for calibration
Blue Dextran	>2000	1 mg/mL
Thyroglobulin	669	0.5 mg/mL
Apoferritin	443	0.25 mg/mL
B-Amylase	200	0.5 mg/mL
Alcohol Dehydrogenase	150	0.7 mg/mL
Bovine Serum Albumin	66	0.5 mg/mL
Ovalbumin	44	0.5 mg/mL
Carbonic Anhydrase	25	0.5 mg/mL

*M_r: Relative Molecular Weight

3.1.10 Protein Crystallization

3.1.10.1 *Materials and Equipment*

Materials/Equipment	Vendor
CrystalQuick™ LP plate (96 well) (Ref. No. 609171)	Greiner Bio-One, Germany
22 mm Circular cover slips-plain (Catalog.No. CSL-104)	Jena Bioscience GmbH, Germany
CryoMount Set (LithoLoops) (0.2/0.4 mm)	Molecular Dimensions Limited (MDL), UK
SuperClear™ plate (24 well) (Catalog.No. CPL-132)	Jena Bioscience GmbH, Germany
mosquito® crystal (Liquid handler)	TTP LabTech Inc, USA

3.1.10.2 *Crystallization Screens*

The following commercial screens were generated in-house by Emilia Gärtner (RG Fischer) using the manufacturer's recipes:

Screens	Vendor/Reference
Wizard™ Classic Screen	Molecular Dimensions Limited, UK
Wizard™ Classic Screen 2	Molecular Dimensions Limited, UK
Crystal Screen	Hampton Research, USA
Crystal Screen 2	Hampton Research, USA
Natrix	Hampton Research, USA
Natrix2	Hampton Research, USA
Midas™	Molecular Dimensions Limited, UK (Grimm et al. 2010)
JJS	Non- commercial/In-house prep

3.1.11 Software and Servers

Software/Server*	Provider/Reference
Wormbase, Version: WS265	www.wormbase.org
UniProtKB	(UniProt 2019)
PyMOL Molecular Graphics System	Schrödinger LLC, USA
Protparam*	(Gasteiger et al. 2003)
PSIPRED	(Jones 1999)
T-coffee Multiple sequence alignment	(Notredame, Higgins, and Heringa 2000)
Origin 8.6	OriginLab Corporation, USA
ESPript 3.0	(Robert and Gouet 2014)
ImageJ 1.52a	National Institutes of Health, USA
PHENIX 1.14 - 3247	(Adams et al. 2010)
BALBES 1.1.5	(Long et al. 2008; Robert and Gouet 2014)

3.2 Methods

3.2.1 Molecular Biology Methods

3.2.1.1 *Polymerase Chain Reaction (PCR)*

Open reading frames (ORFs) and/or DNA fragments of interest were PCR amplified using the appropriate oligonucleotides bearing the desired 5’ and 3’ restriction sites (Section 3.1.3). cDNA mix (for the *C. elegans* ORFs) or DNA plasmids (for *H. sapiens* genes, Chari et al., 2008) were used as PCR templates. The reaction set-up for a 50 µL PCR reaction mix using the KAPA Hifi DNA polymerase (Section 3.1.4.1) and the accompanying amplification programme was as follows:

PCR reaction set-up	
Template DNA	10-50 ng
Forward Oligo (10 µM)	2.5 µL
Reverse Oligo (10 µM)	2.5 µL
dNTP mix (10 mM)	2.5 µL
KAPA Buffer (5X)	10 µL
KAPA HiFi DNA Polymerase (1U/µL)	1 µL
ddH_2O	ad 50 µL

PCR programme		
Temperature	Time	Cycles
95 ºC	3 min	1x
95 ºC	25 s	30x
64 ºC	30 s	
72 ºC	15-60 s/kb	
72 ºC	1 min/kb	1x
4 ºC	∞	

Based on the size of the amplification products, the appropriate system of agarose gel electrophoresis (percentage of agarose, TBE/TAE; Section 3.1.8.3) was used. Following which, the amplicons were purified by a PCR clean-up commercial kit (Section 3.1.2).

3.2.1.2 *Restriction Digestion and Ligation*

Vectors and PCR products were restricted digested using the desired restriction enzymes (2U) (Section 3.1.4.1) in appropriate reaction buffers at 37°C overnight. The following table details the cloning sites used for generating the expression plasmids in this work:

Vector**	**Cloning sites (5'/3')**
pET21a	*Nde*I/*Xho*I
pET28a	*Nde*I/*Xho*I
pETM11	*Nco*I/*Xho*I
pETM13	*Nco*I/*Xho*I
pETM13_AV	*Nde*I/*Xho*I
pETM30_AV	*Nde*I/*Xho*I
pETM41*/pETM41	*Nco*I/*Not*I
pGEX6P1_AV	*Nde*I/*Xho*I

*/**Refer Section 3.1.5

These restricted digested samples were then purified by agarose gel electrophoresis (Section 3.1.8.3) with subsequent silica column purification (Section 3.1.2). Ligation was carried out with a 3-fold molar excess of insert relative to vector in the presence of T4 DNA ligase (0.5-1U) (Section 3.1.4.1).

3.2.1.3 *Poly-cistronic cloning*

A major advantage of the pET and pETM vector series (except for pETM41 vector) is the possibility of iterative cloning and re-shuffling of gene cassettes. Introduction of the *Nhe*I restriction site with the reverse oligo during PCR amplification (Section 3.2.1.1) was exploited to generate di-/tri-cistronic expression vectors. While, the recipient expression vector would be hydrolysed using the enzymes *Nhe*I and *Xho*I, the donor expression vector would be restricted digested with *Xba*I and *Xho*I enzymes (*Nhe*I and *Xba*I are isocaudamers). The resulting donor vector fragment would include the promotor and ribosome binding sequence (rbs) sequences along with the desired gene(s) (Figure 3.1).The thus generated DNA fragments would be purified by agarose gel electrophoresis and ligated (Section 3.2.1.2). The resultant expression plasmid can now serve as the recipient vector for other coding sequences.

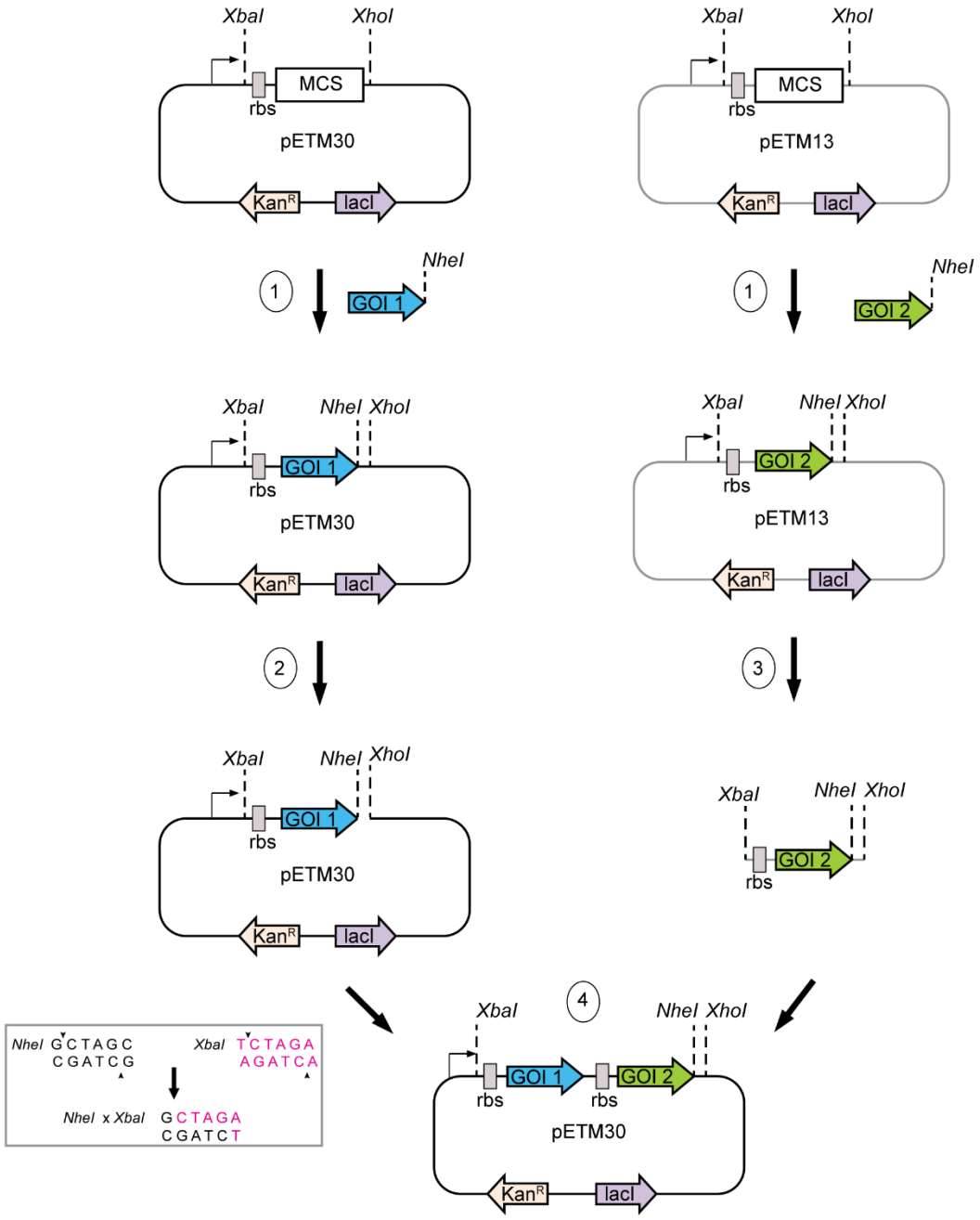


Figure 3.1 Poly-cistronic cloning strategy
Schematics of poly-cistronic cloning using the pET and pETM vector series. (1) Initially, in a donor vector (here, pETM30 having an N-terminal His 6 -GST tag is shown on the left) and a recipient vector (here, pETM13 is shown on the right), gene-of-interest 1 (GOI 1) and gene-of-interest 2 (GOI 2) are introduced in the multiple cloning site (MCS) by conventional cloning. Note that the *Nhe*I restriction endonuclease site is introduced in the 3' end of the GOI 1 and GOI 2 by PCR during the conventional cloning process. (2) The donor vector (left) is then linearized with the restriction endonucleases NheI and XhoI, (3) whereas the recipient vector is hydrolysed with the restriction endonucleases *Xba*I and *Xho*I to excise the GOI 2 with the 5' rbs (ribosome binding site) and T7 promoter sequence. (4) Since *Nhe*I and *Xba*I are isocaudamers, the fragment with GOI 2 obtained in (3) can be ligated with the linearized donor vector from step (2) (see inset on left), forming an expression cassette containing a N-terminal His 6 -GST-GOI 1 followed by GOI 2. Due to the inactivation of the restriction sites used for ligation, the resulting vector can iteratively serve as a donor vector (for *Nhe*I/*Xho*I linearization) for other coding sequences. A similar strategy can be followed for the pET vector series as well.

3.2.1.4 *Transformation, Screening for positive clones and DNA Sequencing*

For transformation of ligation products (Section 3.2.1.2), plasmids or mutagenesis products (Section 3.2.1.7), 100 µL of chemically competent *E. coli* cells (Section 3.1.7) was used. The chemically competent cells were thawed on ice, followed by incubation with the appropriate amount of plasmid/vector/product for 30 min Cells were transformed by heat shock at 42°C for 45 s and immediately transferred on ice to recover for 2-3 min. Subsequently, 800 µL LB was added and cells were incubated shaking (950 rpm) at 37°C for 60 min in a thermomixer. Subsequently, the cells were harvested by brief centrifugation for 1 min at 3,000 rpm and the concomitant cell pellet was resuspended with a small volume of the already present medium (~100 µL). This cell suspension was plated on LB agar plates containing the appropriate antibiotic (Section 3.1.8.1) for selection. These transformation plates were then incubated at 37°C over-night.

For screening of positive clones, single colonies from the transformation plates were used to inoculate 5 mL LB cultures with the appropriate antibiotic. Plasmids were then isolated and purified using the commercial kit (Section 3.1.2) based on the manufacturer's instructions.

The isolated plasmids were subsequently subjected to analytical restriction digestion and agarose gel electrophoresis. Positive clones, thus identified, were further verified by DNA sequencing (EUROFINS, Germany).

3.2.1.5 *Overlap Extension PCR*

Deletion mutagenesis was achieved by over-lap extension PCR. In the primary amplification round the two desired segments of the gene are generated independently by PCR (Section 3.2.1.1). This is accomplished by using a combination of the gene specific oligo with the mutagenizing oligo (each with complementary orientations). The mutagenizing primers with terminal complementarity ensures that the two independently amplified DNA fragments from the target gene can be fused into a single product by primer extension during the secondary PCR amplification round. The gene specific oligo provides the terminal restriction sites of choice (Vallejo, Pogulis, and Pease 2008). The resultant PCR inserts were assessed by agarose gel electrophoresis, followed by conventional cloning by restriction digestion and ligation (Section 3.2.1.2).

3.2.1.6 *Oligo-Annealing for construction of pGEX6P1_AV vector*

Single stranded sense and anti-sense oligonucleotides (Section 3.1.3) corresponding to the desired multiple cloning site (MCS) were generated with nucleotide overhangs for *EcoRI* and *XhoI* restriction sites (Figure 3.2). The annealing reaction was set-up by mixing 1 µL (100 pmol/µL) of each oligo with 8 µL of 1X annealing buffer (10 mM Tris, 7.5, 50 mM NaCl and 1 mM EDTA) in a total reaction volume of 10 µL. This reaction mix was then incubated at 95°C for 5 min, following which the sample was allowed to gradually cool-down to room-temperature. Subsequently, the sample was set-up for ligation with *Eco*RI/*Xho*I digested pGEX6P1 vector and proceeded with transformation and positive clone screening (Section 3.2.1.4).

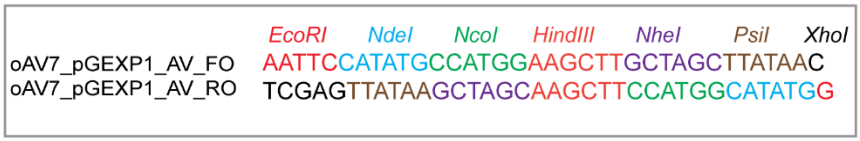


Figure 3.2 Generation of pGEX6P1_AV vector
The MCS of pGEX6P1 (Section 3.1.5) was replaced with a desired MCS by oligo-annealing to create the pGEX6P1_AV vector (Section 3.1.5). The forward and reverse oligonucleotides with 5' *Eco*RI (in red) and 3' *Xho*I (in black) overhangs flanking the desired MCS is illustrated here.

3.2.1.7 *Site directed mutagenesis*

Single point mutations of interest were introduced by the Quick-Change™ site-directed mutagenesis system (Stratagene, USA). Initially, two PCR reactions were set-up (Section 3.2.1.1), wherein the target plasmid was amplified with the appropriate mutant forward and reverse oligonucleotides (Section 3.1.3) separately with 10 cycles of amplification (Section 3.2.1.1). After this initial amplification, the products from these two independent PCR reactions were mixed 1:1, supplemented with 1U of KAPA Hifi DNA polymerase (Section 3.1.4.1) and 0.5 µL of 10 µM dNTP mix and subjected to another round of PCR amplification with 18 cycles. The resultant PCR product was treated with 2U of *Dpn*I enzyme, to remove the methylated paternal DNA template. 10-20 % of this reaction mix was transformed into *E. coli* DH5α cells (Section 3.1.7) and positive clones were identified by gene sequencing (Section 3.2.1.4).

3.2.2 Protein Biochemistry Methods

3.2.2.1 *Recombinant protein expression and purification – Part I*

Sequential (co-expression) or single transformation of mono-, di-or tri-cistronic expression plasmids was performed as described (Section 3.2.1.4).The choice of bacterial

expression strains (Section 3.1.7) varied: *E. coli* BL21 (DE3) cells were used for both the *S. pombe* and *H. sapiens* orthologues' expression and *E. coli* BL21 (DE3) pLysS pRARE cells for the *C. elegans* expression plasmids. The transformed bacterial cells were cultured in TB medium (Sections 3.1.8.1 and 3.1.8.2) until the OD_{600} reached 0.8 units. Expression was then induced by addition of 0.5 mM Isopropyl β-D-1-thio-galactopyranoside (IPTG) and carried out at 15°C for 18 h.

The bacterial cells were harvested by centrifugation at 5000 rpm for 20 min and the resultant cell pellets were re-suspended in IMAC A buffer (Section 0) supplemented with protease inhibitors (Section 3.1.8.6). Following cell lysis by sonification (Sonifier® 250-Branson, USA), the lysate was clarified by ultracentrifugation for 45 min and 45,000 rpm at 4°C (Rotor 45Ti, Beckman Coulter, Krefeld, Germany). The concomitant supernatant was incubated with either Ni-NTA or GSH resin (Section 3.1.9.1) pre-equilibrated with IMAC A, followed by extensive washing of the resin with buffer IMAC A and subsequent elution (IMAC B or GSH elution buffer) (Section 0). The eluates were treated with TEV or PreScission™ protease (Section 3.1.4.2) at 4°C for tag removal (wherever feasible). TEV protease cleavage was performed during simultaneous dialysis (Section 3.1.8.7) with a ratio of 1:50 of protease: protein. PreScission™ protease cleavage was achieved by supplementing the protein bound resin (made to a 1:1 slurry with IMAC A buffer) with 1:20 ratio (protease:protein). Eluates/concomitant flow-through fractions (after tag removal) containing the proteins were concentrated and subjected to size exclusion chromatography (SEC) (Section 3.1.9.2). Peak fractions were analysed by SDS-PAGE (Section 3.1.8.4) and the desired fractions were pooled, concentrated, flash-frozen and stored at -80°C until further use.

3.2.2.2 *Recombinant protein expression and purification – Part II*

The expression, purification and/or reconstitution of the following protein complexes (for the corresponding expression plasmids, see Section 3.1.6) was followed as described in the accompanying references:

Recombinant Proteins	References
Sm D1/D2	Kambach et al.,1999
Sm E/F/G	Kambach et al.,1999
Sm B^{1-174}/D3	Kambach et al.,1999
6S complex	Grimm et al., 2013
Sm sub-core	Chari et al., 2008
Dm Gemin2	Grimm et al., 2013

3.2.2.3 *In vitro transcription and translation of Gemin8*

C. elegans (*Ce*) Gemin8 (fl) and Gemin8$^{57-199}$ (ΔN) variant constructs were *in vitro* transcribed and translated with [^{35}S]-Methionine labelling with the TNT® T7 Quick coupled Transcription/Translation system according to the manufacturer's instructions (Promega, Germany).

3.2.2.4 *In vitro protein binding assays*

In case of the GST binding assays, recombinant GST fusion proteins were immobilized on GSH beads (Section 3.1.9.1) and incubated with two molar excess of Sm hetero-oligomers in binding buffer (HEPES, pH 7.0, 150 mM NaCl, 2 mM DTT and protease inhibitors) at 4°C for 2 h. Subsequently, the beads were washed extensively with the binding buffer by centrifugation at 5,000 rpm for 5 min at 4°C. The bound proteins were then eluted with the GSH elution buffer (Section 0) and resolved by SDS-PAGE (Tris-Tricine system, Section 3.1.8.4.3) and visualized by Coomassie staining (Section 3.1.8.4.6).

For the MBP binding assays, 40 µg of MBP fusion proteins immobilized on 20 µL Amylose resin were incubated with *in vitro* transcribed and *in vitro* translated [^{35}S]-Methionine labelled Gemin8 transcripts in binding buffer (HEPES, pH 7.0, 150 mM NaCl, 2 mM DTT and protease inhibitors) at 4°C for 3 h. The beads were then washed initially with a high salt buffer (HEPES, pH 7.0, 300 mM NaCl, 2 mM DTT and protease inhibitors) followed by washes with the binding buffer. Bound proteins were then eluted with 1X SDS sample buffer, resolved by SDS-PAGE (Bis-Tris system,

Section 3.1.8.4.5) and analysed by Coomassie staining. [^{35}S]-Met labelled proteins were detected by autoradiography of the dried gel.

3.2.3 X-ray Crystallography

3.2.3.1 *Protein Crystallogenesis*

Diffraction quality protein crystals were obtained following trials with drop-size, precipitant and pH variations of the initial crystallization condition (reservoir solution) at the indicated protein concentration (prior to 1:1 reservoir solution mixing) by sitting-drop vapour diffusion at 18°C:

Protein/Complex	**Concentration**	**Crystallisation Condition (Reservoir Solution)**
Hs Gemin7$^{31\text{-}131}$:6 (fl)	20 mg/mL	100 mM Sodium acetate, pH 4.7 18 % (w/v) 2-methyl-2, 4-pentanediol (MPD)
Hs Gemin8$^{191\text{-}230}$:7$^{46\text{-}131}$:6$^{1\text{-}92}$	30 mg/mL	100 mM 2-(N-morpholino) ethanesulfonic acid 200 mM NaCl 30 % Jeffamine ED2003
Sp Gemin8ΔN30ΔL:7:6	30 mg/mL	50 mM HEPES, pH 6.0 40 mM Magnesium Acetate 30 % (w/v) 2-methyl-2, 4-pentanediol (MPD)
Ce MBP*-YG$^{182\text{-}207}$	20 mg/mL	35 % Glycerol ethoxylate 200 mM Lithium citrate

The protein crystals were cryo-protected by gradual transfer from the reservoir solution (see above) to the following cryo-protectant solutions before being flash frozen in liquid nitrogen:

Protein/Complex	**Cryo-protectant**
Hs Gemin7$^{31-131}$:6 (fl)	100 mM Sodium acetate, pH 4.7 35 % (w/v) 2-methyl-2, 4-pentanediol (MPD)
Hs Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$	100 mM 2-(N-morpholino) ethanesulfonic acid 200 mM NaCl 35 % Jeffamine ED2003
Sp Gemin8ΔN30ΔL:7:6	50 mM HEPES, pH 6.0 40 mM Magnesium Acetate 40 % (w/v) 2-methyl-2, 4-pentanediol (MPD)
Ce MBP*-YG$^{182-207}$	35 % Glycerol ethoxylate 200 mM Lithium citrate 15 % Polyethylene glycol (PEG)

3.2.3.2 *Data collection, structure determination and refinement*

The data sets for each of the protein crystals were collected at either the ESRF (Grenoble, France) or the HZB (Berlin, Germany) beamlines and processed with XDS (Kabsch 2010). The structures were solved either by molecular replacement with PHASER (McCoy et al. 2007) and/or *de novo*. Automated refinement was performed in PHENIX (Section 3.1.11) until R/R$_{free}$ factors converged. The crystallographic data, processing and refinement parameters are summarized in Tables 4.2 and 4.3. Structure determination and processing was executed by Dr. Clemens Grimm (RG Fischer) from the *Hs* Gemin7$^{31-131}$:6 (fl), *Hs* Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$ and *Ce* MBP*-YG$^{182-207}$ protein crystals. Structural elucidation attempts for the second *Sp* Gemin8ΔN30ΔL:7:6 data set was performed by Aravindan Viswanathan (RG Fischer). The final figures were generated using PyMOL (Schrödinger, LLC). For individual contributions pertaining to structural elucidations, refer Appendix 7.5.

4 Results

4.1 Introductory notes

The *Homo sapiens* (*Hs*) SMN complex has been reconstituted *in vitro* using the combined application of bacterial and insect cell expression systems (Neuenkirchen et al. 2015). However, the highly oligomeric nature of this complex, aggregation-prone propensity and low yields have been handicapping for X-ray crystallographic studies till date. In an attempt to circumvent these inherent difficulties and to establish sub-complexes with a high crystallization propensity, this dissertation work initially employed the nematode *Caenorhabditis elegans* (*Ce*) SMN complex alongside their human counterparts for biochemical and crystallographic studies.

A homology search for the protein components constituting the SMN complex inventory in the nematode genome identified SMN and Gemins 2, 3, 6, 7 and 8 orthologues while Gemins 4-5 and unrip were lacking (Figure 4.1 A) (Kroiss et al. 2008). With a modular mode of interactions constituting the human SMN complex (Figure 4.1 B) (Otter et al. 2007), an analogous scenario is likely for the nematode SMN complex (Figure 4.1 C). In common with its human counterparts, mutual biochemical interactions of *Ce* SMN and *Ce* Gemin2 has been established by yeast two-hybrid assay and co-immunoprecipitations (Burt, Towers, and Sattelle 2006; Miguel-Aliaga et al. 2000). However, the interactions of the Gemins 3, 6 and 7 and 8 within the nematode SMN complex remains to be characterised (Figure 4.1 B).

In conjunction with the nematode and human SMN sub-complexes, the recently established *Schizosaccharomyces pombe* SMN complex (unpublished data, RG Fischer) was probed (in part) to define protein interaction domains for crystallographic pursuits (Section 4.5.3).

Notably, sequence alignments and secondary structure prediction algorithms of the SMN, Gemin2 and Gemins 6-8 orthologues revealed pronounced conservation of functional regions/domains essential for modular interactions within the SMN complex. Furthermore, these orthologues feature shorter intermittent loops and unstructured regions and are thus amenable for protein crystallographic studies.

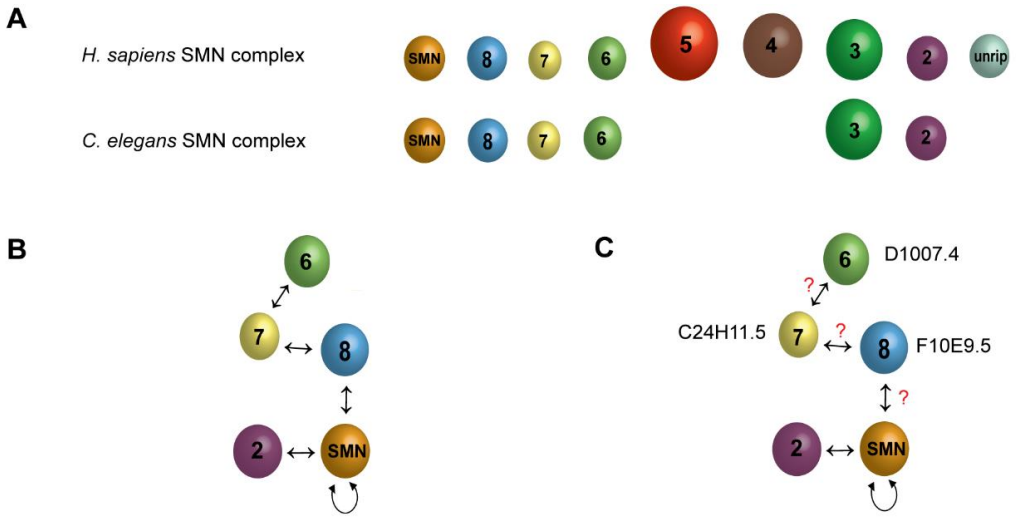


Figure 4.1 Conservation of the multi-subunit SMN complex in *H. sapiens* and *C. elegans*
(A) Screening of the complete *C. elegans* revealed most of the orthologues of human SMN complex components barring Gemins 4, 5 and unrip (Kroiss et al., 2008).
(B) Schematics of all interactions within a minimal *H. sapiens* SMN complex (Otter et al., 2007) shown to be sufficient orchestrating Sm core assembly and its concomitant transfer to U snRNA *in vitro* (Chari et al., 2008 and Neuenkirchen et al., 2015).
(C) Schematics of all putative interactions that could occur within a *C. elegans* SMN complex (lacking Gemin3) analogous to the pentameric *H. sapiens* SMN complex **(B)**. Thus far, only the *C. elegans* orthologues of SMN-Gemin2 interaction has been reported (Burt et al., 2006). Note: The Wormbase database gene accession IDs for the putative *C. elegans* Gemins 6, 7 and 8 orthologues are indicated alongside the said Gemins.

4.2 The Gemin7:6 sub-complex within the SMN complex

4.2.1 The *Homo sapiens* (*Hs*) Gemin7:6 heterodimer

Initially, the human (*Hs*) Gemins 7 and 6 proteins of the SMN complex inventory were investigated via structural and biochemical means to establish compliance with previous studies (Ma et al. 2005; Otter et al. 2007). The earlier reported structural model of the truncation variant Gemin7$^{31\text{-}131}$ together with full-length (fl) Gemin6 revealed a sub-complex that dimerized in a head-to-tail manner. Although within this structural model the electron density for the largely unstructured C-terminal domain of Gemin6 (residues 87-167) was not sufficiently well defined, the essential interaction domains were defined (Figure 4.2 A) (Ma et al. 2005).

Centred on these findings, recombinant full-length (fl) and several associated truncation variants of *Hs* Gemins 7 and 6 (Figure 4.2 B) were biochemically assessed for Gemin7:6 sub-complex formation by analytical SEC and SDS-PAGE (Figure 4.2 C).

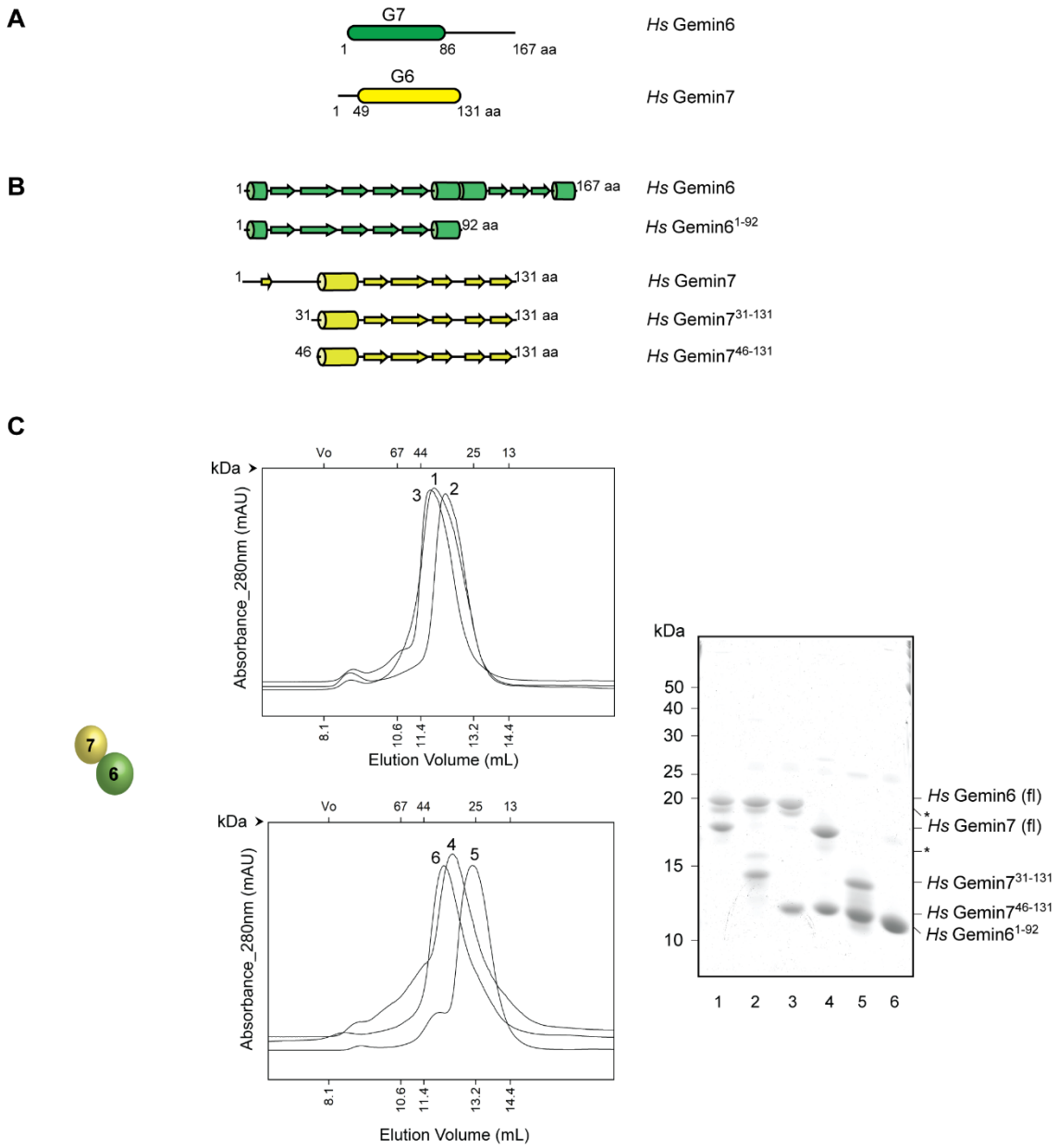


Figure 4.2 Recapitulating the *H. sapiens* Gemin7:6 sub-complex
(A) Schematics of domain organization in *H. sapiens* (*Hs*) Gemins (G) 7 and 6 as reported by Ma et al., 2005. The Sm-like fold domains that interact to form the dimeric Gemin7:6 sub-complex are indicated.
(B) Schematic representation of the *Hs* Gemin6 and Gemin7 full-length (fl) and associated truncation mutants secondary structural elements (based on PSIPRED algorithm output). The α-helices, β-sheets and coils are displayed as cylinders, arrows and straight lines respectively.
(C) Purification of *Hs* Gemin7:6 (fl) dimer and associated truncation mutants. Recombinantly reconstituted protein complexes were analysed by SEC (HiLoad® 13/300 Superdex® 75 GL) and SDS-PAGE analysis. *Left panel*: Schematic of the formed dimeric protein complex; *Middle panel*: SEC elution profiles of the reconstituted Gemin7:6 dimeric complexes and associated truncation mutants. Upper middle panel shows the SEC elution profiles of Gemin7:6 (fl) [M_r = 33.5 kDa, V_E = 11.74 mL], Gemin7$^{31-131}$:6 (fl) [M_r = 30.4 kDa, V_E = 12.23 mL] and Gemin7$^{46-131}$:6 (fl) [M_r = 28.7 kDa, V_E = 11.84 mL] respectively. The lower middle panel shows the elution profiles of Gemin7 (fl):6$^{1-92}$ [M_r = 25.2 kDa, V_E = 12.40 mL], Gemin7$^{31-131}$:6$^{1-92}$ [M_r = 22.1 kDa, V_E = 13.07 mL] and Gemin7$^{46-131}$:6$^{1-92}$ [M_r = 20.4 kDa, V_E = 12.09 mL] complexes labelled 1-6 respectively. The elution positions of globular molecular weight calibration standards are indicated on the upper X-axis for reference; *Right panel*: SDS-PAGE of the SEC elution peak fractions of the corresponding protein complexes (lanes 1-6) visualized on a coomassie stained 15 % Tris-Tricine gel. Bands marked with an asterisk indicate protein degradation products.

Markedly, the Gemin7$^{46-131}$:Gemin6 (fl) (Figure 4.2 C, upper middle panel, profile 3) and Gemin7$^{46-131}$:Gemin6$^{1-92}$ (Figure 4.2 C, lower middle panel, profile 6) complexes elute at an earlier elution volume (near the 44 kDa marker) than predicated (near the 25 kDa marker) (Figure 4.2 C, lower middle panel, profile 6). Although size exclusion chromatography (SEC) analysis cannot be decisive, these elution profiles suggest either a molecular mass equivalent to a hetero-tetramer or that these subcomplexes adopt an elongated shape and consequently a large Stokes radius. The former notion is more appealing given the speculated self-association propensity for these Gemins (Ma et al. 2005) arising due to their similar structural architecture. Figure 4.2 C (right panel) depicts the SDS-PAGE analysis of the SEC peak fractions of the established Gemin7:6 complexes (lanes 1-6) demonstrating complexes' integrity.

To ascertain the crystallisation propensity of the established Gemin7:6 subcomplex variants, crystallization trails employing the in-house screens were pursued (Section 3.1.10.2). However, only the previously crystallized Gemin7$^{31-131}$:6 (fl) subcomplex (Figure 4.2 C, right panel, lane 2) yielded diffraction quality crystals under the reported parameters albeit at a higher precipitant concentration (18 % 2-methyl-2, 4-pentanediol, MPD) (Section 3.2.3) (Ma et al. 2005). Using the previously published atomic resolution structure (PBD ID: 1y96) as a molecular replacement model (Section 3.2.3.1), a solution with two instances of the Gemin7:6 dimer in the asymmetric unit (AU) was obtained with subsequent refinement to a 3.1 Å resolution by Dr. C. Grimm (RG Fischer) (Figure 4.3 A). The crystallographic data, results of phasing and refinement are summarized in Table 4.2 and Table 4.3 respectively.

In contrast to the published structure belonging to the crystallographic group $P4_22_22_2$, the current crystal form belongs to the crystallographic group $P6_122$ and features a significantly different crystal packing since the former group represents an orthorhombic lattice symmetry and the latter is of hexagonal symmetry. The final refined model includes residues 1-86 and 49-131 for Gemin6 and Gemin7 respectively. Both Gemins exhibit a similar architecture with a five stranded β-sheet flanked by N-terminal α helices, reminiscent of the Sm-fold in Sm proteins (Figure 4.3 A). Notably, with the limited atomic resolution of the present model (Figure 4.3 B, in black), Gemin7's β3 (Asp111-Gln106) and β4 (Ala115-Arg120) strands were refined slightly shorter than the earlier structural model (β3 beginning with Asp111-Gln108 and β4 having Val113-Arg120) (Figure 4.3 B, in red) (Ma et al. 2005). Also, the α1 helix of Gemin6 appears to be slightly less well-defined comprising of only residues Phe9-Tyr15, in contrast to

an earlier reported extended helix of residues Ser2-Tyr15 (Figure 4.3 B) (Ma et al. 2005). Nevertheless, the independent hetero-dimers interact via the β4:β5 interface of Gemins 6 and 7 respectively, affording to form a continuous 10 stranded β sheet (Figure 4.3), concordant with the previously published structural model (Section 1.3, Figure 1.11 C) (Ma et. al., 2005). No significant differences could be discerned between the two structures given an overall RMSD (root mean squared deviation of C_α atomic position co-ordinates) value of 0.613 (RMSD <1.0 implies largely identical in conformation) (Figure 4.3 B) (Kufareva and Abagyan 2012).

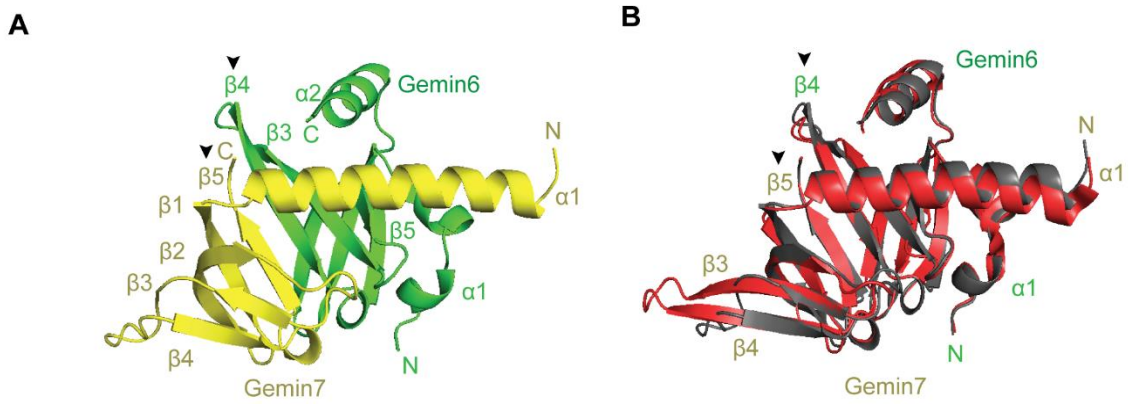


Figure 4.3 *H. sapiens Hs* Gemin7:6 sub-complex structure
(A) Structural model of *Hs* Gemin7:6 sub-complex, with Gemin7 and Gemin6 depicted in yellow and green respectively. The dimerization interface involves the Gemin7 β5-Gemin6 β4 strands (indicated with arrow heads) forming a continuous 10 sheet β-barrel. The respective N- and C-termini are labelled.
(B) Superposition of the Gemin7:6 dimer (in red) from Ma et al., 2005 (PDB ID: 1y96) and the current model (in black). The dimerization interface appears intact (indicated with arrow heads). Minor differences (as indicated) are observed in the regions of β3 and β4 strands of Gemin7, resulting in a more compact β sheet. Also, the α1 helix of Gemin6 demonstrates variation in the structural model (see text for details).

4.2.2 Characterisation of the nematode *C. elegans* Gemin7:6 sub-complex

Encouraged by the effective biochemical and structural recapitulation of the previously reported *H. sapiens* dimeric Gemin7:6 sub-complex (Section 4.2.1), an equivalent characterisation of the presumed *C. elegans* Gemin7 (C24H11.5, isoform b) and Gemin6 (D1007.4) orthologues was pursued. Alignment of the nematode candidate orthologues with their human counterparts revealed comparable and/or conserved residues that could contribute to identical secondary structural elements as shown in Figure 4.4.

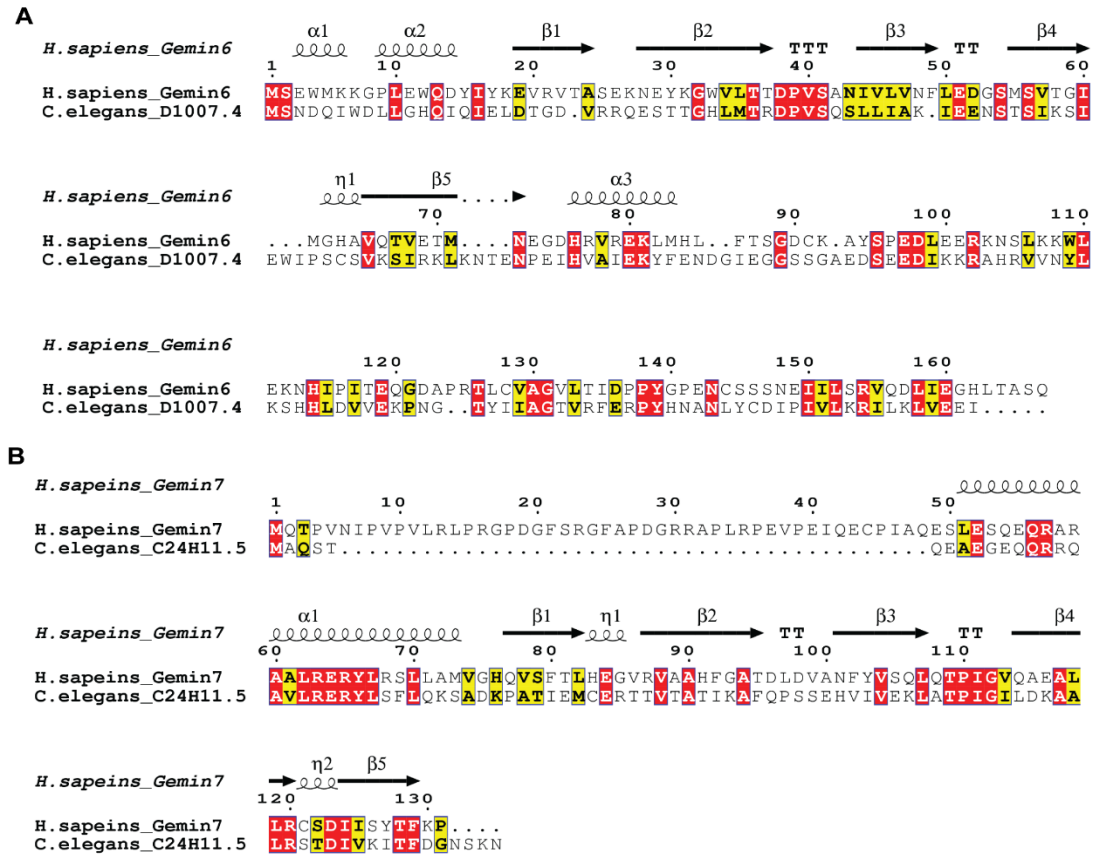


Figure 4.4 Conserved regions of the putative *C. elegans* Gemin6 and 7 orthologues
Alignment of the *C. elegans* putative orthologues D1007.4 and C24H11.5 (isoform b) with *H. sapiens* Gemin6 and Gemin7 in **(A)** and **(B)** respectively. Secondary structure elements of the *H. sapiens* Gemins available from the previously described structural model in Section 4.2 are shown above the sequences. α-helices and 3_{10}-helices (η) are displayed as squiggles, β-strands are rendered as arrows, strict β-turns as TT letters and strict α-turns as TTT. Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).

Particularly, the N-terminal region of *Hs* Gemin7 (residues 1-31) preceding strand β1 is lacking in the nematode counterpart. Also, as described in Section 4.2.1(Figure 4.2 C and Figure 4.3 A), with the C-terminal domain of *Hs* Gemin6 (residues 87-167) being expendable for Gemin7 interaction, a similar truncation mutant of *Ce* Gemin6, *Ce* Gemin6$^{1\text{-}94}$ was included for crystallographic pursuits. Thus, centred on these observations, the recombinant *Ce* Gemin7:6 (fl) and Gemin7:6$^{1\text{-}94}$ sub-complexes (Figure 4.5 A) would be analogous to the human crystallized experimental construct, *Hs* Gemin7$^{31\text{-}131}$:6$^{1\text{-}92}$ (fl) (Figure 4.2 C, lane 2) and the resultant structural model, *Hs* Gemin7$^{46\text{-}131}$:6$^{1\text{-}92}$ (Figure 4.2 C, lane 6 and Figure 4.3 A) respectively. The recombinant *Ce* Gemin7:6 sub-complexes displayed mono-disperse SEC elution profiles on the analytical HiLoad® 13/300 Superdex® 200 GL column, albeit their early retention volumes in comparison to globular protein standards (Fig-

ure 4.5 B, middle panel). This could be plausibly attributed to interactions of independent Gemin7:6 dimers to form tetramers, given their identical structural architecture (Ma et al. 2005). Unfortunately, neither the *Ce* Gemin7:6 (fl) nor the Gemin7:6$^{1-94}$ sub-complexes resulted in protein crystals for crystallographic studies. As evidenced in Figure 4.5 B, for the *Ce* Gemin7:6$^{1-94}$ sub-complex (lane 2), the C-terminal region of *Ce* Gemin6 (residues 95-168) is expendable for its interaction with Gemin7, thereby conforming to identical biochemical characteristics with the *H. sapiens* sub-complex (Section 4.2.1, Figure 4.2 B).

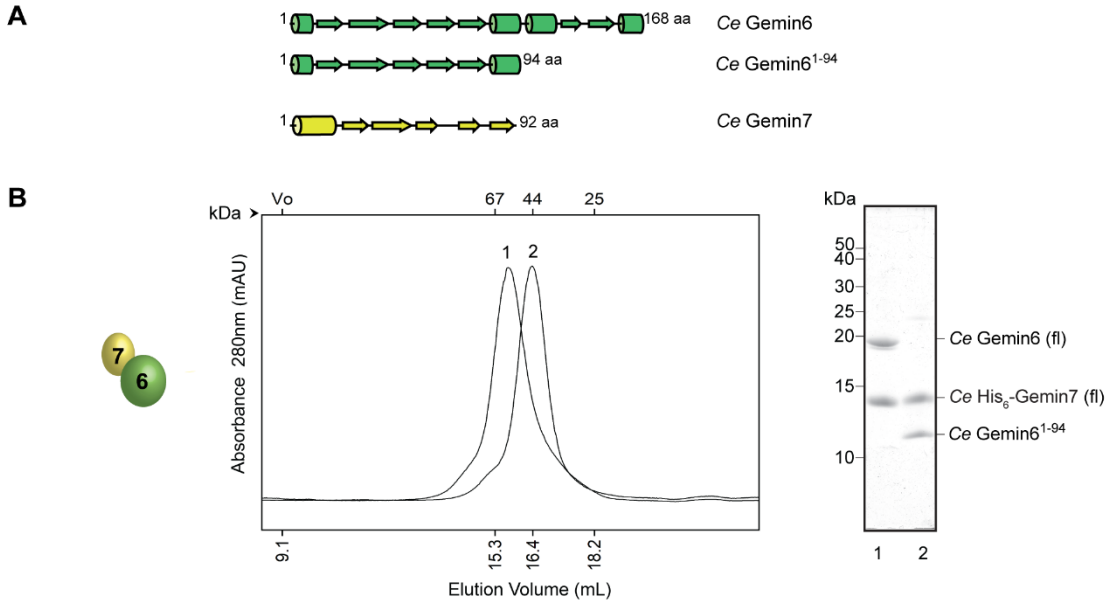


Figure 4.5 Reconstitution of the *C. elegans* Gemin7:6 dimeric complex
(A) Schematics of domain architecture of *C. elegans* (*Ce*) Gemin6 and Gemin7 full-length (fl) orthologues and associated truncation mutant. The secondary structural elements (based on PSIPRED output): α-helices, β-sheets and coils are depicted as cylinders, arrows and straight lines respectively, to scale.
(B) Purification of *Ce* Gemin7:6 protein complex. Protein orthologues co-expressed in bacterial cells were initially purified by IMAC followed by SEC (HiLoad® 13/300 Superdex® 200 GL). *Left panel*: Schematic of formed protein complex; *Middle panel*: SEC elution profiles of the reconstituted *Ce* Gemin7:6 (fl) [M_r = 31.5 kDa, V_E = 15.68 mL] and Gemin7:6$^{1-94}$ [M_r = 22.4 kDa, V_E = 16.38 mL] complexes indicated as (1) and (2) respectively. The elution positions of the conventional globular proteins used as SEC calibration standards are indicated on the upper X-axis for reference; *Right panel*: the corresponding complexes (lanes 1 and 2) visualized on a coomassie stained 15 % Tris-Tricine gel.

4.3 The Gemin8:7:6 sub-complex

4.3.1 The nematode Gemin8:7 interaction domains

The peripheral Gemin7:6 module is incorporated into the heteromeric SMN complex via its association with Gemin8 (Otter et al. 2007). Intending to gain structural insights on this mode of interaction, the trimeric *Ce* Gemin8:7:6 module was established. The putative *C. elegans* Gemin8 orthologue F10E9.5 (Figure 4.6 A) was co-expressed in

bacterial cells as an N-terminal hexa-histidine fusion protein with *Ce* His_6 -7:6 sub-complex. The supposed *Ce* Gemin8 orthologue indeed forms a trimeric complex with *Ce* Gemin7:6 eluting between the 67 kDa - 44 kDa markers on the analytical SEC column (Figure 4.6 B, middle panel, elution profile 2 in black) consistent with the predicated molecular weight of 55.2 kDa. The trimeric *Ce* Gemin8:7:6 elutes slightly earlier (V_E = 15.36 mL) than *Ce* Gemin7:6 sub-complex (V_E = 15.68 mL) with the latter behaving as a hetero-tetramer in solution (Section 4.2.2). However, the broad elution peak of the trimeric complex overlays partly with the SEC elution profile of the Gemin7:6 complex suggestive of a heterogeneous population of sub-complexes in solution: The highly degradation prone *Ce* Gemin8 (fl) within the Gemin8:7:6 trimeric complex (Figure 4.6 B, right panel, lane 2), influences the stoichiometry of the individual Gemins leading to either Gemin6-8 trimeric or Gemin7-6 species.

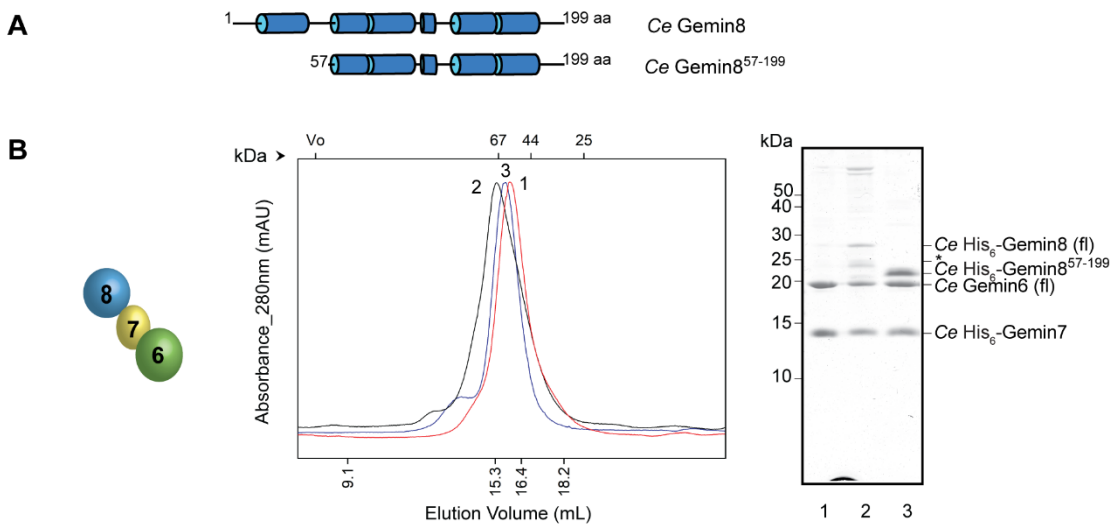


Figure 4.6 *C. elegans* Gemin8:7:6 sub-complex
(A) Schematic representation of *C. elegans* (*Ce*) Gemin8 full-length (fl) orthologue and associated N-terminal truncation mutant. The secondary structural elements (based on PSIPRED output): the α-helices and coils are depicted as cylinders and straight lines respectively, to scale.
(B) IMAC purified *Ce* Gemin8:7:6 and associated truncation mutant trimeric complexes were assessed by SEC (HiLoad® 13/300 Superdex® 200 GL). *Left panel*: Schematic of formed protein complex; *Middle panel*: Overlay of SEC elution profiles of the reconstituted *Ce* Gemin8:7:6 (fl) (in black (2); M_r = 55.2 kDa, V_E = 15.36 mL) and Gemin8$^{57-199}$:7:6 (in blue (3); M_r = 50.3 kDa, V_E = 15.54 mL) complexes in comparison to *Ce* Gemin7:6 complex (in red (1); M_r = 31.5 kDa, V_E = 15.68 mL). The elution positions of the globular protein SEC calibration standards are indicated on the upper X-axis for reference; *Right panel*: Coomassie stained SDS-PAGE gel (15 % Tris-Tricine) of the reconstituted trimeric protein complexes, *Ce* Gemin8:7:6 (fl) (lane 2) and *Ce* Gemin8$^{57-199}$:7:6 (lane 3) with *Ce* Gemin7:6 (lane 1) as a reference. Asterisk denotes protein degradation products.

Otter et al. in 2007 first demonstrated that the C-terminus of Gemin8 (residues 128-242) harbours the interaction interface for Gemin7 within the *Hs* SMN complex. A similar scenario could be conceived for the nematode counterpart. Indeed, the N-terminal truncation mutant, Gemin8$^{57-199}$ (Figure 4.6 A) appears to retain its contact with

the Gemin7:6 sub-complex (SEC and SDS-PAGE analysis shown in Figure 4.6 B), thus implying that the predicted nematode Gemin8 orthologue also contacts the *Ce* Gemin7 directly via its C-terminal region. Although, *Ce* Gemin7:6 behaves as a hetero-tetramer in solution as described in Section 4.3, this is not evident in the presence of Gemin8 (Figure 4.6 B). Notably, within the trimeric *Ce* Gemin8$^{57\text{-}199}$:7:6 sub-complex, *Ce* Gemin8$^{57\text{-}199}$ exhibited increased protease susceptibility in limited proteolysis experiments (data not shown). Centred on these findings and to enhance crystallization propensity, additional N-terminal *Ce* Gemin8 truncation mutants, Gemin8$^{119\text{-}199}$, Gemin8$^{145\text{-}199}$, Gemin8$^{145\text{-}187}$ (Figure 4.7 A) were generated as N-terminal GST fusion proteins. Notably, the truncation mutant Gemin8$^{145\text{-}187}$ is similar to Gemin8$^{145\text{-}199}$ variant, only that the former is devoid of the terminal 12 amino acids (188-199 residues) predicted to be unstructured by secondary structure prediction algorithms (Figure 4.7 A). These truncation mutants were then analysed biochemically for their ability to interact with Gemin7.

To this end, following bacterial co-expression of GST-Gemin8 truncation mutants with His_6-Gemin7:6 sub-complexes, these protein sub-complexes were first captured on a GSH resin (Figure 4.7 B, lanes 3, 5 and 7) with Gemin8 truncation mutants serving as the bait to capture the interacting Gemin7:6 sub-complex (Figure 4.7 B, lane 1). Subsequent to GST-tag removal (by PreScission^{TM} protease, Sections 3.1.4.2 and 3.2.2.1), the concomitant flow-through was then incubated on Ni-NTA resin. At this stage, engaging His_6-Gemin7 as the bait it was presumed that only those specific Gemin8 truncation mutants retaining their Gemin7 interaction domain occur in the elution fractions (Figure 4.7 B, lanes 4, 6 and 8). This two-step purification strategy revealed that the C-terminal 145-187 residues of *Ce* Gemin8 alone is adequate for Gemin7 interaction.

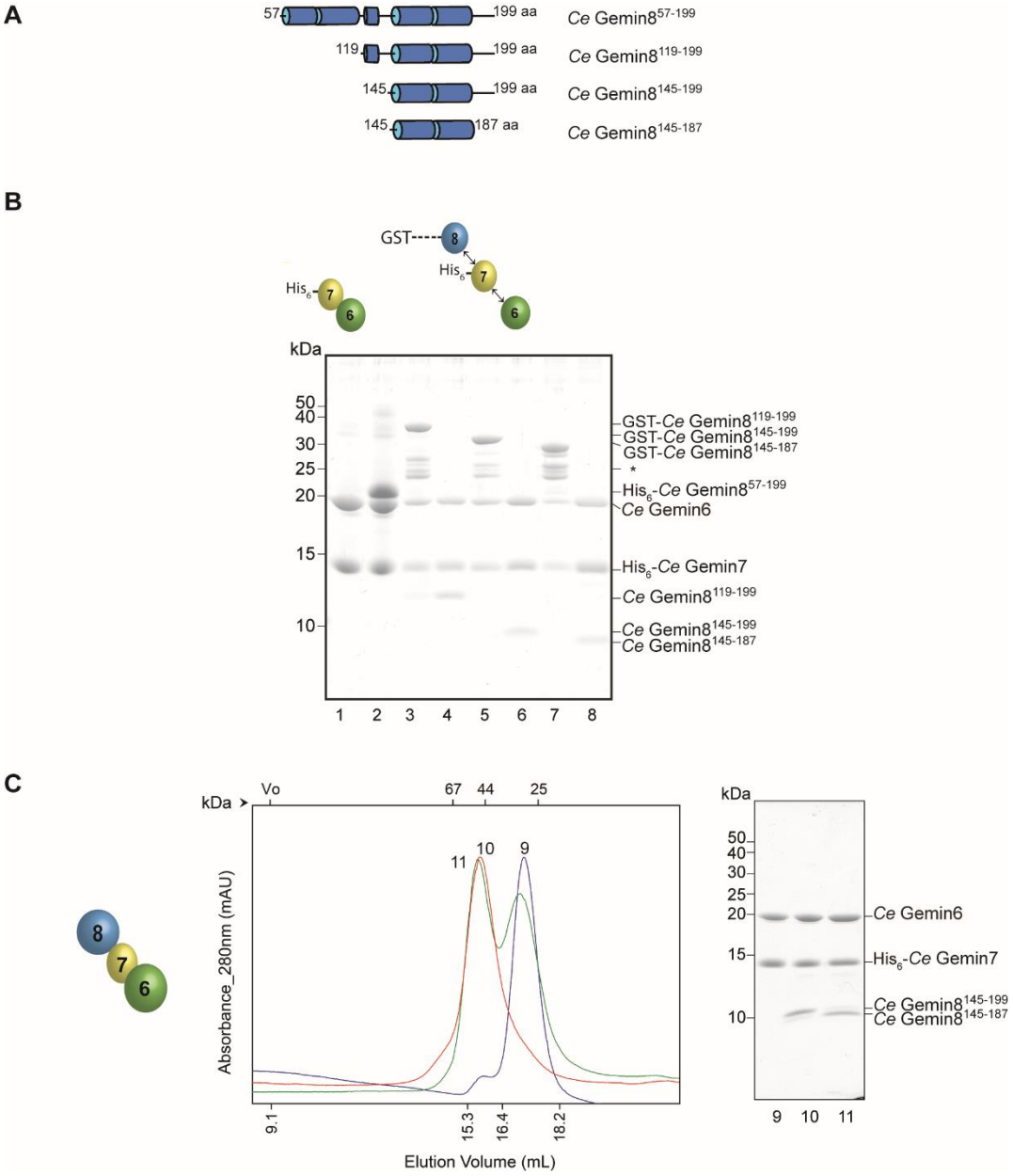


Figure 4.7 Mapping of the *C. elegans* Gemin8:7 interaction interface
(A) Schematic representation of *C. elegans* (*Ce*) Gemin8 N-terminal truncation mutants used. The secondary structural elements (based on PSIPRED output): the α-helices and coils are rendered as cylinders and straight lines respectively, to scale.
(B) An IMAC strategy (see schematic above the SDS-PAGE gel) to identify the minimal region of Gemin8 required for its Gemin7 interaction. GST fusion Gemin8 truncation mutants depicted in (A) were co-expressed with His_6-Gemin7:6 dimer in bacterial cells. The trimeric Gemin8:7:6 complex was first captured by GSH affinity resin (lanes 3, 5 and 7), followed by PreScission™ cleavage of the GST tag (represented as broken lines in the schematic) and the concomitant flow-through was then incubated with Ni-NTA resin wherein Gemin7 serves as the bait to pull down interacting protein components (lanes 4, 6 and 8).Asterisks denote protein degradation products.
(C) The identified minimal trimeric Gemin8:7:6 complexes from (B) represented in lanes 6 and 8 were analysed by SEC (HiLoad® 13/300 Superdex® 200 GL). *Left panel*: Schematic of formed protein complex; *Middle panel*: Overlay of SEC elution profiles of the reconstituted *Ce* Gemin8$^{145-199}$:7:6 (in blue; M_r = 38.8 kDa, V_E = 17.0 mL) and Gemin8$^{145-187}$:7:6 (in green; M_r = 37.2 kDa, V_E = 15.63/16.95 mL) complexes in comparison to *Ce* Gemin7:6 complex (in red; M_r = 31.5 kDa, V_E = 15.68 mL). The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference; *Right panel*: Coomassie stained SDS-PAGE gel (15 % Tris-Tricine) of the reconstituted trimeric protein complexes, *Ce* Gemin8$^{145-199}$:7:6 (lane 10) and Gemin8$^{145-187}$:7:6 (lane 11) with *Ce* Gemin7:6 (lane 9) as a reference. Note: The associated elution profiles are numbered similarly.

Prior to crystallization trials, the thus established Gemin8$^{145-199}$:7:6 and Gemin8$^{145-187}$:7:6 sub-complexes (Figure 4.7 B, lanes 6 and 8) were subjected to analytical SEC analysis (Figure 4.7 C) on the HiLoad® 13/300 Superdex® 200 GL sizing column to assess complex homogeneity. Unlike the Gemin8$^{144-199}$:7:6 sub-complex, the comparable Gemin8$^{145-187}$:7:6 sub-complex, only devoid of the C-terminal 12 amino acid (aa) unstructured stretch in Gemin8 (Figure 4.7 A), appears to be hetero-disperse (Figure 4.7 C, middle panel, elution profiles in blue and green respectively). The elution peak at 15.38 mL for the Gemin8$^{145-187}$:7:6 sub-complex occurs in close proximity to the elution profile of *Ce* Gemin7:6 sub-complex (V_E = 15.68 mL) (Figure 4.7 C, middle panel). Also, SEC runs with re-injections of either of the peak fractions of Gemin8$^{145-187}$:7:6 on an analytical HiLoad® 13/300 Superdex® 200 GL column (at identical molar concentrations) resulted in similar hetero-disperse elution profiles (data not shown). Moreover, with the predicted molecular weight of 37.2 kDa for this trimeric sub-complex, the earlier elution peak in comparison to the globular molecular weight calibration standards, could imply a higher order oligomeric species. Taken together, this data suggests that the Gemin8$^{145-187}$:7:6 sub-complex exists in a dynamic trimeric-hexameric species equilibrium *in vitro*. Nevertheless, these results define the C-terminal 145-187 residues of the nematode Gemin8 as the direct interaction interface for Gemin7.

4.3.2 The *H. sapiens* Gemin8:7:6 sub-complex

4.3.2.1 *Hs Gemin8:7 interaction interface*

Despite the effective biochemical characterisation of the nematode (*Ce*) Gemin8:7 interface, the inconsequential crystallogenesis attempts for the same (Section 4.3.1), encouraged extending similar biochemical and crystallographic studies for the human orthologues.

For the reconstitution of the trimeric Gemin8:7:6 trimeric complex, full-length (fl) Gemin8 protein was co-expressed with Gemin7:6 dimeric complex as described in Section 3.2.2. However, the co-expression of the fl-Gemin8 comprising of N-terminal His_6-/GST-/MBP-tags with Gemin7:6 sub-complex posed challenges: either insoluble or no recombinant expression was observed. At this juncture, grounded on the domain mapping data from the *C. elegans* Gemin8:7:6 sub-complex (Section 4.3.1), analogous *Hs* Gemin8 N-terminal truncation mutants (Figure 4.8 A) were established to generate crystallization prone sub-complexes. As anticipated, the extended N-terminal

truncation mutant, *Hs* Gemin8$^{103-242}$ (Figure 4.8 A, akin to the *Ce* Gemin8$^{57-199}$ variant), retained its Gemin7 binding domain resulting in the formation of a stable Gemin8:7:6 trimeric complex (Figure 4.8 B, elution profile 2, lane 2). Furthermore, as conceived in Figure 4.8 C (middle panel: elution profile 4, right panel: lane 4), the minimal *Hs* Gemin8$^{191-230}$ C-terminal variant (Figure 4.7 A, comparable to the minimal *Ce* Gemin8$^{144-187}$ variant) alone bears the necessary and sufficient Gemin7 binding domain as demonstrated by the resultant trimeric *Hs* Gemin8$^{191-230}$:7:6 sub-complex. These results augment the current work on the nematode orthologues described in Section 4.3.1.

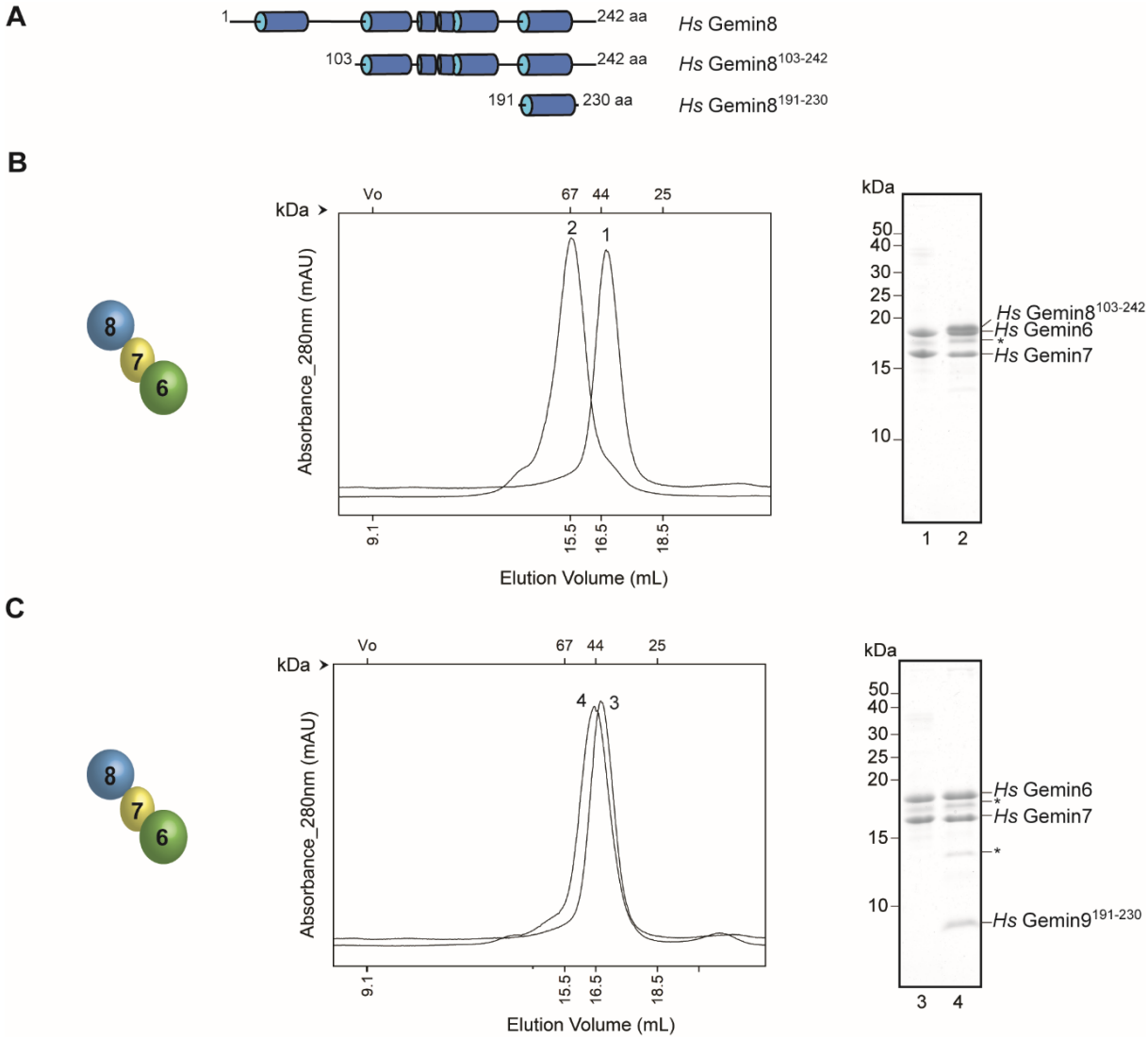


Figure 4.8 The *H. sapiens* (*Hs*) Gemin8:7 interface
(A) Schematic domain representation of *Hs* Gemin8 full-length and truncation variants based on PSIPRED algorithm (Jones, 1999) output. The α-helices and coils are rendered as cylinders and straight lines respectively, to scale.
(B) Complex integrity of the recombinantly purified *Hs* Gemin8$^{103-242}$:7:6 sub-complex was analysed by SEC (HiLoad® 13/300 Superdex® 200 GL) and SDS-PAGE. *Left panel*: Schematic of the formed trimeric protein complex; *Middle panel*: SEC elution profiles of the Hs Gemin8$^{103-242}$:7:6 complex (M_r = 50 kDa) [V_E = 15.52 mL; 2] in comparison to Gemin7:6 (M_r = 33.5 kDa) [V_E = 16.66 mL; 1]. *Right panel*: SDS-PAGE of the SEC elution peak fractions of Gemin7:6 (lane 1) and *Hs* Gemin8$^{103-242}$:7:6 (lane 2) sub-complexes visualized on a coomassie stained 15 % Tris-Tricine gel.
(C) Recombinant *Hs* Gemin8$^{191-230}$:7:6 sub-complex was subjected to SEC (HiLoad® 13/300 Superdex® 200 GL) analysis for ascertaining complex homogeneity. *Left panel*: Schematic of the formed trimeric protein complex; *Middle panel*: SEC elution profiles of the *Hs* Gemin8$^{191-230}$:7:6 complex (M_r = 38.1 kDa) [V_E = 16.57 mL; 3] in comparison to Gemin7:6 (M_r = 33.5 kDa) [V_E = 16.66 mL; 4]. *Right panel*: Coomassie stained SDS-PAGE gel (15 % Tris-Tricine) of Gemin7:6 (lane 3) and Gemin8$^{191-230}$:7:6 complexes (lane 4).
The elution volumes for the conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference. Bands marked with an asterisk denote degradation products.

4.3.2.2 *Delineation of the minimal Gemin8 binding domain on Hs Gemin7*

The studies thus far focussed on delineating the Gemin8:7 interface employing the nematode and human orthologues revealed conserved domains, with full-length Gemin7 associating with the minimal Gemin8 C-terminal truncations. However, sequence alignments of *Ce* Gemin7 and *Hs* Gemin7 demonstrate that while their Gemin6 interaction interfaces were largely conserved, most of the predicted N-terminal unstructured region of the *Hs* Gemin7 was lacking in the former (Figure 4.4 B). This necessitated defining the Gemin7:8 interface within the human SMN complex using the previously established *Hs* Gemin7:6 sub-complex truncation variants (Section 4.2.1, Figure 4.2 A) for enhancing crystallization propensity. The minimal *Hs* Gemin8$^{190\text{-}230}$ (Figure 4.8 A) and *Hs* Gemin6$^{1\text{-}92}$ (Figure 4.2 A) variants, both necessary and sufficient for Gemin7 association (Sections 4.2.1 and 4.3.2) were co-purified with Gemin7 truncation variants (Figure 4.9 A) by Ni-NTA affinity tag purification (Section 3.2.2.1) with subsequent SEC analysis.

The Gemin7$^{31\text{-}131}$:6$^{1\text{-}92}$ sub-complex (Figure 4.9 B, elution profile 1, lane 1) associated with the minimal *Hs* Gemin8$^{190\text{-}230}$ variant, resulting in a homogeneous trimeric complex (Figure 4.9 B, middle panel: elution profile 2, right panel: lane 2). More interestingly, the Gemin7$^{46\text{-}131}$:6$^{1\text{-}92}$ sub-complex (Figure 4.9 B, middle panel: elution profile 3, right panel: lane 3) still retained its interaction with *Hs* Gemin8$^{190\text{-}230}$ (Figure 4.9 B, middle panel: elution profile 4, right panel: lane 4).These results imply that the largely unstructured N-terminus (residues 1-45) of *Hs* Gemin7 is non-essential for Gemin8 association, in context of the SMN complex architecture.

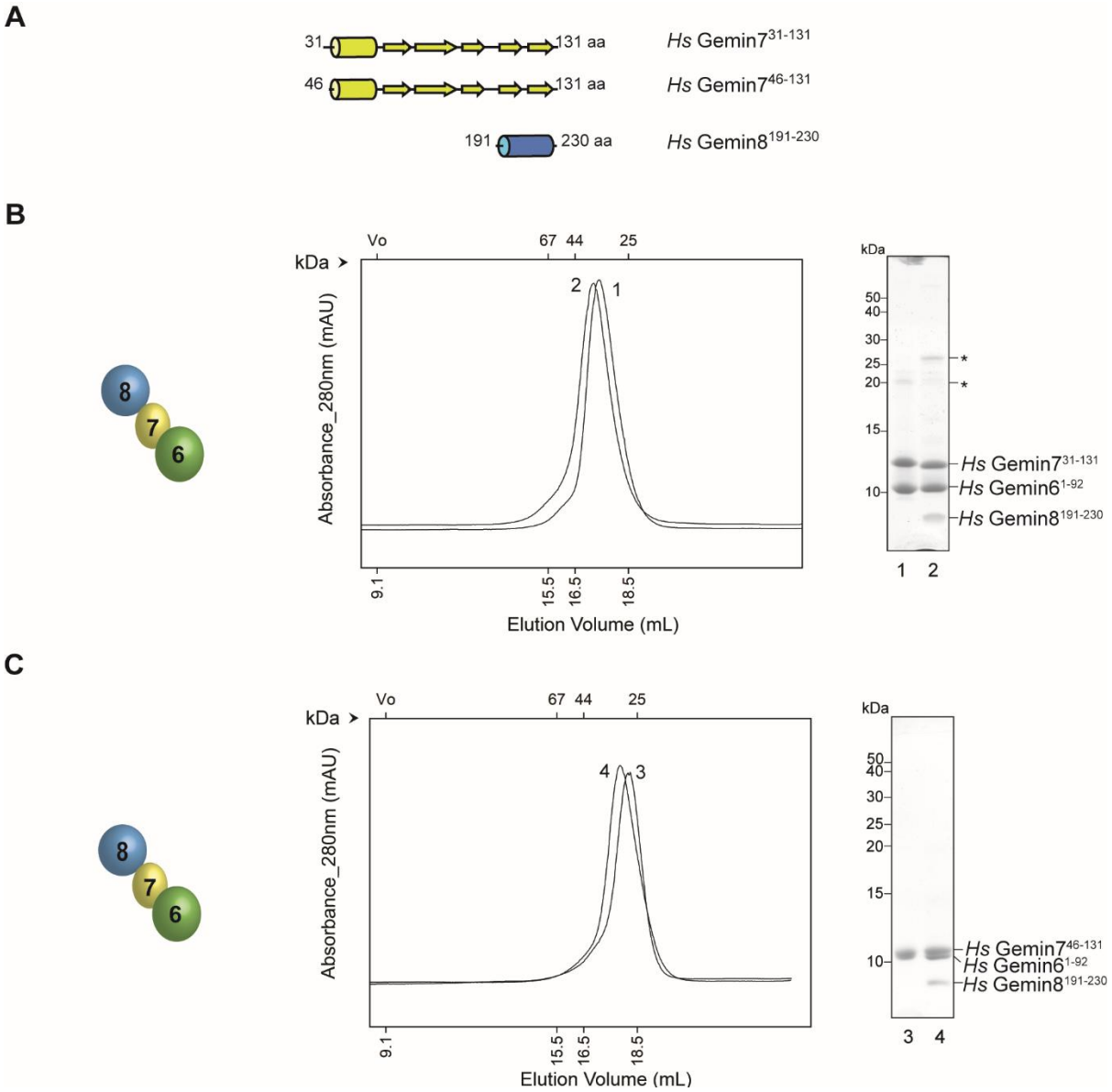


Figure 4.9 Delineation of Gemin7's Gemin8 binding domain
(A) Schematic domain representation of *Hs* Gemin7 and Gemin8 truncation variants based on PSIPRED algorithm (Jones, 1999) output. The α-helices, β-sheets and coils are rendered as cylinders, arrows and straight lines respectively, to scale.
(B) Complex integrity of the recombinantly purified *Hs* Gemin8$^{191-230}$:7$^{31-131}$:6$^{1-92}$ sub-complex was analysed by SEC (HiLoad® 13/300 Superdex® 200 GL) and SDS-PAGE. *Left panel*: Schematic of the formed trimeric protein complex; *Middle panel*: SEC elution profiles of the *Hs* Gemin8$^{191-230}$:7$^{31-131}$:6$^{1-92}$ complex (Mr = 26.4 kDa) [V_E = 17.20 mL; 2] in comparison to Gemin7$^{31-131}$:6$^{1-92}$ (M_r = 22.1 kDa) [V_E = 17.40 mL; 1]. *Right panel*: SDS-PAGE of the SEC elution peak fractions of Gemin7$^{31-131}$:6$^{1-92}$ (lane 1) and *Hs* Gemin8$^{191-230}$:7$^{31-131}$:6$^{1-92}$ (lane 2) sub-complexes visualized on a coomassie stained 15 % Tris-Tricine gel. Bands marked with an asterisk indicate non-specific proteins.
(C) Recombinant *Hs* Gemin8$^{191-230}$:7$^{46-131}$:6$^{1-92}$ sub-complex was subjected to SEC (HiLoad® 13/300 Superdex® 200 GL) analysis. *Left panel*: Schematic of the formed trimeric protein complex; *Middle panel*: SEC elution profiles of the *Hs* Gemin8$^{191-230}$:7$^{46-131}$:6$^{1-92}$ complex (M_r = 24.7 kDa) [V_E = 17.26 mL; 3] in comparison to Gemin7$^{46-131}$:6$^{1-92}$ (M_r = 20.4 kDa) [V_E = 18.20 mL; 4]. *Right panel*: Coomassie stained SDS-PAGE gel (15 % Tris-Tricine) of Gemin7$^{46-131}$:6$^{1-92}$ (lane 3) and Gemin8$^{191-230}$:7$^{46-131}$:6$^{1-92}$ complexes (lane 4).
The elution volumes for the conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference.

4.3.3 Structural insights into the minimal *Hs* Gemin8:7:6 sub-complex

The minimal *Hs* Gemin8$^{191-230}$:7$^{46-131}$:6$^{1-92}$ sub-complex (Section 4.3.2.2, Figure 4.9 C) yielded well-diffracting crystals (Section 3.2.3) and the summary of the crystallographic data collection and refinement statistics is provided in Table 4.2 and Table 4.3 respectively. The structure of this trimeric complex was solved using the Gemin7:6 (PDB ID: 1y96, Ma et. al., 2005) dimer as the molecular replacement model and the Gemin8$^{190-230}$ fragment could be traced and built in the resulting electron density map. The resulting model could be refined to an R_{free}/R_{work} of 1.5 Å (Figure 4.10) and included residues 1-86, residues 47-131 and residues 191-227 for Gemin6, Gemin7 and Gemin8 respectively.

The C-terminus 191-227 residues of Gemin8 adopts a helix (α1)-turn-helix (α2) motif. The N-terminal α helix of *Hs* Gemin7 (residues Glu49-Met73) flanking the five stranded β-sheet interacts with both α1 and α2 helices of Gemin8. Within the Gemin8$^{190-230}$:Gemin7$^{46-131}$:6$^{1-92}$ trimeric sub-complex, the Gemins 7 and 6 retain their Sm-like interaction interface (Section 4.2.1) forming the continuous 10 stranded β-sheet by their outermost β strands (Figure 4.10).

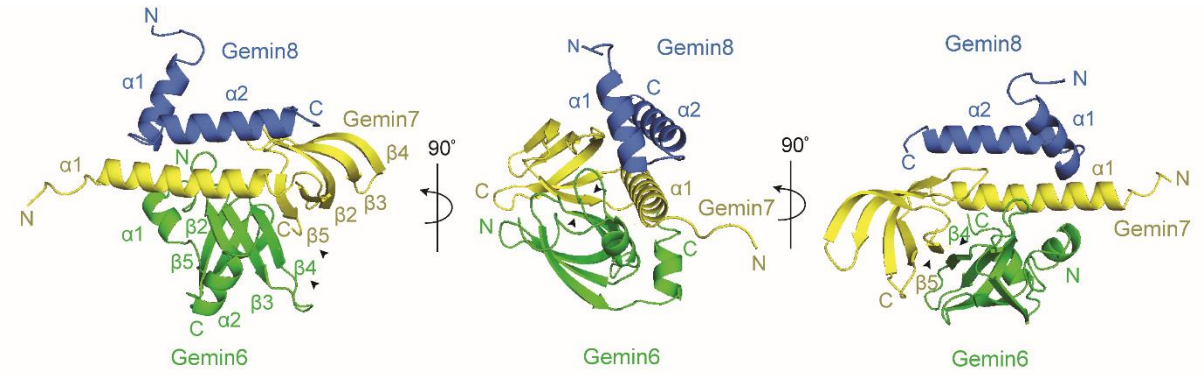


Figure 4.10 Molecular architecture of *Hs* Gemin8:7:6 sub-complex
Structural model of *Hs* Gemin8:7:6 sub-complex with Gemin8, Gemin7 and Gemin6 depicted in blue, yellow and green respectively. The C-terminus of Gemin8 consisting of a helix (α1)-turn-helix (α2) motif engages with the N-terminal α1 helix of Gemin7.The dimerization interface between Gemin7 (β5):Gemin6 (β4) strands (indicated with arrow heads) forming a continuous 10 sheet β-barrel remains intact. The respective N- and C- termini are labelled. Differential rotation views are indicated by the arrow and the degree of the rotation.

The interaction of Gemin8 and Gemin7 is mediated by hydrogen bonding and salt-bridge interactions (Figure 4.11 B and C). The side chain carbonyl oxygen of Glu216 (α2 chain) in Gemin8 and the side chain amino hydrogens of Arg63 (α1 chain) in Gemin7 form hydrogen bonds (Figure 4.11 B). In another instance, the main chain carbonyl oxygen of Arg203 (α1 chain) in Gemin8 engages with the side chain amino

hydrogen atoms of Gln56 (α1 chain) in Gemin7 establishing hydrogen bonds (Figure 4.11 B). In addition, the side chain of Glu216 (α2 chain) in Gemin8 forms salt bridge interactions with that of Gemin7's Arg63 (α1 chain) (Figure 4.11 C).

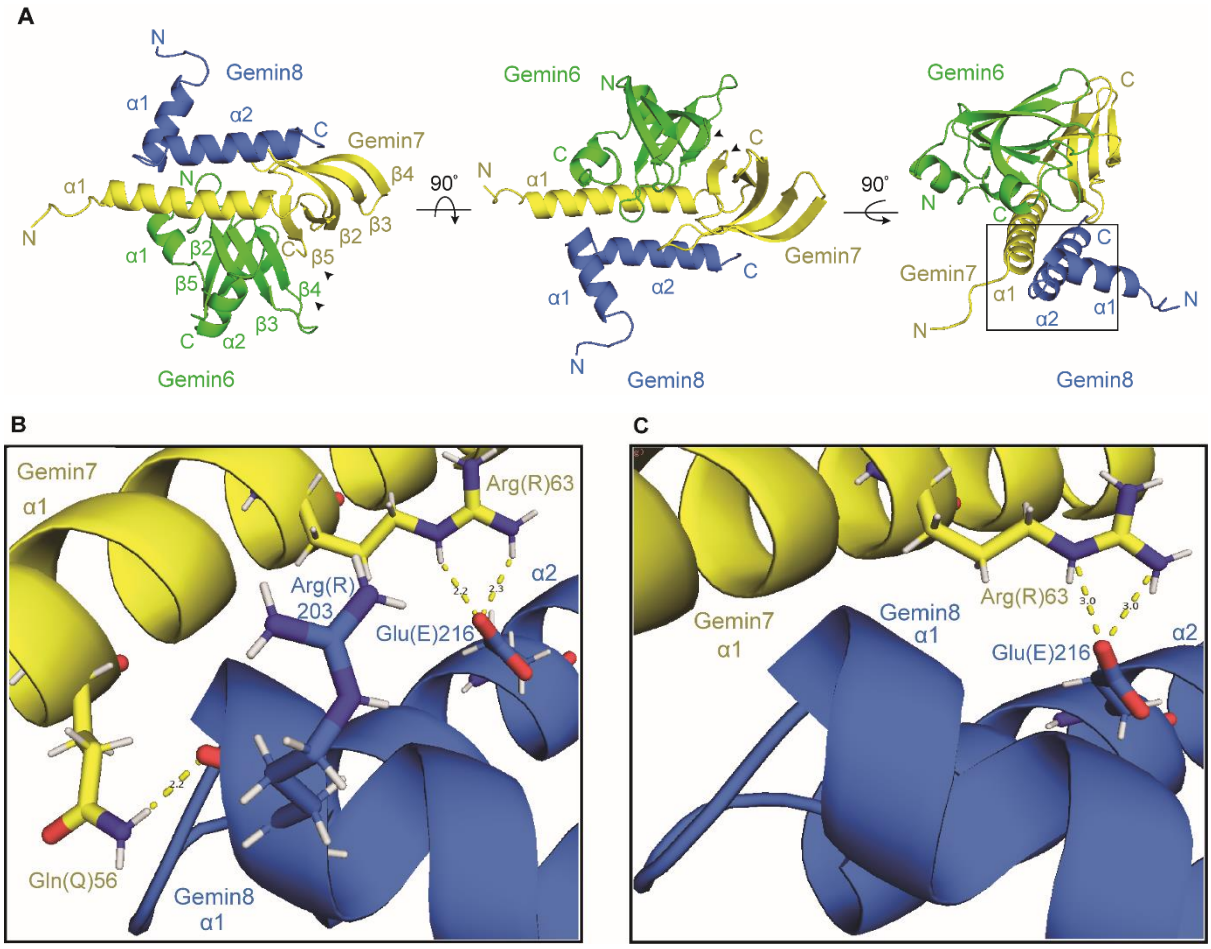


Figure 4.11 The Gemin8:7 interaction interface
(A) Differential rotation views of the Gemin8:7:6 sub-complex. Gemin8, Gemin7 and Gemin6 are represented in blue, yellow and green respectively. The respective N- and C-termini are labelled. The view presented in the right panel has been magnified (boxed region) in **(B)** and **(C)** for representational purposes.
(B) Within the Gemin8:7 interface, the three instances of hydrogen bonding are depicted here: Gemin8's Arg (R) 203 (in α1 chain) side-chain interacts with Gln (Q) 56 (on α1 chain) of Gemin7, while the Glu (E) 216 on α2 chain of Gemin8 engages with Gemin7's Arg (R) 63 (hydrogen atoms are depicted in grey).
(C) Also, within this interface, the Glu (E) 216 and Arg (R) 63 residues of Gemin8 (α2 chain) and Gemin7 (α1 chain) establish salt-bridge interactions via their oxygen (red) and nitrogen (blue) atoms.

4.3.4 The fission yeast *S. pombe* (*Sp*) Gemin8:7:6 sub-complex

Having demonstrated the conserved nature of Gemin8:7 interface in both nematodes and humans, a similar possibility was explored in the recently elucidated *S. pombe* SMN complex (unpublished data, RG Fischer). Employing a series of recombinant *Sp* Gemin8 N-terminal deletion variants (detailed in Section 4.5.3), the minimal *Sp* Gemin8 ΔN30ΔL variant (Figure 4.12 A) consisting only of the C-terminal residues

115-166 was sufficient to interact with the *Sp* Gemin7:6 sub-complex (Figure 4.12 B, middle panel: elution profile 2, right panel: lane 2) forming a homogenous trimeric sub-complex, akin to the nematode (Section 4.3.1) and human (Section 4.3.2) counterparts. Of note, similar to the *Ce* Gemin7:6 sub-complex (Section 4.2.2), the *Sp* Gemin7:6 exists as higher order oligomer (hetero-dimer, Figure 4.12 B, middle panel: elution profile 1, right panel: lane 1) in solution as well.

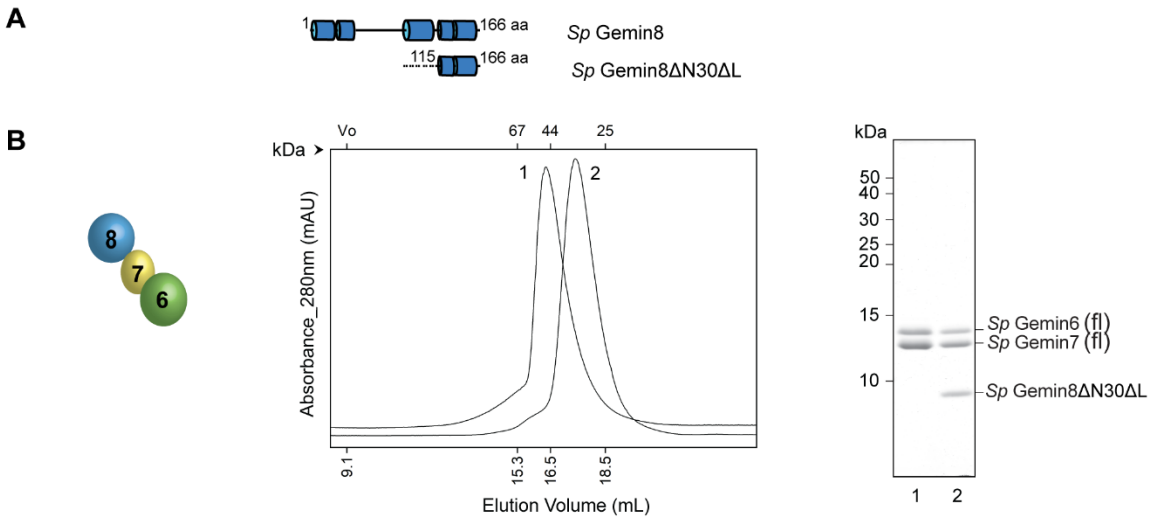


Figure 4.12 A minimal *S. pombe* Gemin8:7:6 trimeric complex
(A) Schematic of domain architecture of *S. pombe* (*Sp*) Gemin8 full-length (fl) and Gemin8ΔN30ΔL truncation mutant. The secondary structural elements (based on PSIPRED output), the α-helices and the unstructured regions (coils) are depicted as cylinders and straight lines respectively, to scale. The Gemin8ΔN30ΔL variant bears residual amino acids of a short linker which is shown here as dotted lines.
(B) Purification of *Sp* Gemin8ΔN30ΔL:7:6 protein sub-complex. Protein orthologues co-expressed in bacterial cells were initially purified by IMAC followed by SEC (HiLoad® 13/300 Superdex® 200 GL).*Left panel*: Schematic of formed protein complex; *Middle panel*: SEC elution profiles of the reconstituted *Sp* Gemin8ΔN30ΔL:7:6 (fl) complex (M_r = 27.9 kDa, V_E = 17.39 mL) in comparison to the *Sp* Gemin7:6 sub-complex (M_r = 20.9 kDa, V_E = 15.61 mL) indicated as (1) and (2) respectively. The elution position of the conventional globular proteins used as SEC calibration standards are indicated on the upper X-axis for reference; *Right panel*: the corresponding complexes (lanes 1 and 2) visualized on a coomassie stained 15 % Tris-Tricine gel.

The *Sp* Gemin8 ΔN30ΔL:7:6 trimeric complex yielded well-diffracting protein crystals at a 2.1 Å resolution, however due to time constrains at the beamline the entire data set could not be collected resulting in only 70 % completeness. Consequently, automated model building with PHENIX pipeline (Adams et al. 2010) was ineffectual (personal communication, Dr. C. Grimm, RG Fischer). However, while the online automated molecular replacement BALBES pipeline (Long et al. 2008), enabled an initial solution with R_{free} of 0.589, this could not be further refined given the incomplete data set. Although, these protein crystals were reproducible and a second attempt at crystallographic data collection being successful (2.7 Å resolution), technical glitches with solution refinement yet again hindered structural elucidation (see Section 7.8).

4.4 A putative role for the Gemin7:6 sub-complex during the heptameric Sm core assembly

The Gemins 6 and 7 of the SMN complex hetero-dimerize via their Sm-like interfaces (Section 4.2.1, Figure 4.3), reminiscent of the human Sm hetero-oligomers D1/D2, B/D3 (Kambach et al. 1999) and E/F/G (Plessel, Luhrmann, and Kastner 1997; Camasses et al. 1998) (Section 1.1.2, Figure 1.3). Also of note, GST pull-down experiments indicated that Gemin6 and Gemin7 could independently associate with individual Sm proteins with varying affinities given their identical structural features (Ma et al. 2005). Therefore, it was presumed such a likely engagement could occur during the distinct step-wise assembly of the heptameric Sm core (Sm D1/D2/E/F/G/B/D3) on the SMN complex (Section 1.2.3, Figure 1.11). To address this hypothesis, *in vitro* binding assays (Section 3.2.2.3) wherein GSH resin immobilized recombinant proteins (Gemin2 and Gemin8$^{190\text{-}230}$:7:6) served as the 'bait' to capture Sm hetero-oligomers (Sm D1/D2, Sm E/F/G and Sm B$^{1\text{-}174}$/D3), Sm core assembly intermediates (6S complex and Sm sub-core) and Sm-core (Figure 4.13).

As the placement of affinity tags on either Gemin7 or Gemin6 for the binding assays might jeopardize the experimental outcome, the recombinant His$_6$-GST-*Hs* Gemin8$^{190\text{-}230}$:Gemin7:6 sub-complex (Section 4.3.2) was chosen for the following *in vitro* binding assay. With Gemin2 of the SMN complex alone being sufficient to bind the 6S intermediate, Sm sub-core intermediate and Sm core, the previously established recombinant *Drosophila melanogaster* (*Dm*) Gemin2 served as an effective control for the experimental set-up (Section 1.3.2, Figure 1.11 A, E and F) (Chari et al. 2008; Zhang et al. 2011; Grimm et al. 2013).

In these binding assays (Figure 4.13), Gemin2 (Figure 4.13 A, lane 1) did not engage with the individual Sm hetero-oligomers: Sm D1/D2 (Figure 4.13 A, lane 8) Sm E/F/G (Figure 4.13 A, lane 9) and Sm B$^{1\text{-}174}$/D3 (Figure 4.13 A, lane 10). However, as anticipated, Gemin2 alone directly bound the Sm core assembly intermediates (1) the stalled intermediate, 6S complex (pIcln-SmD2/D1/E/F/G) (Figure 4.13 A, lane 11), but was incapable of releasing the assembly chaperone, pIcln, from the Sm hetero-oligomers (Grimm et al. 2013; Chari et al. 2008) and (2) the open pentameric Sm sub-core (Sm D1/D2/E/F/F) (Figure 4.13 A, lane 12) (Zhang et al. 2011). In case of the Sm core components binding (Figure 4.13 A, lane 7), Gemin2 retained its association to

five of the Sm proteins (Sm D1/D2/E/F/G) albeit with no detection of Sm B/D3 (Figure 4.13 A, lane 13). Sm B/D3 was previously shown to interact with SMN (Brahms et al. 2001; Friesen and Dreyfuss 2000; Friesen, Massenet, et al. 2001), with its association to the Sm-core being stabilized by U snRNA binding (Urlaub et al. 2001).

In contrast, the *Hs* Gemin8$^{190-230}$:7:6 sub-complex (Figure 4.13 B, lane 14) displayed no binding to the open Sm sub-core assembly intermediate (Figure 4.13 B, lane 25). This observation being corroborated by the lack of binding to the individual Sm hetero-oligomers as well (Figure 4.13 B, lanes 21, 22, 23). Also, as expected the trimeric *Hs* Gemin8$^{190-230}$:7:6 sub-complex does not engage with either the 6S complex (Figure 4.13 B, lane 24) or the Sm core components (Figure 4.13 B, lane 26).

Cumulatively under the tested experimental conditions, the isolated Gemin7:6 dimer despite possessing the Sm- like interfaces most likely does not engage with the Sm proteins during the step-wise Sm core assembly on the SMN complex.

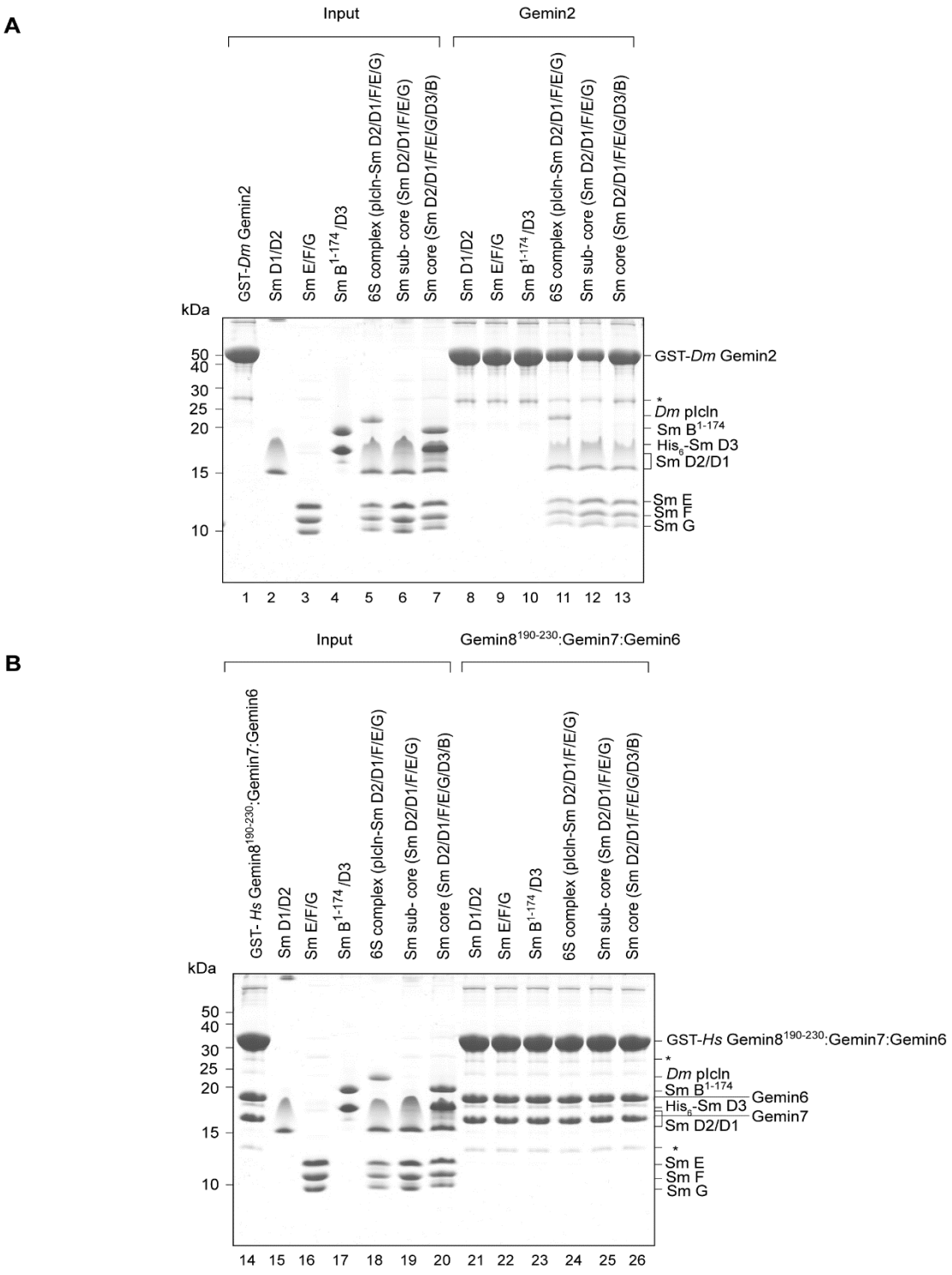


Figure 4.13 Gemin7:6 sub-complex does not engage with Sm protein hetero-oligomers
In vitro binding assays comparing Gemin2 and Gemin7:6 binding to Sm hetero-oligomers during Sm core assembly
(A) Recombinant GST-*Dm* Gemin2 (lane 1) immobilized on GSH resin was used as a control for the binding assays involving recombinant Sm protein hetero-oligomers (lanes 2-4), U snRNP assembly incompetent 6S complex (lane 5), Sm sub-core (lane 6) and Sm core (lane 7).
(B) Recombinant GST-*Hs* Gemin8$^{190-230}$:7:6 (lane 14) immobilized on GSH resin was tested for binding to recombinant Sm protein hetero-oligomers (lanes 15-17), U snRNP assembly incompetent 6S complex (lane 18), Sm sub-core (lane 19) and Sm core (lane 20).
Following elution from the GSH resin, the samples (5 % from the elution) were resolved on a 15 % Tris-Tricine gel system and visualized by coomassie staining. Asterisks indicate degradation products. *Dm*: *Drosophila melanogaster*; *Hs*: *Homo sapiens*.

4.5 The Gemin8: SMN interaction interface

Thus far, having established the conserved nature of Gemin8:7 interaction interface both biochemically and structurally, Gemin8's engagement with the oligomeric SMN in building up the heteromeric SMN complex (Figure 4.1) was examined. *In vitro* translation assays have demonstrated that the N-terminus of *Hs* Gemin8 (residues 1-124) contacts the *Hs* SMN C-terminus (residues 242-294) harbouring the oligomerisation inducing YG-box domain (Otter et al. 2007). Notably, given the clinical relevance of SMN's YG-box domain as a major hot-spot for Spinal Muscular Atrophy (SMA) mutations (Section 1.3.1.1), addressing Gemin8's engagement with SMN bearing these patient mutations merited investigation. To this end, a recombinant SMN YG-box biochemical handle bearing SMA patient mutations was established to investigate their association with Gemin8 in isolation.

4.5.1 The MBP-YG-box biochemical handle

4.5.1.1 *Establishing the Hs MBP-YG fusion model system*

SMN's C-terminal YG-box domain (Section 1.3.1, Figure 1.10) is both necessary and sufficient to influence SMN's ability to self-associate in forming complexes with unusually large hydrodynamic sizes (Lorson et al. 1998; Martin et al. 2012; Gupta et al. 2015). The biochemical and structural nature of the human (*Hs*) SMN oligomerization was explored by establishing the SMN YG-box domain in isolation as a C-terminal Maltose binding protein (MBP) fusion protein (Figure 4.14 A) (Martin et al. 2012). The MBP-YG$^{263\text{-}294}$ fusion protein structural model (PDB ID: 4GLI) adopts the glycine zipper architecture, with YG-box inter-subunit interactions characterised by the conserved tyrosine residues on one YG-helix packing against the i+3 glycine residues on the opposing strand (Figure 4.14 B) (Martin et al. 2012).

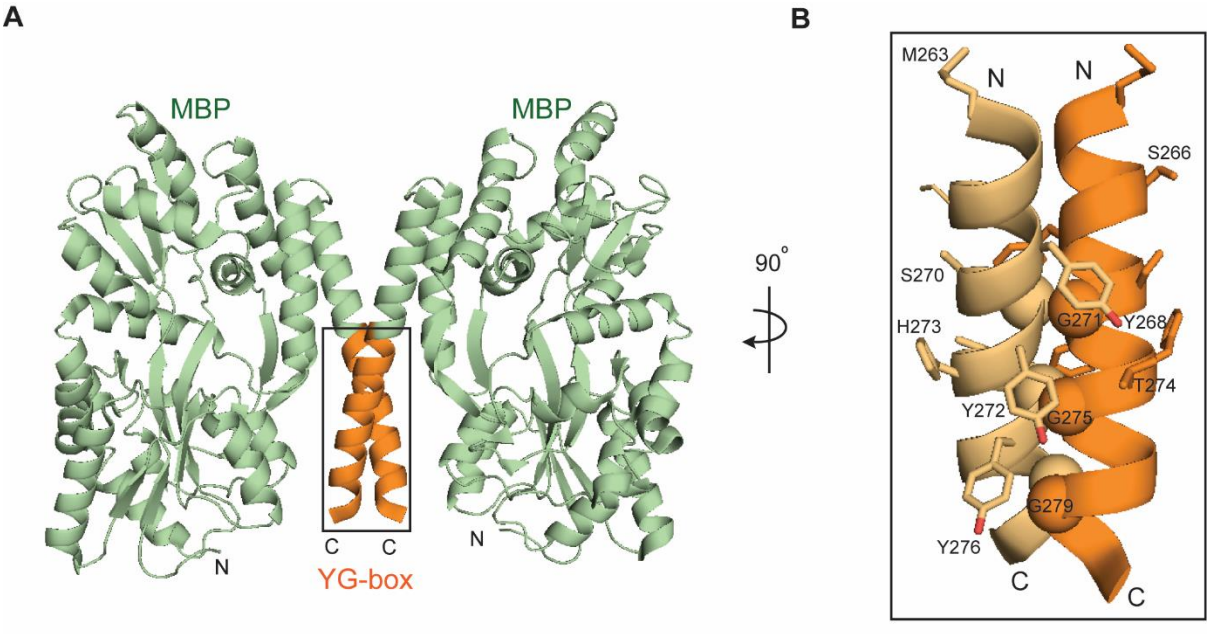


Figure 4.14 Structure of the *Hs* SMN YG-box domain
(A) Structure of the MBP-human SMN YG-box fusion protein (PDB ID: 4GLI). The tightly coupled fusion protein assembles as dimer with SMN's YG-box (in orange) (residues M263-R281) forming a continuous helix with the C-terminal of the MBP tag (in wheat). The dimeric assembly contacts observed are almost exclusively between the YG-box helices. The N- and C-termini of the MBP-fusion protein are indicated. *(PDB ID: 4GLI)
(B) The YG-box dimer interface. The boxed region in **(A)**, the YG-box helical dimer is depicted in this panel. The conserved tyrosine residues (sticks) of the YxxxYxxxY motif (Y268, Y272, Y276) on one YG helix (wheat) interface against the conserved glycine residue (spheres) of the GxxxGxxxG motif (G271, G275, G279) on the other interacting YG helix (orange). The residues mapped to SMA patient mutations studied in this dissertation (M263, S266, S270, Y272, H273 and T274) indicated here were substituted in the longer MBP-*Hs* YG$^{252-284}$ fusion constructs and the consequential influence on SMN's oligomerisation behaviour was determined by SEC (Section 4.5.1.3).

Martin et al. (2012) reported that while the tightly coupled MBP-YG$^{263-294}$ fusion protein formed dimers were amenable for crystallization, the longer linker variant MBP-YG$^{252-294}$ existed at a higher oligomeric status equivalent to those observed for native SMN in solution. Centred on these findings, comparable MBP tagged SMN's YG-box domains (Figure 4.15 A) were established for our experimental purposes, however, with the following variations: (1) within the reported MBP-*Hs* YG$^{263-294}$ dimeric structural model, the terminal 13 residues (282-294 aa) were not involved in forming the glycine zipper motif, therefore the terminal residues (285-294 aa) were excluded in the current recombinant protein constructs and, (2) to control the oligomeric status of the thus generated fusion proteins, the pETM41 expression vector bearing the N-terminal MBP tag was modified (pETM41*, see Appendix 7.3) by shortening the linker distance between the MBP protein and multi-cloning site (MCS) (Figure 4.15 B). The resultant recombinant MBP fusion YG-box proteins (henceforth referred to as 'MBP*-YG') (Figure 4.15 D, middle panel: lanes 2 and 3) were subjected to SEC analyses for establishing the oligomeric status (Figure 4.15 D, right panel).

Both the minimal MBP*-*Hs* YG$^{263-284}$ (analogous to the crystallized human MBP-YG fusion, PDB ID: 4GLI) and the longer MBP*-*Hs* YG$^{252-284}$ variants formed stable dimers in solution (at similar molar concentrations of 150-200 µM) (Figure 4.15 D). This was in contrast to the previously reported study, wherein the longer linker variant MBP-YG$^{252-294}$ formed stable tetramers in solution (Martin et al., 2012). To circumvent this issue, the SMN YG-box variants were established in the original pETM41 expression system, affording a longer MBP-SMN from the vector (Figure 4.15 C), to enable higher order oligomerisation states as reasoned earlier (Martin et al., 2012). Interestingly, both the longer MBP-*Hs* YG$^{252-284}$ and shorter MBP-*Hs* YG$^{263-284}$ fusion proteins (Figure 4.15 E, middle panel: lanes 5 and 6) largely formed similar higher order oligomers (octamers) (Figure 4.15 E, left panel: elution profiles in red and green respectively) albeit a small population of dimeric MBP-YG fusions (based on their retention volumes on a HiLoad® 13/300 Superdex® 200 GL) were evident as well. These results substantiate the notion of longer linkers favouring higher oligomeric states mimicking those observed for native SMN *in vitro* (Martin et al. 2012; Gupta et al. 2015). Both the empty pETM41 and modified pETM41* vectors (Section 3.1.5, Appendix 7.3) behaved as monomeric species in solution (at molar concentrations of 200 µM) attributing the oligomerisation behaviour of the fusion proteins solely to SMN's YG-box domain (Figure 4.15 D and E, left panel: elution profiles 1 and 4 respectively).

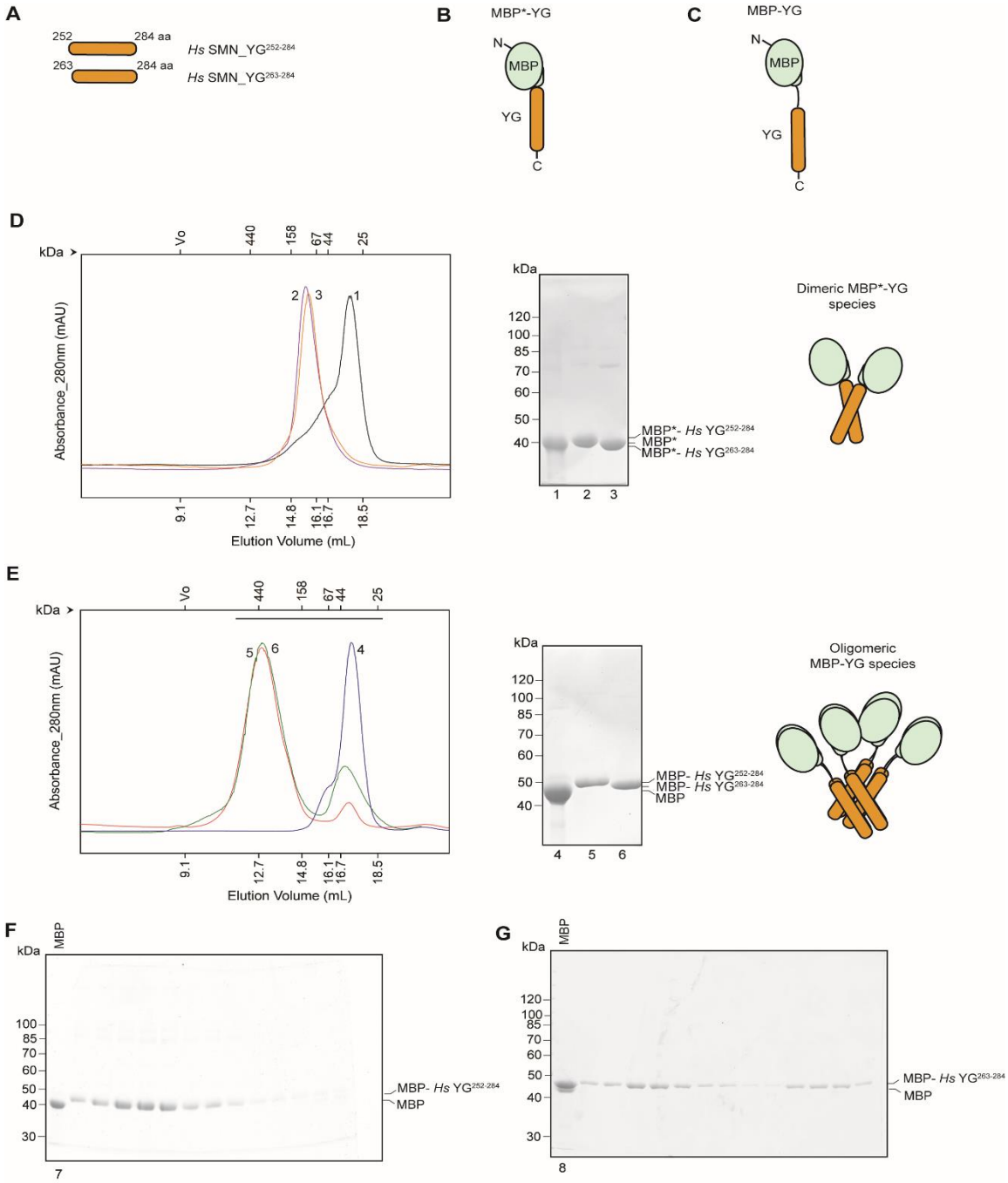


Figure 4.15 The *Hs* MBP-YG box biochemical handle

(A) The *Hs* SMN YG box region was established in isolation as soluble MBP (Maltose binding protein)-tagged fusion proteins as an experimental handle for the experiments (Martin et al., 2012). To control the extent of SMN YG box oligomers formed, the MBP-SMN linker lengths were varied by employing variable lengths of the *Hs* SMN YG box region as indicated in the schematic.

(B) and **(C)** illustrate the tightly coupled MBP*-YG fusion (in pETM41* vector) and longer linker bearing MBP-YG fusion proteins respectively. Note the longer linker region afforded by the original pETM41 vector in (C) enables higher oligomerisation states. The respective N- and C- termini are indicated.

(D) Oligomerisation status of MBP*-YG fusion proteins determined by SEC on an analytical HiLoad® 13/300 Superdex® 200 GL column. *Left panel*: SEC elution profiles of MBP*-*Hs* $YG^{252-284}$ (M_r = 46.4 kDa; V_E = 15.54 mL) and MBP*-*Hs* $YG^{263-284}$ (M_r = 44.9 kDa; V_E = 15.71 mL) in comparison to MBP* alone (M_r = 41.8 kDa; V_E = 17.81 mL) in purple (2), orange (3) and black (1) respectively. *Middle panel*: Coomassie stained SDS-PAGE gel (8 % Tris-Glycine) of MBP* (lane 1), MBP*-*Hs* $YG^{252-284}$ (lane 2) and MBP*-*Hs* $YG^{263-284}$ (lane 3) respectively. *Right panel*: Schematic of the formed dimeric MBP*-YG fusion proteins based on the SEC retention volumes.

(E) Oligomerisation status of MBP-YG fusion proteins determined by SEC on an analytical HiLoad® 13/300 Superdex® 200 GL column at 200 µM. *Left panel*: SEC elution profiles of MBP-*Hs* $YG^{252-284}$ (M_r = 49.1 kDa) and MBP-*Hs* $YG^{263-284}$ (M_r = 47.6 kDa) in comparison to MBP alone (M_r = 44.5 kDa; V_E = 17.22 mL) in red (5), green (6) and

blue (4) respectively. *Middle panel*: Coomassie stained SDS-PAGE gel (8 % Tris-Glycine) of MBP (lane 4), MBP-*Hs* $YG^{252-284}$ (lane 5) and MBP-*Hs* $YG^{263-284}$ (lane 6) respectively. The SDS-PAGE lanes and the associated elution profiles are numbered identically for ease of reference. *Right panel*: Schematic of the formed octameric MBP-YG fusion proteins based on the SEC retention volumes. The black line above the elution profile indicates the samples from SEC applied to SDS-PAGE shown in **(F)** and **(G)**. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference.
(F) and **(G)** SDS-PAGE (8 % Tris-Glycine) of elution profile of MBP-*Hs* $YG^{252-284}$ (seen in **(E)**, elution profile 5, in red) and MBP-*Hs* $YG^{263-284}$ (see in **(E)**, elution profile 6, in green) respectively. MBP fusion tag alone (lanes 7 and 8) are indicated. The black line in the chromatogram indicates the range of SEC elution fractions analysed with SDS-PAGE.

4.5.1.2 *Structure of the Ce SMN YG-box dimer*

SMN's C-terminal YG-box domain harbours two significantly conserved overlapping motifs across divergent organisms: YxxGYxxGYxxG and SxxxSxxxSxxxT, that contribute to the glycine zipper architecture (Figure 4.16 A). However, the subtle residue variations within this domain, for instance, the third YxxG repeat in *C. elegans* (*Ce*) SMN (Wormbase ID: C41G7.1b) is YxxA whilst an LxxG variation occurs in the fission yeast (*Sp*) and fly (*Dm*) SMN proteins (Figure 4.16 A). This prompted examining the consequential structural differences (if any) relative to the previously published human SMN and fission yeast SMN YG-box dimer (PDB IDs: 4GLI and 4RG5 respectively) (Martin et al. 2012; Gupta et al. 2015).

To this end, a MBP *Ce* SMN's YG-box fusion (MBP*-*Ce* $YG^{182-207}$) (Figure 4.16 B) was generated for crystallisation pursuits as described for the human SMN YG-dimer (Section 4.5.1.1) (Martin et al. 2012). This dimeric *Ce* MBP*-YG-box fusion protein (Figure 4.16 C) yielded diffraction quality crystals (Section 3.2.3) and the associated crystallographic data and refinement statistics are presented in Table 4.2 and Table 4.3 respectively. Structural elucidation of the nematode YG-box fusion protein was enabled by molecular replacement using the human MBP-YG fusion structural model (PDB code: 4GLI) (Martin et al. 2012) with subsequent refinement to a 2.2 Å resolution (Section 3.2.3.2).

In the MBP*-*Ce* $YG^{182-207}$ structure (Figure 4.16 D), as anticipated, SMN's YG-box domain moulds into a right-handed glycine zipper helical dimer consisting of the most highly conserved residues (Met182-Lys203) with the no observed electron density for the disordered C-terminal four residues (Asn204-Asn207). The first two conserved tyrosine residues' (Tyr187 and Tyr191) side chains pack against the preceding i+3 conserved glycine (Gly190 and Gly195) backbone in the opposing helix (Figure 4.16 E), akin to the human and fission yeast structural models (PDB IDs: 4GLI and 4RG5, respectively).

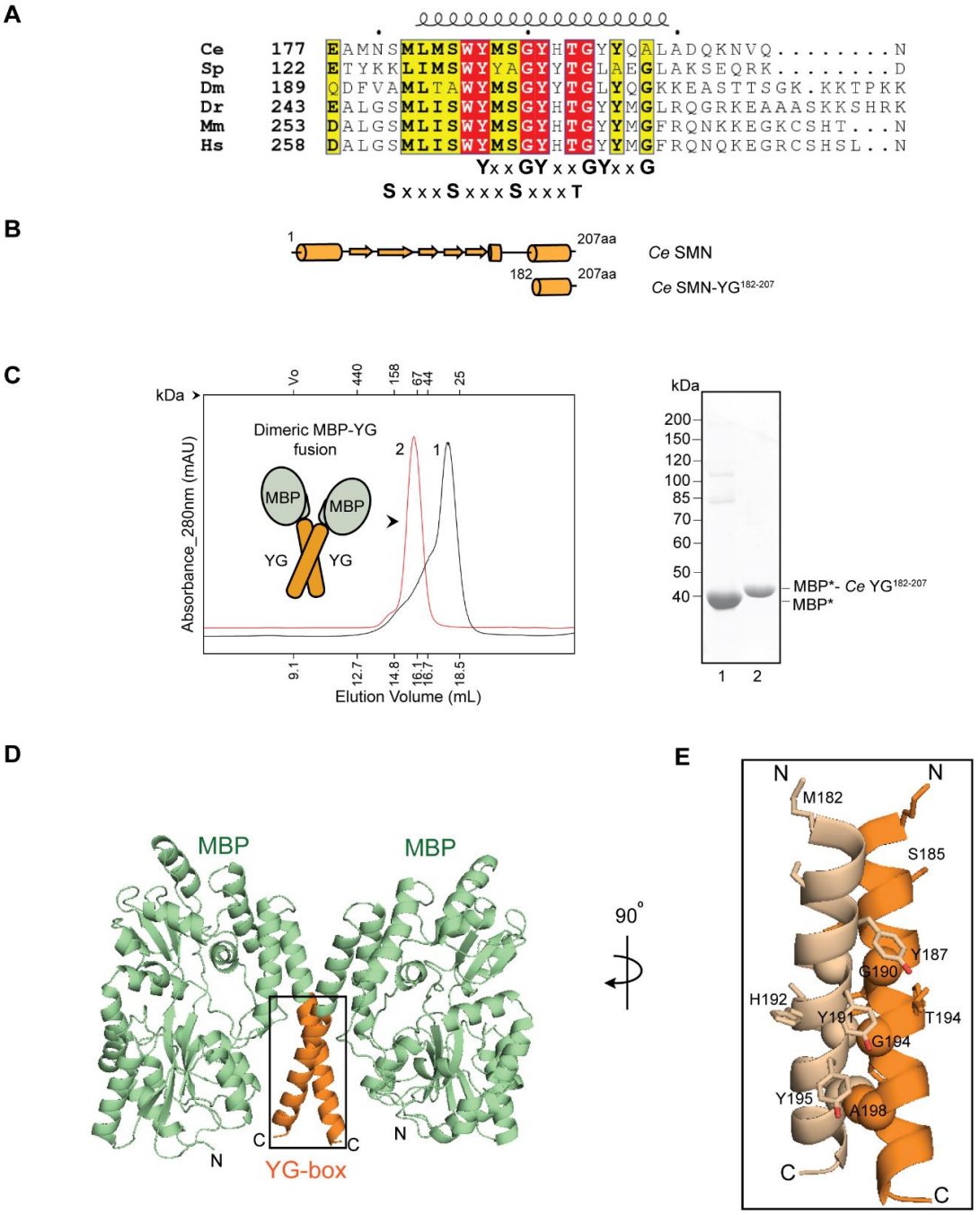


Figure 4.16 Structural conservation of SMN YG-box module in *C. elegans* (*Ce*)
(A) Alignment of SMN's YG-box sequences from diverse organisms. Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. The conserved S, G and Y motifs are indicated. The secondary structure element (α-helix) of the *H. sapiens* YG-box (PDB: 4GLI) is displayed above the sequences as squiggles. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).
Ce: *C. elegans*, Sp: *S pombe*, Dm: *D. melanogaster*, Dr: *D. rerio*, Mm: *M. musculus*, Hs: *H. sapiens* (also, see Section 7.5)
(B) Schematic of domain architecture of *C. elegans* (*Ce*) SMN full-length (fl) and minimal C-terminal YG-box module used for crystallization. The secondary structural elements (based on PSIPRED output), the α-helices and the unstructured regions (coils) are depicted as cylinders and straight lines respectively, to scale.
(C) Oligomerisation status of MBP*-YG fusion protein determined by SEC on an analytical HiLoad® 13/300 Superdex® 200 GL column. *Left panel*: SEC elution profiles of MBP*-*Ce* YG$^{182-207}$ (M_r= 44.9 kDa; V_E = 15.91 mL) in comparison to MBP* alone (M_r = 41.8 kDa; V_E = 17.81 mL) in orange (2) and black (1) respectively. Inset illustrates the tightly couple N-terminal MBP tagged YG-box that only allows formation of a dimeric species *in vitro*. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference. *Right panel*: Coomassie stained SDS-PAGE gel (8 % Tris-Glycine) of MBP* (lane 1), MBP*-*Ce* YG$^{182-207}$ (lane 2).

(D) Structure of the MBP-YG-box fusion protein. The tightly coupled fusion protein assembles as dimer with SMN's YG-box (in orange) (residues M183-K203) forming a continuous right-handed helix with the C-terminal of the MBP tag (in wheat). The N- and C-termini of the MBP-fusion protein are indicated.
(E) Dimerization interface of the nematode SMN YG-box. The boxed region in **(D)**, the YG-box helical dimer is depicted in this panel. The conserved tyrosine residues (sticks) of the YxxxYxxxY motif (Y187, Y191, Y195) on one YG helix (wheat) interface against two of the highly conserved glycines (G190, G194), with the third glycine of the GxxxGxxxG motif, being replaced by an alanine residue (A198) here on the other interacting YG helix (orange). The residues mapped to the SMA patient mutations studied in this dissertation - M181, S185, Y191, H192 and T273 - indicated here were substituted in the longer MBP-*Ce* $YG^{162-207}$ fusion construct and the consequential influence on SMN's oligomerisation behaviour was determined by SEC (Section 4.5.1.3).

Notably, in the nematode SMN the third tyrosine residue of the $(YxxG)_3$ motif, Tyr195, interfaces minimally with the alanine residue (Ala198) alternate on the opposing YG helix, a variation to the third YxxG repeat of the classical glycine zipper motif (Figure 4.16 E). The conserved serine (here, the structural model only includes the last two serines, Ser185 and Ser189) and threonine residues (Thr194) of the SxxxSxxxSxxxT motif, engage in intra-helical hydrogen bonds and hydrogen bonding to the carbonyl oxygens of the i-3 residues. An RMSD of 0.539 (RMSD values of <1.0 imply largely identical conformation) computed after optical superposition of the human and nematode YG-box structures further emphasises an identical molecular architecture (Kufareva and Abagyan 2012).

4.5.1.3 *Spinal Muscular Atrophy (SMA) patient mutations influence SMN's YG-box derived oligomerisation*

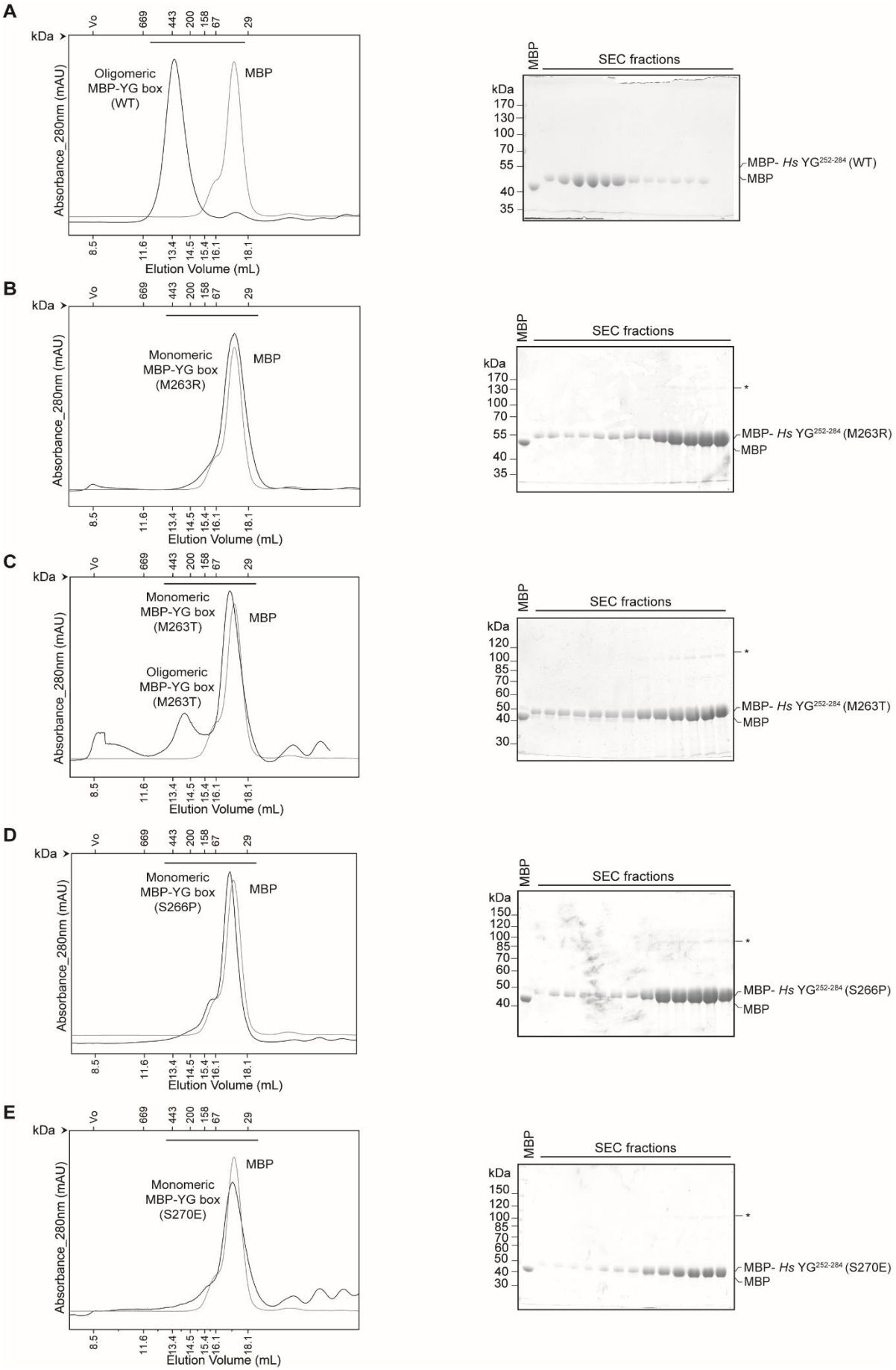
Defective SMN oligomerisation arising due to Spinal Muscular Atrophy (SMA) patient mutations (Section 1.3.1.1) in the self-interacting YG-box module was earlier investigated by SEC-MALS employing the highly oligomeric MBP-human SMN YG-box fusion protein (Martin et al. 2012). To recapitulate these findings as well as intending to extend them to the nematode SMN (Section 4.5.1.2), the oligomeric status of the established MBP-YG model system (Figure 4.15 E and F) bearing a common set of SMA YG-box mutations (albeit with certain exceptions for the human and nematode biochemical handles) was determined using SEC and the results are presented in Figure 4.17 (for *Hs* SMN) and Figure 4.18 (for *Ce* SMN) and summarized in Table 4.1.

The recombinantly purified *Hs* MBP-$YG^{252-284}$ wild-type (WT) (Figure 4.17 A) and *Ce* $YG^{169-207}$ WT (Figure 4.18 A) fusion proteins exclusively formed octamers in solution as observed on the sizing column (HiLoad® 13/300 Superdex® 200 GL).

The patient mutation S181G located immediately N-terminal to M182 (the first YG-box residue in the nematode structural model, Figure 4.16 E) does not disrupt the oligomerisation status of SMN (Figure 4.18 B), reflecting an identical behaviour of the

equivalent SMA mutation in the human MBP model studied by Martin et al., in 2012. Also, the recently published fission yeast MBP-YG structural model (PDB ID: 4RG5) includes several residues at the N-terminus of the YG-box in comparison to the human and nematode models that extend into a right handed helix as well. Given the highly conserved nature of the YG-box module, Ser181 residue (in nematodes) would occur on a solvent exposed surface of the YG-helix and, thus the S181G substitution would not affect SMN self-association.

In contrast, the patient mutations of the first YG-box residue in both the human (Figure 4.14 B) and our nematode (Figure 4.16 E) structural models: M263 (Figure 4.17 B and C) and M182 (Figure 4.18 C and D) respectively, impairs SMN self-association, albeit variably (see Table 4.1). Likewise, the S226P and S185P substitutions of the first conserved serine residue of the SxxxSxxxSxxxT motif within the YG-box module (Figure 4.16 A), resulted in a monomeric MBP-YG fusions in solution (Figure 4.17 D and Figure 4.18 E). It must be noted that an alanine substitution of the second *Hs* Ser270 residue in this motif was earlier reported to form large oligomers much like the WT (Martin et al. 2012), consistent with the presence of an alanine variant in fission yeast at this position within the conserved YG-box surface (Figure 4.16 A). However, substitution of the Ser270 with a bulker amino acid (S270E) prevented SMN oligomerisation (Figure 4.17 E). These substitutions are indeed consistent with the molecular architecture of the YG-box dimer since these conserved serine residues although not participating directly in the YG-dimer interface form a network of hydrogen bonds within each YG-helix contributing to helix stabilization (Section 4.5.1.2; Figure 4.14 B and Figure 4.16 E).


A
kDa
Vo
669
443
200
158
67
29
Oligomeric
MBP-YG box
(WT)
MBP
Absorbance_280nm (mAU)
8.5
11.6
13.4
14.5
15.4
16.1
18.1
Elution Volume (mL)
MBP
SEC fractions
kDa
170
130
100
70
55
40
35
MBP- Hs YG252-284 (WT)
MBP
B
Monomeric
MBP-YG box
(M263R)
MBP
MBP- Hs YG252-284 (M263R)
C
Monomeric
MBP-YG box
(M263T)
Oligomeric
MBP-YG box
(M263T)
MBP- Hs YG252-284 (M263T)
D
Monomeric
MBP-YG box
(S266P)
MBP- Hs YG252-284 (S266P)
E
Monomeric
MBP-YG box
(S270E)
MBP- Hs YG252-284 (S270E)

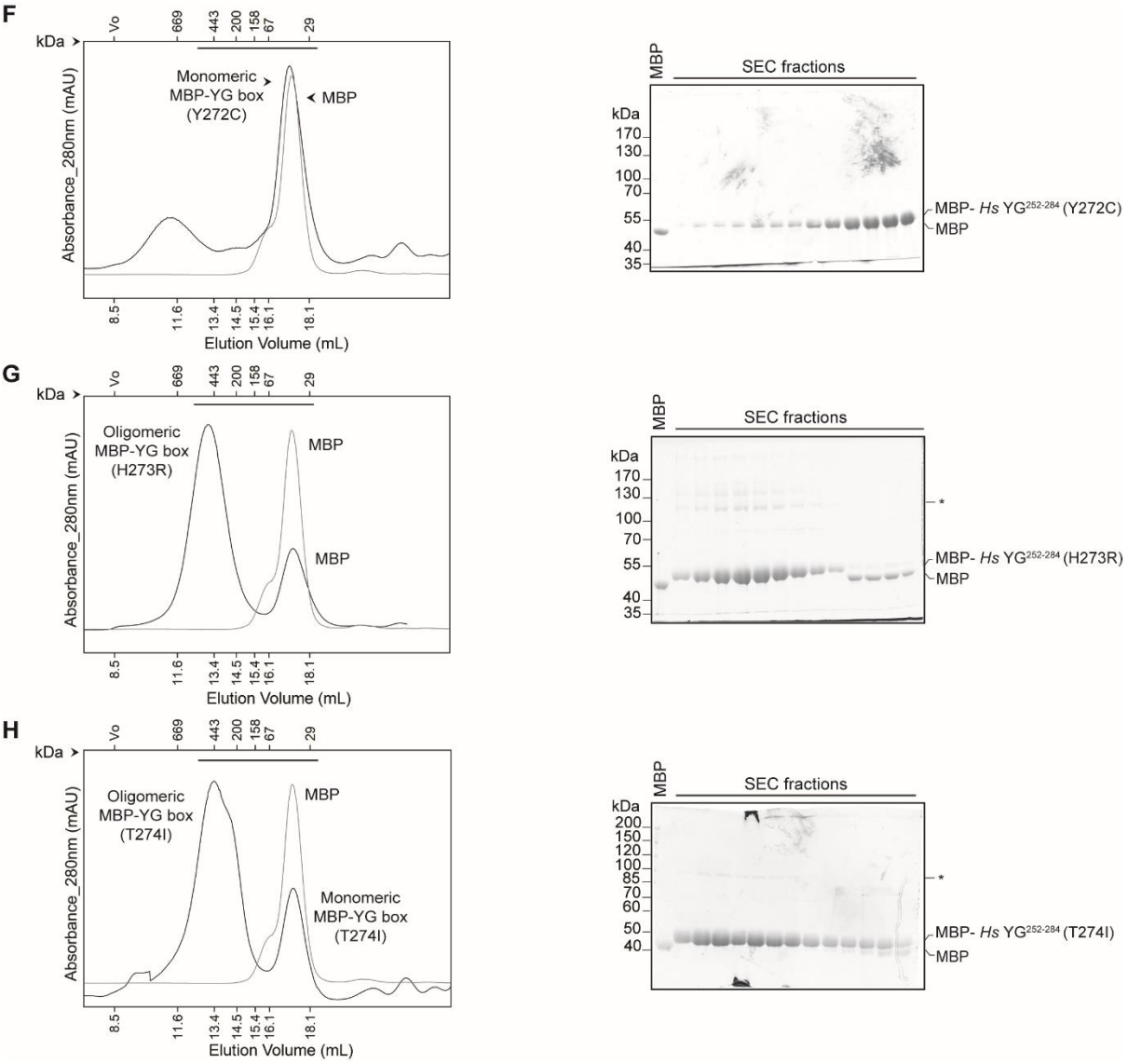
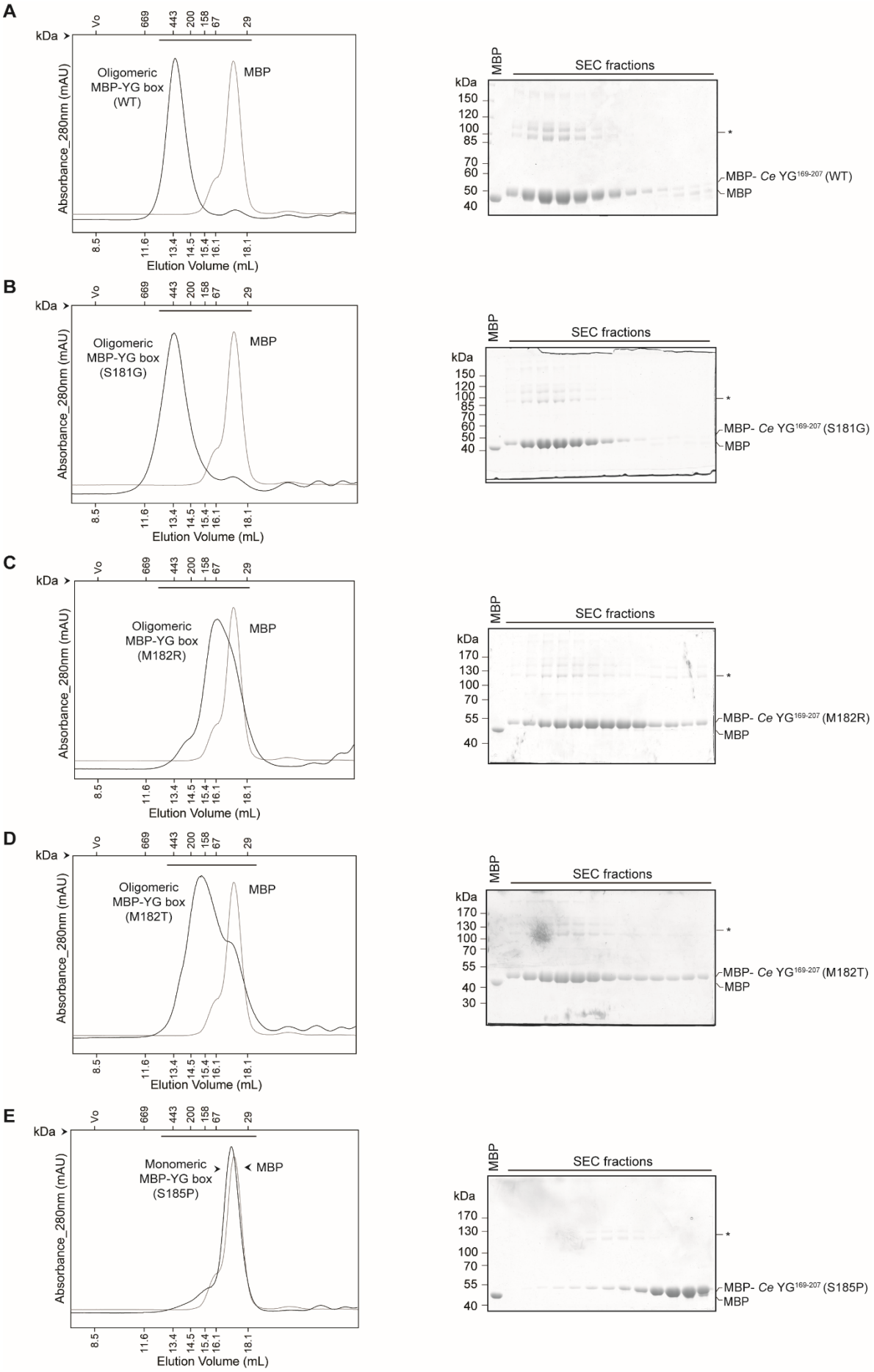


Figure 4.17 Effects of SMA patient mutations on MBP-*Hs* YG$^{252-284}$ oligomerisation
(A-H) SEC and SDS-PAGE analysis of SMA patient mutations bearing MBP-YG fusions.
(A) WT, **(B)** M263R**, (C)** M263T, **(D)** S266P, **(E)** S270E**, **(F)** Y272C, **(G)** H273R, **(H)** T274I. **Non-SMA patient mutation.
Left panel: SEC analysis of MBP-*Hs* YG$^{252-284}$ WT construct (M_r = 49.1 kDa; V_E = 13.24 mL) and mutants shown in black in comparison to the MBP alone elution profile (M_r = 44.5 kDa; V_E = 17.22 mL) in grey. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference. The black line above chromatogram depicts the range of the SEC fractions analysed by SDS-PAGE. *Right panel*: SDS-PAGE (8 % Tris-Glycine) of SEC elution profiles of the MBP fusions including the MBP protein as a reference for gauging any degradations. PageRuler™ Prestained Protein ladder (170-10 kDa) (Section 3.1.8.4.2) was used as a reference for SDS-PAGE gels in (A), (B), (F) and (G). PageRuler™ Unstained Protein ladder (200-10 kDa) (Section 3.1.8.4.2) was used for SDS-PAGE gels in (C), (D), (E) and (H). *Non-specific protein bands.

The highly conserved second tyrosine residue of the YxxGYxxGYxxG motif (Figure 4.16 A), imperative for the classical YG-dimerization interface (Figure 4.14 B and Figure 4.16 E), when substituted with SMA patient mutations (Y272C and Y191C in human and nematode respectively) led to impaired oligomerisation status as evidenced by their higher retention volumes on the HiLoad® 13/300 Superdex® 200 GL column (Figure 4.17 F and Figure 4.18 F). On the contrary, the subsequent solvent-

exposed histidine residue that does not participate in the dimerization interface when substituted with SMA patient mutations (*Hs* H273R/ *Ce* H192R) (Figure 4.17 G and Figure 4.18 G) resulted in oligomerisation competency similar to that of the wild-type YG box (Figure 4.17 A and Figure 4.18 A).

The T274I (human; Figure 4.17 H)/ T193I (nematode; Figure 4.18 H) SMA substitution of the threonine residue within the $(YxxxG)_3$ motif displayed only a minor oligomerisation defect forming oligomers spanning octamers, hexamers (not distinct on the sizing column used) and monomers. This highly conserved threonine residue is pivotal in establishing multiple inter-helical and intra-helical contacts. Substitution of threonine's hydroxyl group with the hydrophobic ethyl group of isoleucine disrupts the intra-helical hydrogen bonding, whilst their identical methyl groups maintain the hydrophobic contacts established with second tyrosine of the $(YxxxG)_3$ motif on the opposing helix, thus consistent with the observed oligomeric behaviour (Martin et al. 2012).


A
kDa
Vo
669
443
200
158
67
29
Oligomeric MBP-YG box (WT)
MBP
Absorbance_280nm (mAU)
8.5
11.6
13.4
14.5
15.4
16.1
18.1
Elution Volume (mL)
MBP
SEC fractions
kDa
150
120
100
85
70
60
50
40
*
MBP- Ce YG169-207 (WT)
MBP
B
Oligomeric MBP-YG box (S181G)
MBP- Ce YG169-207 (S181G)
C
Oligomeric MBP-YG box (M182R)
MBP- Ce YG169-207 (M182R)
D
Oligomeric MBP-YG box (M182T)
MBP- Ce YG169-207 (M182T)
E
Monomeric MBP-YG box (S185P)
MBP- Ce YG169-207 (S185P)

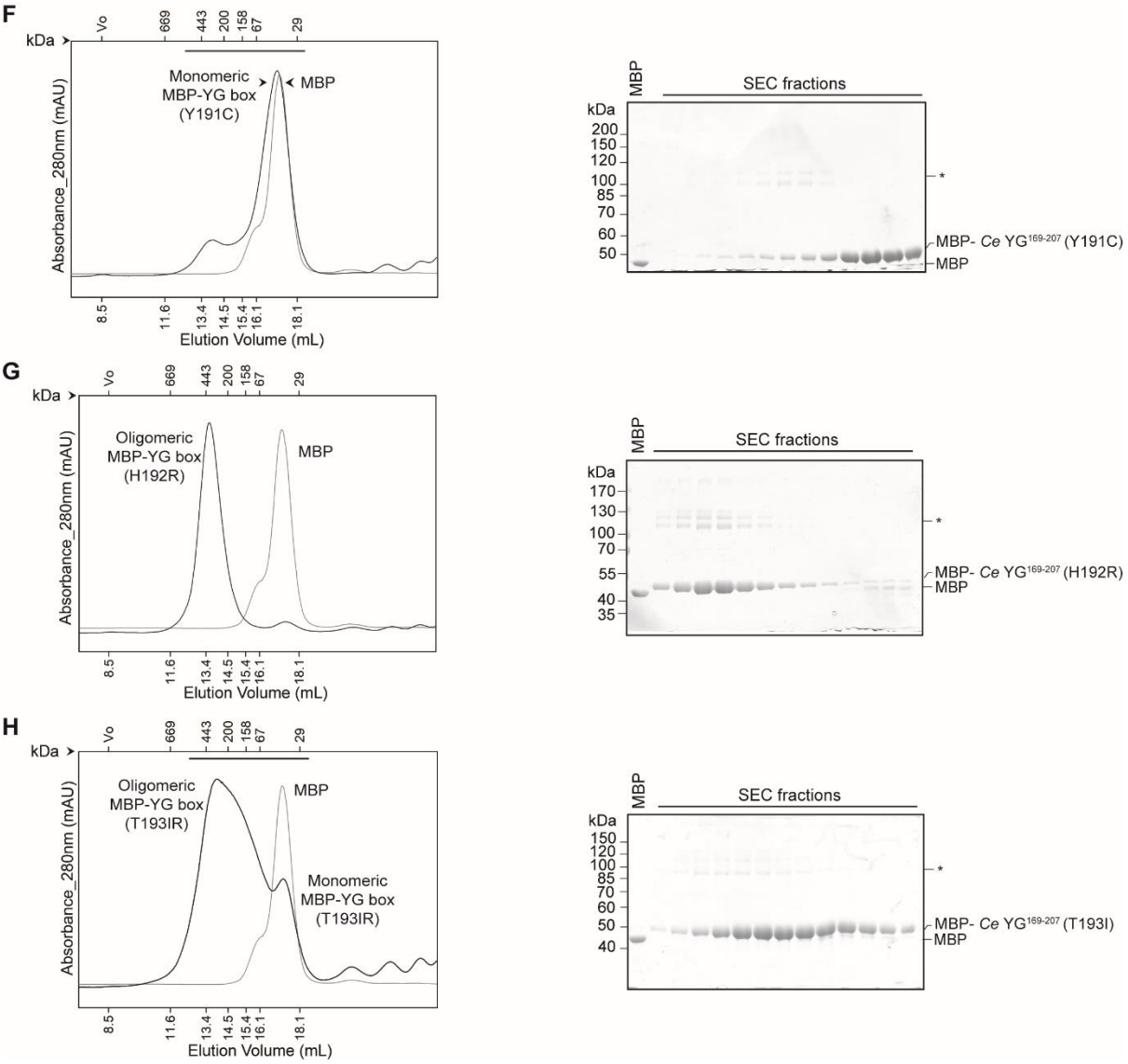


Figure 4.18 Effects of SMA patient mutations on MBP-*Ce* YG$^{169-207}$ oligomerisation
(A-H) SEC and SDS-PAGE analysis of SMA patient mutations bearing MBP-YG fusions.
(A) WT, **(B)** S181G, **(C)** M182R, **(D)** M182T, **(E)** S185P, **(F)** Y191C, **(G)** H192R, **(H)** T193I.
Left panel: SEC analysis of MBP-*Ce* YG$^{169-207}$ WT construct (M_r = 48.9 kDa; V_E = 13.24 mL) and mutants shown in black in comparison to the MBP alone elution profile (M_r = 44.5 kDa; V_E = 17.22 mL) in grey. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference. The black line above chromatogram depicts the range of the SEC fractions analysed by SDS-PAGE. *Right panel*: SDS-PAGE (8 % Tris-Glycine) of SEC elution profiles of the MBP fusions including the MBP protein as a reference for gauging any degradations. PageRuler^{TM} Unstained Protein ladder (200-10 kDa) (Section 3.1.8.4.2) was used as a reference for SDS-PAGE gels in (A), (B), (F) and (H). PageRuler^{TM} Prestained Protein ladder (170-10 kDa) (Section 3.1.8.4.2) was used for SDS-PAGE gels in (C), (D), (E) and (G). *Non-specific protein bands.

The oligomeric states predicted here for these MBP-fusion proteins are based on their retention volumes on the analytical HiLoad® 13/300 Superdex® 200 GL column in comparison to globular SEC calibration standards and are consistent with the SEC-MALS analysis reported by Martin et al., 2012. Taken together, the established MBP-YG fusions of the human and nematode orthologues containing SMA patient mutations demonstrating identical biochemical and structural behaviour reflects the highly conserved nature of SMN's YG-box domain (summarized in Table 4.1).

Table 4.1: Oligomeric status of MBP-*Hs* YG$^{252-284}$ and MBP-*Ce* YG$^{169-207}$ harbouring SMA patient mutations
[a] Oligomerisation states were categorised based on the relative elution volumes in comparison to the standard globular standards on the analytical HiLoad® 13/300 Superdex® 200 GL column.
[b] For each patient mutation, Type refers to the SMA severity which is also dependent on the *SMN2* gene copy number present in the patient. Adapted from Burghes and Beattie (2009).
ND: not determined. *: Mutant studied only in the nematode SMN system and the data for comparison is taken from Martin et al., 2012. **: Non-patient mutant oligomeric status examined only in the human SMN.

H. sapiens variant	*C. elegans* variant	Oligomeric status[a]		Type (*SMN2* copy number)[b]
		H. sapiens	*C. elegans*	
WT	WT	Octamer (Tetramer*)	Octamer	Normal
S262G*	S181G	Tetramer*	Octamer	III (1)
M263R	M182R	Monomer	Tetramer	I (1)
M263T	M182T	Monomer	Tetramer	II (2)
S266P	S185P	Monomer	Monomer	II (2)
S270E**	NA	Monomer	NA	Non-patient
Y272C	Y191C	Monomer	Monomer	I (2)
H273R	H192R	Octamer	Octamer	II (ND)
T274I	T193I	Monomer-Octamer	Monomer-Octamer	III (1)

4.5.2 Gemin8's interaction with SMN YG-box module is affected in the presence of SMA patient mutations

Gemin8 plays an integral role in building the core architecture of the heteromeric SMN complex by linking the SMN:Gemin2 and Gemin7:6 sub-complexes (Section 1.3, Figure 1.6) (Otter et al. 2007). Earlier *in vitro* experiments and the recently defined *S. pombe* SMN complex (unpublished data, RG Fischer) have delineated Gemin8's SMN interacting surface as the latter's highly conserved YG-box domain (Otter et al. 2007; Chari et al. 2008). Consequently, examining how Gemin8's engagement with

the oligomeric SMN is affected in the context of Spinal Muscular Atrophy (SMA) patient mutations was warranted.

For the aforementioned experimental purposes, biochemical handles of SMN's YG-box domain and Gemin8 had to be generated to study their association competency in isolation. While recombinant MBP-YG module system bearing SMA patient mutations could be readily established using the human and nematode SMN orthologues, generating a similar isolated Gemin8 biochemical handle was ineffectual. Even, generating a stable Gemin8 (fl) within the trimeric Gemin8:7:6 sub-complex was futile given the high susceptible to degradation or the absence of expression or insoluble expression despite multiple strategies of varying affinity tags, employing minimal truncation variants and/or expression strains (Sections 4.3.1 and 4.3.2). To resolve this issue, *in vitro* translated [^{35}S] labelled Gemin8 was employed to analyse its interaction with the MBP-YG fusion proteins harbouring SMA patient mutations. With the N-terminus of Gemin8 reported to be the binding site for SMN, [^{35}S] labelled Gemin8 lacking its SMN binding domain (ΔN) was also included in the subsequent interaction studies. However, only the nematode Gemin8 (fl) and the N-terminal truncation variant, Gemin8$^{57\text{-}199}$ (ΔN) constructs generated [^{35}S] labelled translation products. Reasoning from the previously described experiments wherein both the human and nematode SMN bearing the SMA patient mutations display identical oligomeric behaviour (Figure 4.17 and Figure 4.18 respectively, also see Table 4.1), immobilized nematode YG-box system (MBP-*Ce* YG$^{169\text{-}207}$; Figure 4.18) variants was used as bait to study their interaction with [^{35}S] labelled Gemin8 (fl) and Gemin8$^{57\text{-}199}$ (ΔN) translation products (Figure 4.19).

Robust binding of Gemin8 (fl) with the immobilised MBP-*Ce* YG$^{169\text{-}207}$ WT was detected but not for the N-terminal domain lacking Gemin8 variant (Figure 4.19 A and B, lanes 9 and 30 respectively), Gemin8$^{57\text{-}199}$ (ΔN). This implies *Ce* SMN's YG-box module as the direct binding site for *Ce* Gemin8's N-terminal domain (Figure 4.19 A and B, lanes 16 and 36 respectively), consistent with earlier reports (Otter et al. 2007; Chari et al. 2008) and the recently elucidated fission yeast SMN complex biochemistry (Section 4.5.3). While none of the SMA patient harbouring YG-box modules displayed association with the Gemin8$^{57\text{-}199}$ (ΔN) variant (Figure 4.19 A and B, lanes 16-20 and lanes 36-49 respectively), Gemin8 (fl) binding competency however varied. Also, the [^{35}S] labelled fl-Gemin8 and Gemin8$^{57\text{-}199}$ (ΔN) proteins showed no

association with MBP alone (Figure 4.19 A, lanes 9 and 15; Figure 4.19 B, lanes 29 and 35).

The solvent exposed Ser181 residue substitution (S181G) of the conserved SxxxSxxxSxxxT motif in the YG-zipper (Figure 4.16 A) that retained the oligomeric status similar to the wild-type (WT) fusion (compare in Figure 4.18, panels A and B), showed reduced Gemin8 binding (Figure 4.19 A, lane 11). Interestingly, the second serine (solvent-inaccessible residue) in this motif when substituted with the oligomerisation defective S185P mutation, no association of Gemin8 (fl) could be detected Figure 4.19 A, lane 14).

Interestingly, although patient mutations at Met182: M182R and M182T both disrupted SMN oligomerisation (Figure 4.18 C and D, respectively), only the Type II SMA variant M182T retained Gemin8 binding similar to the WT in contrast to the reduced association observed for the Type I SMA variant M182R (compare in Figure 4.19 A, lanes 13 and 12 respectively). While the M182T substitution only replaces the hydrophobic methionine side chain to a polar uncharged residue, the M182R variation introduces the positively charged side chain of arginine and, thus the plausible difference in Gemin8 association among the two SMA missense mutations at the identical residue and, consequently variable clinical SMA phenotypes.

Similarly, for the highly conserved tyrosine of the $(YxxG)_3$ motif in the dimer interface Y191 (Figure 4.16 A and E), the SMA Y191C mutant lacking SMN oligomerisation (Figure 4.18 F) also displays reduced binding to Gemin8 (fl) (Figure 4.19 B, lane 31). This is consistent with an earlier report of the human analogous SMN Y272C mutant showing ineffective incorporation into the SMN complex (Otter et al. 2007).

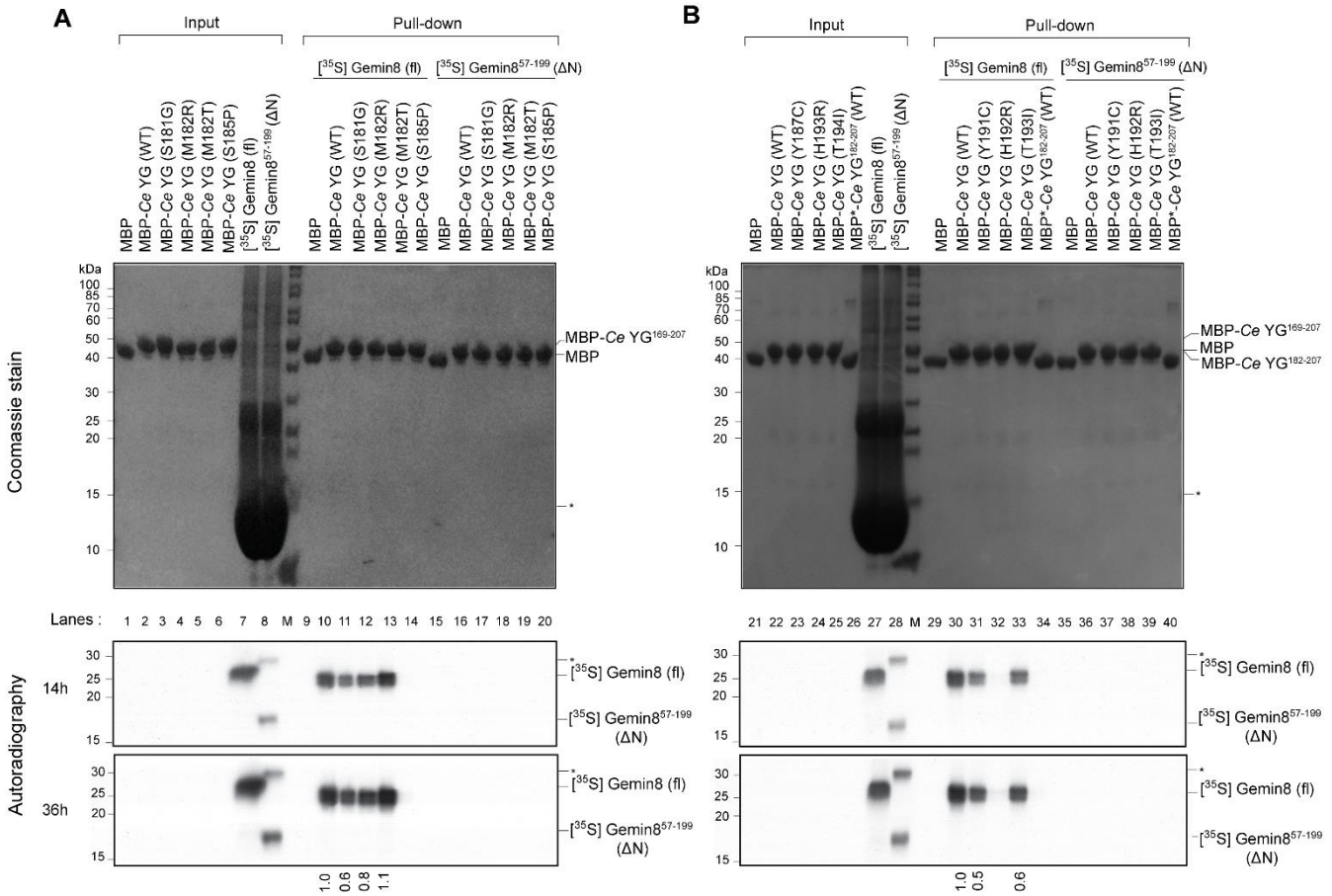


Figure 4.19 Gemin8 binding to SMN is affected by SMA patient mutations
(A and **B)** MBP-pulldown assay of SMN YG-box with Gemin8.
Recombinant MBP-fusions of *Ce*-YG$^{169-207}$ WT (lane 2 and 23) and indicated SMA patient mutants (lanes 3, 4, 5, 6, 23, 24 and 25), the dimeric crystallized MBP*-*Ce* YG$^{182-207}$ (lane 26) (see Figure 4.16) along with MBP as negative control (lanes 1 and 21) were immobilized on amylose resin and incubated with *in vitro* translated, [^{35}S]-labelled Gemin8 full-length (fl) (M_r = 23.3 kDa and Gemin8$^{57-199}$ (ΔN) (M_r = 16.6 kDa). After extensive washing, the bound proteins were eluted by boiling the beads in 1X SDS-loading buffer and resolved by SDS-PAGE on a 13.5 % Bis-Tris gel for visualisation by coomassie staining (top panels, lanes 1-40) and subsequently by autoradiography of the dried gels (middle and lower panels, lanes 1-40). The dried gels were subjected to 14 h (middle) and 36 h (lower) exposure periods and the relative quantifications are indicated below the respective lanes. 10 % input of [^{35}S]-labelled Gemin8 (fl) (lanes 7 and 27) and Gemin8$^{57-199}$ (ΔN) (lanes 8 and 28) are indicated. The total panel shows 5 % of the proteins used for binding. M: PageRuler^{TM} Unstained Protein ladder (200-10 kDa) (Section 3.1.8.4.2).

Interestingly for the H192R mutant despite demonstrating similar oligomeric status to that of the wild-type (WT) YG-box (Figure 4.18 G), no Gemin8 binding was detected (Figure 4.19 B, lane 32). The His193 residue does not participate in the classical YG dimerization interface and, is thus solvent exposed and available for Gemin8 association with any substitutions affecting this binding. On the contrary, for the mildly oligomerisation defective T193I mutant forming a wide range of oligomeric states, only slightly reduced Gemin8 (fl) binding was observed (Figure 4.19 B, lane 33).

Further, to address the boundaries of the YG-box domain in the context of Gemin8 binding, a shorter MBP-YG fusion, MBP*-*Ce* YG$^{182-207}$ (Figure 4.16), was also included in this interaction assay. This crystallized MBP*-*Ce* YG$^{182-207}$ variant differed from the MBP-*Ce* YG$^{169-207}$ construct in possessing a shorter linker between the MBP tag (for details on the MBP* construct, see Section 4.5.1.1) as well a shorter YG-box module. Interestingly, the dimeric MBP*-*Ce* YG$^{182-207}$ fusion although lacking any of the SMA mutations, showed no engagement with Gemin8 (fl) (Figure 4.19 B, lane 34). This could arise due to the steric hindrance posed by the close proximity of the tag or the shorter YG-module only beginning with the Met183 residue. The latter notion might be plausible, given that one of the residues,S181, lacking in the short *Ce* YG$^{182-207}$ fusion is essential for Gemin8 association since a SMA mutation at this residue shows reduced Gemin8 binding (Figure 4.19 A, lane 11).

In conclusion, Gemin8's N-terminus is essential for interaction with SMN's YG-box domain. Additionally, these results imply that SMN's oligomeric status probably does not directly influence Gemin8 binding but rather the accessibility and/or relevance of the residues involved within SMN's YG-box domain.

4.5.3 The fission yeast *S. pombe* (*Sp*) Gemin8: SMN interaction interface

The preceding experiments (including those not presented here) establish considerable difficulties encountered in biochemically establishing an isolated and stable Gemin8 biochemical handle and/or a Gemin8:7:6 sub-complex using either the human or nematode orthologues for structurally charting the SMN: Gemin8 interface. Therefore, in an attempt to circumvent these inherent difficulties for crystallographic pursuits, the fission yeast *Schizosaccharomyces pombe (S. pombe)* SMN complex was probed.

Till date, the Yab8p and Yip1p proteins were the only known functional orthologues of the human SMN complex components, SMN and Gemin2 respectively,

in the fission yeast (Hannus et al. 2000). Adding to this minimal complex, are the recently elucidated *Sp* Gemins 6, 7 and 8 (unpublished data, RG Fischer and Section 4.3.4). Akin to the pentameric human SMN complex (Figure 4.1 B), biochemical studies have also lately established an analogous pentameric *Sp* SMN complex comprising of the fission yeast SMN, Gemin2, Gemin8, Gemin7 and Gemin6 proteins (unpublished data, RG Fischer).

To permit structural studies to proceed, minimal variants of *Sp* SMN and *Sp* Gemin8 harbouring the binding domains that are necessary and sufficient for establishing stable sub-complexes were employed. The minimal *Sp* SMN (*Sp* SMNΔL) includes the N-terminal Gemin2 (G2) binding domain (residues 1-35) linked to the C-terminal Gemin8 (G8) binding domain (residues 120-152 harbouring the oligomerisation inducing YG-box) by 9 aa linker (Figure 4.20 A and B). SEC analysis of this 8.1 kDa minimal variant indicates a higher order oligomeric species in solution with an apparent molecular mass of >50 kDa (hexamer) based on calibration with globular standards on a Superose® 6 10/300 GL column (Figure 4.20 C). This is in broad agreement with the SAXS measurements of *Sp* SMNΔL (unpublished data, RG Fischer), thus excluding an unusually large Stokes radius for the early SEC elution time.

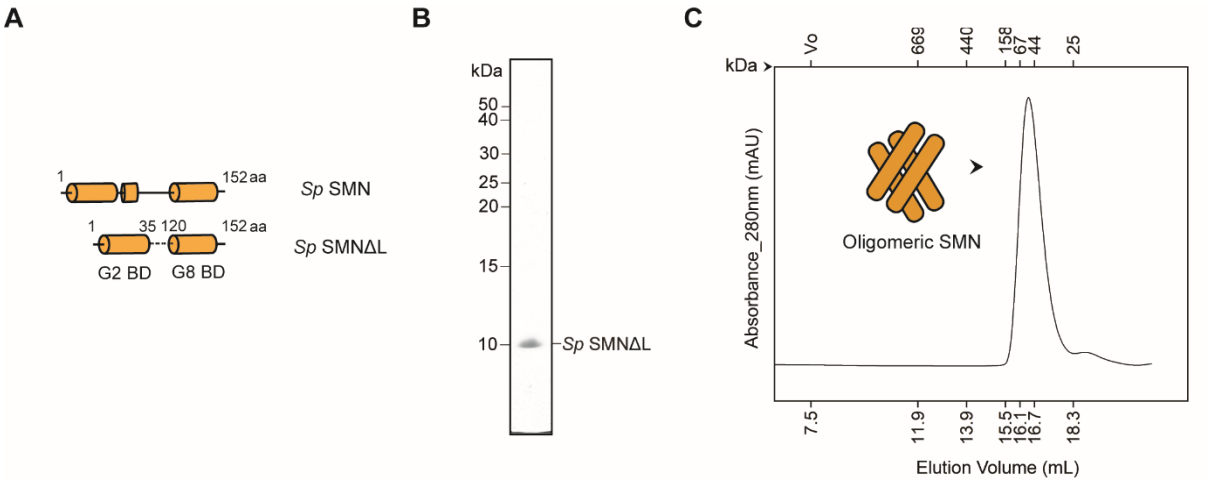


Figure 4.20 The minimal *S. pombe (Sp)* SMNΔL variant
(A) Schematic of the domain architecture of the minimal *Sp* SMNΔL in comparison to its full-length counterpart. The minimal *Sp* SMNΔL includes the N-terminal Gemin2 (G2) binding domain (BD) and the C-terminal Gemin8 (G8) binding domain (BD) harbouring of the oligomerization inducing YG box. The secondary structural elements (based on PSIPRED output): α-helices and coils are rendered as cylinders and straight lines respectively, to scale. The binding domains within the minimal variant are linked by a 9 amino acid (aa) linker which is shown here in dotted lines.
(B) The recombinantly established minimal *Sp* SMNΔL protein visualised on a coomassie stained 15 % Tris-Tricine gel.
(C) The highly oligomeric status of *Sp* SMNΔL based on SEC analysis on a Superose® 6 10/300 GL column (M_r = 8.1 kDa; V_E = 16.45 mL). The inset in **(C)** offers a schematic visualization of the oligomeric species formed as speculated in Gupta et al., 2015. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference.

Similarly, the minimal *Sp* Gemin8 variant (*Sp* Gemin8ΔL) included only the N-terminal SMN (residues 1-35) and the C-terminal Gemin7 (residues 115-166) binding domains sufficient for establishing a stable tetrameric *Sp* SMN complex of SMN and Gemins 6-8 (Figure 4.21 A). In an attempt to clearly define the Gemin8's N-terminal SMN binding domain as well as intending to establish stable *Sp* SMN tetrameric sub-complex (SMN:Gemins8-6) variants with high crystallization propensity, N-terminal truncation variants of the minimal *Sp* Gemin8 were generated (Figure 4.21 A). *Sp* Gemin8ΔL and the associated truncation variants co-purified with *Sp* Gemins 6-7 (Figure 4.21 B) forming monomeric sub-complexes as demonstrated by SEC analysis (Figure 4.21 C). On this note, the *Sp* Gemin8ΔN20ΔL variant could not be recombinantly co-expressed with Gemins 6-7 (data not shown).

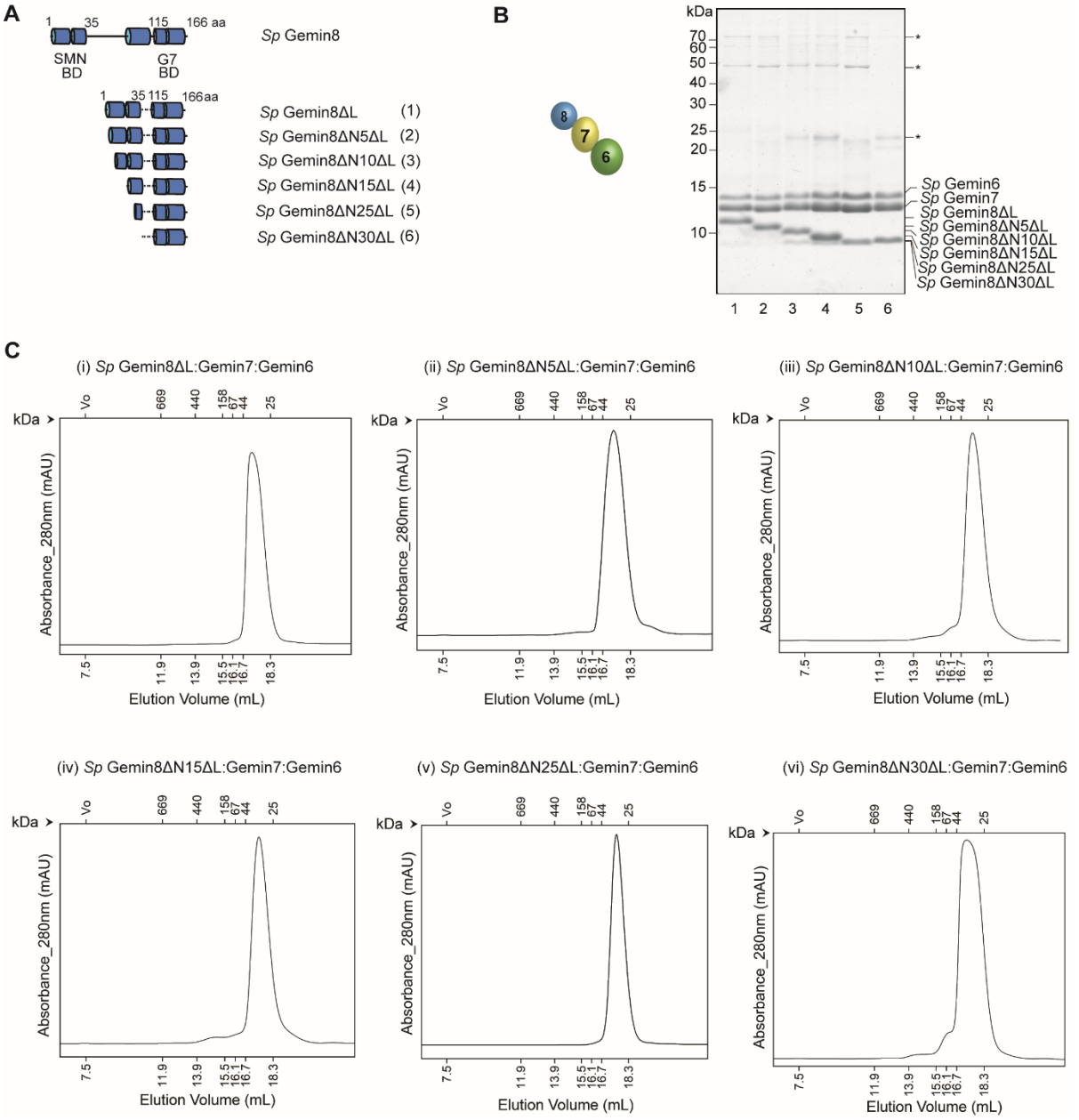
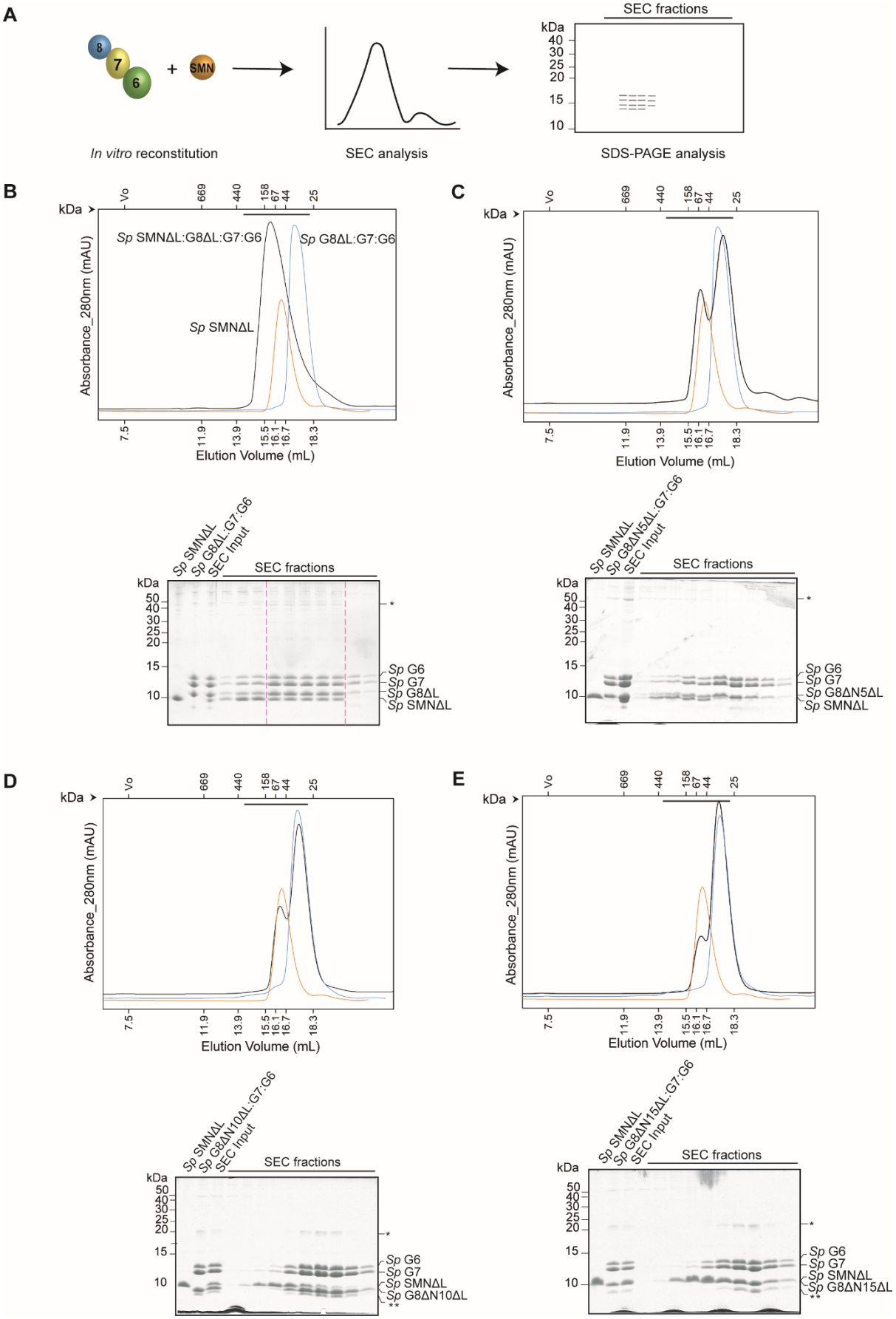


Figure 4.21 N-terminal truncation variants of the *S. pombe (Sp)* Gemin8ΔL:7:6 sub-complex
(A) Schematic of the domain architecture of the minimal *Sp* Gemin8ΔL (1) in comparison to its full-length counterpart. The minimal *Sp* Gemin8ΔL comprises the N-terminal SMN binding domain (BD) and the C-terminal Gemin7 (G7) binding domain (BD). The associated N-terminal truncation variants of the minimal *Sp* Gemin8ΔL are depicted to scale (2-6).The secondary structural elements (based on PSIPRED output): α-helices and coils are rendered as cylinders and straight lines respectively, to scale. The short linker engaging the minimal binding domains is shown as dotted lines.
(B) *Left panel*: Schematic of the formed trimeric sub-complex comprising of Gemins 6-8. *Right panel*: The minimal *Sp* Gemin8ΔL and the associated N-terminal truncation variants were co-expressed in *E. coli* with *Sp* Gemins 7-6 and purified by IMAC. The resultant trimeric complexes were visualised on a coomassie stained 15 % Tris-Tricine gel. The lane numbers correspond to the numbers assigned to the Gemin8 variants depicted in panel **(A)**.
(C) (i-vi) The generated trimeric *Sp* Gemin8ΔL:Gemin7:Gemin6 sub-complex variants demonstrate monodisperse peaks of monomeric nature on a Superose® 6 10/300 GL column. The elution positions of conventional globular protein SEC calibration standards are indicated on the upper X-axis for comparison.

While the minimal *Sp* Gemin8 N-terminal truncation variants retaining their association with *Sp* Gemin7, substantiates the previously detailed Gemin8:Gemin7 interaction interface (Section 4.3), it was pertinent to determine which among these established minimal *Sp* Gemin8ΔL N-terminal truncation variants would be sufficient to engage with *Sp* SMNΔL (Figure 4.21 A) to permit crystallization trials. To address this issue, the recombinant trimeric *Sp* Gemin8ΔL:7:6 and the associated variant sub-complexes thereof were incubated with equimolar amounts of *Sp* SMNΔL followed by SEC and SDS-PAGE analysis for monitoring the putative complexes formed (Figure 4.22 A).

As anticipated the trimeric *Sp* Gemin8ΔL:7:6 sub-complex engages with the minimal *Sp* SMNΔL forming the stable tetrameric *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex as evidenced by the SEC and SDS-PAGE analysis (Figure 4.22 B, top and middle panels). This *in vitro* reconstituted tetrameric complex (Figure 4.22 B, bottom panel) demonstrates an earlier SEC retention volume (Figure 4.22 B, top panel, elution profile in black) in comparison to the oligomeric *Sp* SMNΔL (Figure 4.22 B, top panel, elution profile in orange) and the monomeric *Sp* Gemin8ΔL:7:6 sub-complex (Figure 4.22 B, top panel, elution profile in blue) on an analytical Superose® 6 10/300 GL column. Hence, both the minimal *Sp* SMNΔL and Gemin8ΔL variants harbour the essential binding domains that are sufficient for establishing the tetrameric *Sp* SMN complex comprising of SMN and Gemins 6-8. In contrast, even the loss of the first five N-terminal residues on Gemin8ΔL (Gemin8ΔN5ΔL), although predicted to be unstructured, disrupts SMN association (Figure 4.22 C). Thus, either emphasizing critical role of these residues in SMN engagement or the truncation boundary used here occurs in close proximity to the SMN binding domain, thus hampering stable complexation. Likewise, the other longer N-terminal truncations of Gemin8ΔL spanning by a 5 amino acid window resulted in loss of SMN interaction (Figure 4.22 C, D, E, F and G). Although this attempt in precisely defining the boundaries of the SMN-binding N-terminal domain on Gemin8 for crystallographic pursuits lent itself to be ineffectual, the stable minimal *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex (Figure 4.22 B) was subjected for crystallization trials that unfortunately proved to be inconsequential.


A
8
7
6
+
SMN
In vitro reconstitution
SEC analysis
SEC fractions
kDa
40
30
25
20
15
10
SDS-PAGE analysis
B
C
kDa
Vo
669
440
158
67
44
25
Absorbance_280nm (mAU)
Sp SMNΔL:G8ΔL:G7:G6
Sp G8ΔL:G7:G6
Sp SMNΔL
7.5
11.9
13.9
15.5
16.1
16.7
18.3
Elution Volume (mL)
Sp SMNΔL
Sp G8ΔL:G7:G6
SEC Input
SEC fractions
50
40
30
25
20
15
10
*
Sp G6
Sp G7
Sp G8ΔL
Sp SMNΔL
Sp G8ΔN5ΔL:G7:G6
Sp G8ΔN5ΔL
D
E
Sp G8ΔN10ΔL:G7:G6
Sp G8ΔN10ΔL
Sp G8ΔN15ΔL:G7:G6
Sp G8ΔN15ΔL
**

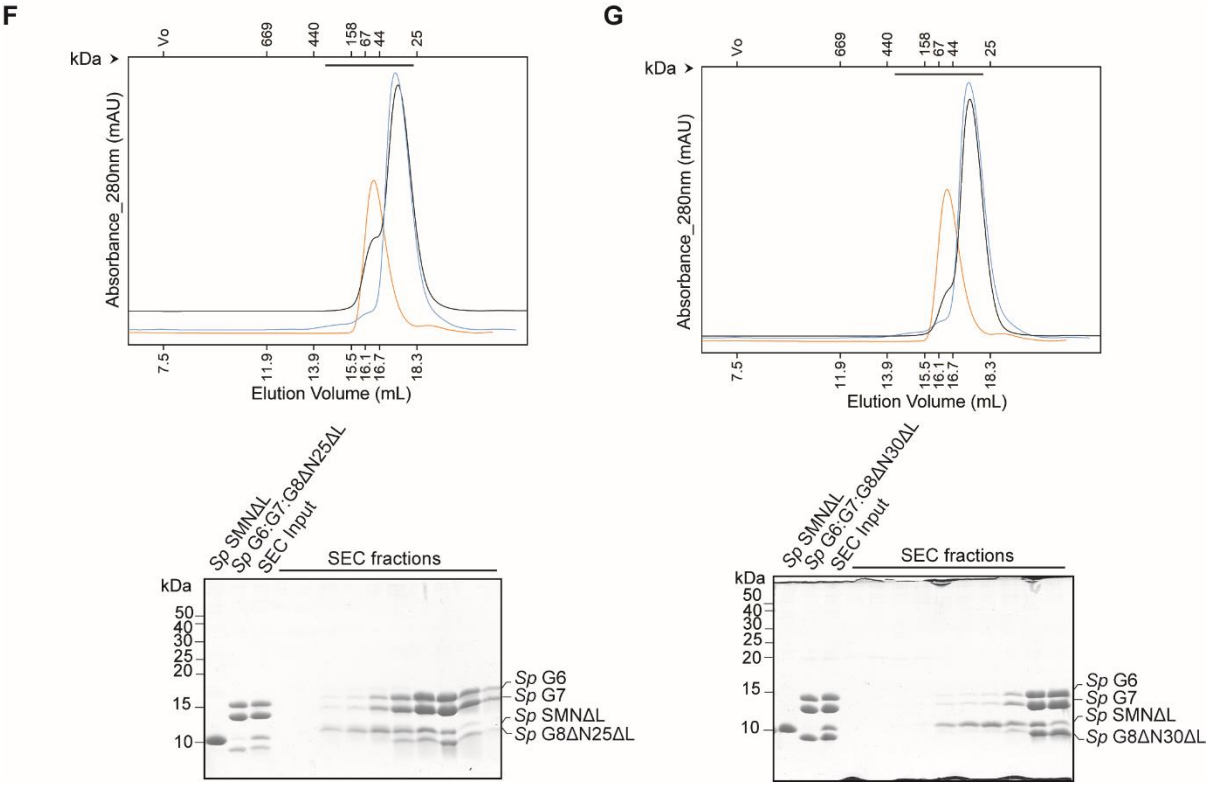


Figure 4.22 Reconstitution of the tetrameric *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex
(A) Schematic of the experimental outline used for monitoring the *in vitro* reconstitution of *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex variants.
(B-G) Top panel: Elution profiles of the reconstituted complexes on an analytical Superose® 6 10/300 GL column. The elution profiles of *Sp* SMNΔL (orange) and *Sp* Gemin8ΔL:7:6 sub-complex and/or associated variants (blue) are depicted alongside the elution profile of the respective formed complexes (black) for comparison. The elution positions of the conventional globular protein SEC calibration standards are indicated on the upper X-axis for reference. The black line above chromatogram depicts the range of the SEC fractions analysed by SDS-PAGE (see bottom panel). Bottom panel: SDS-PAGE analysis of the SEC fractions visualised on a coomassie stained 15 % Tris-Tricine gel. Of note, in **(B)** the stable tetrameric *Sp* SMNΔL:Gemin8ΔL:7:6 sub-complex formed is highlighted. The SEC input refers to the putative *in vitro* reconstituted sub-complexes subjected to SEC analysis. Sample inputs for the *Sp* SMNΔL protein and *Sp* Gemin8ΔL:7:6 sub-complex variants used for the respective *in vitro* reconstitutions are also indicated.

Table 4.2 Crystallographic data collection statistics for *Hs* Gemin7$^{31-131}$:6 (fl) (Section 4.2.1), *Hs* Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$ (Section 4.3.3) and *Ce* MBP*-YG$^{182-207}$ (Section 4.5.1.2)
Numbers in parentheses represent values in the highest resolution shell.

Protein	***Hs* Gemin7$^{31-131}$:6 (fl)**	***Hs* Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$**	***Ce* MBP*-YG$^{182-207}$**
Data Collection			
Wavelength (Å)	0.97625	0.97625	0.918400
Space group	P 6_1 2 2	P 2 2_1 2_1	P 1 2_1 1
Cell dimensions: a, b, c (Å)	107.936, 107.936, 289.09	59.88, 80.597, 82.666	66.084, 74.857, 104.375
Cell dimensions: α, β, γ (º)	90, 90, 120	90, 90, 90	90, 105.961, 90
Molecules in asymmetric unit	2	2	1
Resolution (Å)	49.17 - 3.11 (3.222 - 3.11)	48.07 - 1.521 (1.576 - 1.521)	27.05 - 2.215 (2.294 - 2.215)
Unique reflections	18710	58111	180160
R_{sym} (%)	20.5 (185.9)	6.2 (72.4)	10.7 (192.5)
Mean I/σ (I)	10.67 (1.78)	10.73 (1.19)	7.30 (0.60)
Completeness (%)	99.24 (96.20)	93.55 (70.86)	97.05 (87.48)
Redundancy	0.9173	3.8	3.7

Table 4.3 Crystallographic refinement statistics for *Hs* Gemin7$^{31-131}$:6 (fl) (Section 4.2.1), *Hs* Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$ (Section 4.3.3) and *Ce* MBP*-YG$^{182-207}$ (Section 4.5.1.2)

RMSD: root-mean square deviation. Numbers in parentheses represent values in the highest resolution shell.

Protein	***Hs* Gemin7$^{31-131}$:6 (fl)**	***Hs* Gemin8$^{190-230}$:7$^{46-131}$:6$^{1-92}$**	***Ce* MBP*-YG$^{182-207}$**
Refinement Statistics			
Resolution	49.17 - 3.11 (3.222 - 3.11)	48.07 - 1.521 (1.576 - 1.521)	27.05 - 2.215 (2.294 - 2.215)
Number of reflections	18613 (1748)	58076 (4340)	47405 (4201)
R/R_{free}	0.2657/0.2873	0.2087/0.2441	0.2770/0.3248
Number of atoms: Protein	2696	3395	6098
Number of atoms: ligands/ions	-	-	-
Number of atoms: water	-	399	82
Rmsd: bond lengths (Å)	0.006	0.011	0.011
Rmsd: bond angles (º)	1.24	1.04	1.43
Ramachandran: favoured, allowed, outliers (%)	92.42, 6.97, 0.61	98.04, 1.96, 0.00	94.69, 5.18, 0.13
PDB ID	N/A	N/A	N/A

5 Discussion

The assembly of the spliceosomal U snRNPs is a remarkable feat spread across the nuclear and cytoplasmic compartments requiring the seven Sm proteins brought together to form the Sm core on the U snRNAs (Gruss et al. 2017). Thus, to ensure an ordered Sm core assembly and to preclude illicit RNA loading with the assembled Sm core, the PRMT5 and SMN complexes orchestrate the cytoplasmic U snRNP assembly pathway (Chari and Fischer 2010). During the late cytoplasmic phase of U snRNP biogenesis, the multi-subunit SMN complex gathers the preformed Sm core intermediates from the PRMT5 complex and ensures faithful Sm core assembly on the Sm site of the U snRNAs (Paushkin et al. 2002). While molecular and structural studies have defined the role of the distinct protein subunits and/or individual sub-complexes of the SMN complex (SMN, Gemins 2-8 and unrip), in orchestrating the U snRNP assembly pathway (Zhang et al. 2011; Yong et al. 2010; Chari et al. 2008), how these distinct modules coalesce to form the multi-subunit SMN complex remains opaque. In this pursuit, the present study employing biochemical and structural studies of SMN complex orthologues from divergent organisms (*H. sapiens*, *C. elegans*, *S. pombe*) provides insights on the structural framework of the SMN complex and the influence of SMA patient mutations in the organization of the SMN complex forming subunit modules.

5.1 Conservation of the Gemin7:6 module in nematodes and humans

The Gemins 7 and 6 form a distinct SMN-independent sub-complex both *in vivo* and *in vitro* (Battle et al. 2007; Ma et al. 2005; Carissimi, Saieva, Gabanella, et al. 2006). Therefore, initially the Gemin7:6 module was established as a building block towards understanding the assembly of the SMN complex, full-length and truncation variants of both the *H. sapiens* (*Hs*) and *C. elegans* (*Ce*) Gemin7 and 6 orthologues were expressed for our biochemical and structural studies (Section 4.2). Of all the recombinant human Gemin7:6 variants generated, only the Gemin7$^{31\text{-}131}$:fl-Gemin6 complex generated protein crystals. Notably, X-ray diffraction quality crystals for this complex could only be obtained under the previously reported crystallization conditions albeit with a higher precipitant concentration (Ma et al. 2005). Additionally, our dimeric Gemin7:6 structural model is in excellent agreement with the previously described model (PDB

ID: 1y96) (Ma et al. 2005), evident from the absence of any global structural variations (RMSD= 0.613) (Section 4.2.1, Figure 4.3). The *Hs* Gemins 6 and 7 although lacking sequence similarity with the Sm proteins, share a similar Sm-like fold: a five stranded β sheet flanked by an N-terminal α helix (Figure 1.3). The head-to-tail dimerization of the *Hs* Gemin7:6 sub-complex is afforded via their C- and N-terminal Sm-like folds respectively, wherein the outermost β strands of the Sm fold domains engage to form a continuous 10-stranded anti-parallel β sheet (Figure 4.3). Interestingly, the Sm heteromer-like structural mimicry of the Gemin7:6 sub-complex has been proposed to aid during the step-wise assembly of the Sm-core on the U snRNAs orchestrated by the SMN complex (Ma et al. 2005).

Like their human counterparts, the previously uncharacterised *C. elegans* (*Ce*) Gemin7 and Gemin6 orthologues also form a heterodimer in a SMN-independent manner as discerned from *in vitro* biochemical experiments and SEC analysis. Notably, while the longer N-terminus of the human Gemin7 preceding the Gemin6 interaction domain was absent in the nematode variant, the largely unstructured C-terminus of both the human and nematode Gemin6 was dispensable for Gemin7 interaction (Figure 4.5). Unfortunately for our structural studies the *Ce* Gemin7:6 sub-complexes failed to generate protein crystals, despite several crystallization attempts involving different parameters. However from bioinformatic analysis, it could be discerned that these orthologues share identical secondary structural elements adopting the Sm-like fold with significant amino acid sequence conservation at the predicted Gemin7 (β5 strand): 6 (β4 strand) dimerization interface to their human orthologues (Figure 5.1 and Figure 5.2).

Taken together with the identical biochemical behaviour and (predicted) structural homology, the dimerization of the novel nematode Gemin7:6 module via their conserved Sm-like fold domains analogous to their human counterparts is conceivable. Nevertheless, both the thus established human and nematode Gemin7:6 sub-complex variants facilitated in building up the rest of the core SMN complex for downstream structural studies.

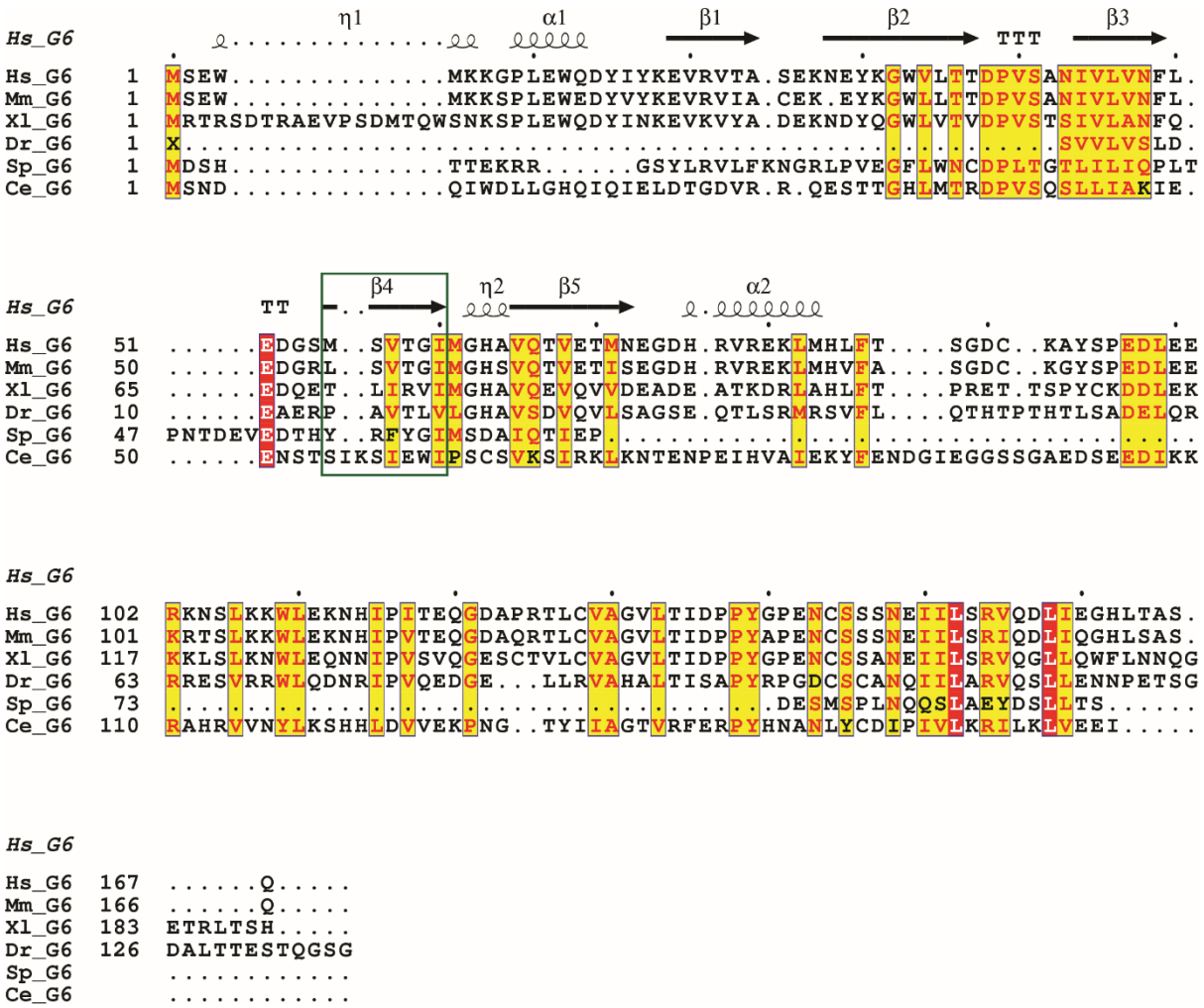


Figure 5.1 Sequence conservation of Gemin6 orthologues in diverse organisms
Alignment of Gemin6 orthologues from diverse organisms. The secondary structure elements of the *H. sapiens* Gemin6 (PDB code: 1y96, chain A) is displayed above the sequences. α-helices and 3_{10}-helices (η) are displayed as squiggles, β-strands are rendered as arrows, strict β-turns as TT letters and strict α-turns as TTT. Gemin6's C-terminus adopts a Sm-like fold consisting of an N-terminal α-helix (α1) followed by five β strands (β1-β5). The β4 strand of Gemin6 (green box) engages with the β5 strand of the Sm-like fold of Gemin7 (see Figure 5.2) for dimerization (Section 4.2.1).
Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).
Hs: *H. sapiens,* Mm: *M. musculus,* Xl: *X. laevis,* Dr: *D. rerio,* Sp: *S pombe,* Ce: *C. elegans* (Refer Section 7.5 for UniProtKB codes).

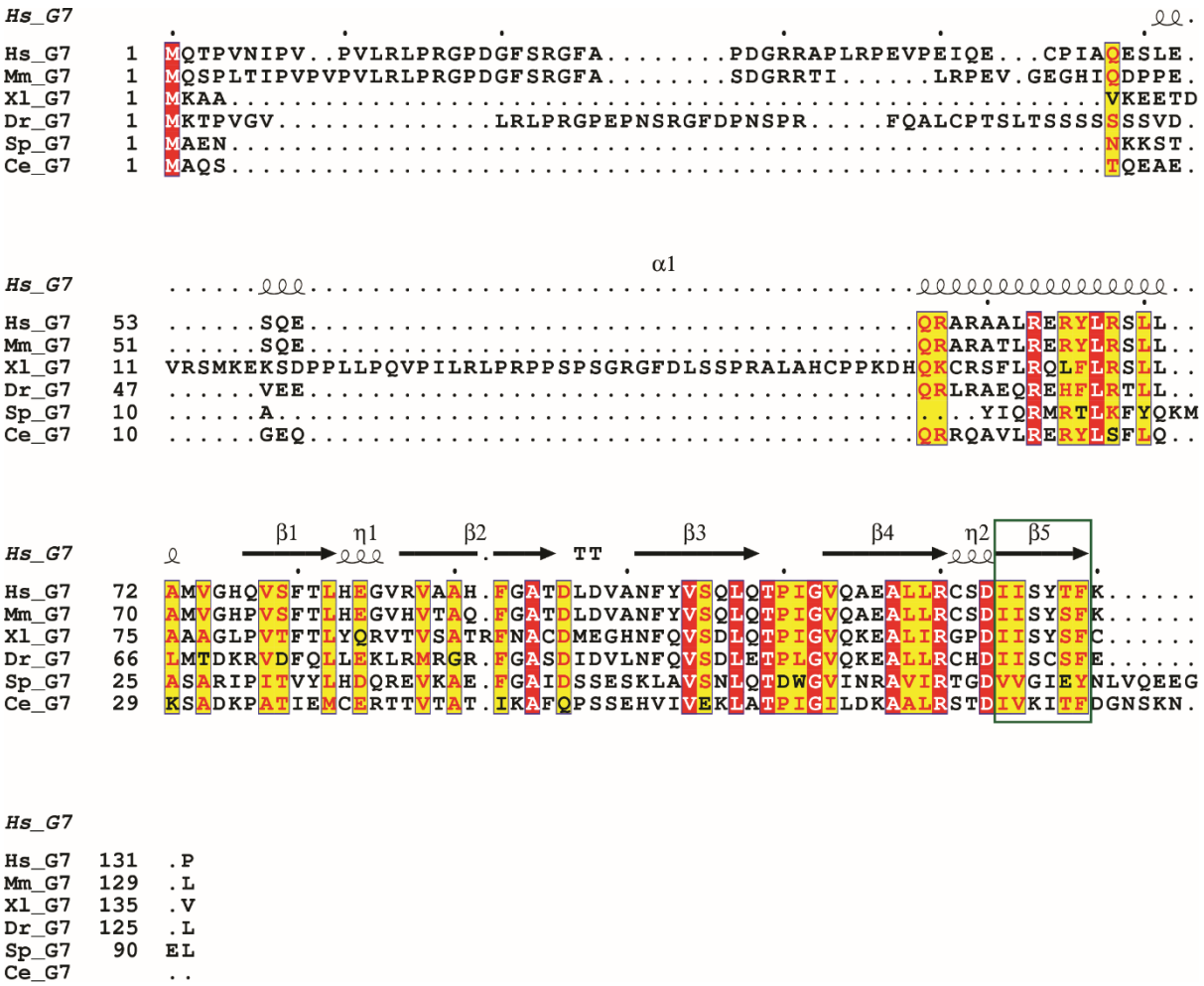


Figure 5.2 Sequence conservation of Gemin7 orthologues in diverse organisms
Alignment of Gemin7 orthologues from diverse organisms. The secondary structure elements of the *H. sapiens* Gemin7 (PDB code: 1y96, chain B) is displayed above the sequences. α-helices and 3_{10}-helices (η) are displayed as squiggles, β-strands are rendered as arrows, strict β-turns as TT letters and strict α-turns as TTT. Gemin7's N-terminus folds into a Sm-like fold domain consisting of an N-terminal α-helix (α1) followed by five β strands (β1-β5). The β5 strand of Gemin7 (green box) engages with the β4 strand of the Sm-like fold of Gemin6 (see Figure 5.1) achieving the Gemin7:6 dimeric sub-complex (Section 4.2.1).
Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).
Hs: *H. sapiens,* Mm: *M. musculus,* Xl: *X. laevis,* Dr: *D. rerio,* Sp: S *pombe,* Ce: *C. elegans* (see Section 7.5 for UniProtKB codes).

5.2 The Gemin7:6 sub-complex tethers to the SMN complex via Gemin8's conserved C-terminal helix

Co-immunoprecipitation and immuno-localization experiments identified Gemin8 as an integral component of the SMN complex (Carissimi, Saieva, Baccon, et al. 2006). However, *in silico* analyses of Gemin8 revealed no distinct protein motifs or significant homology to other proteins to hint at a putative function. Notably, the peripheral Gemin7:6 module is sequestered to the SMN complex via Gemin7's direct association

with Gemin8 (Figure 1.9) (Otter et al. 2007). Therefore in this work, the Gemin8:7 interaction interface was delineated biochemically and structurally by building up on the previously established Gemin7:6 modules with Gemin8 full-length and truncation variants.

Bacterial recombinant expression of both the human and nematode Gemin8 full-length (fl) protein in isolation could not be achieved. Hence for experimental purposes, bacterial co-expression systems of Gemin8 with its cognate binding partner within the SMN complex, namely Gemin7 were established. However, even in the presence of its interaction partner, Gemin7, recombinant expression of the human (fl)-Gemin8 was not feasible. On the other hand, the putative nematode (fl)-Gemin8 protein formed a soluble trimeric complex with Gemin7:6 module, albeit exhibiting a rapid degradation propensity over time. Consequently, N-terminal truncation variants of Gemin8 were generated and assessed for their ability to form stable trimeric complexes with the Gemin7:6 module for the downstream crystallographic studies. Interestingly, Gemin8's C-terminal domain alone was sufficient for direct Gemin7 association (Section 4.3). Furthermore, this biochemical behaviour was shared in the established human, nematode and fission yeast Gemin8:7:6 trimeric sub-complexes.

Among the different trimeric Gemin8:76 sub-complex variants, only the minimal *Hs* Gemin8$^{190\text{-}230}$:7$^{46\text{-}131}$:6$^{1\text{-}92}$ and *Sp* Gemin8ΔN30ΔL:7:6 sub-complexes yielded X-ray diffraction quality protein crystals. However, only the structure of the *Hs* Gemin8$^{190\text{-}230}$:7$^{46\text{-}131}$:6$^{1\text{-}92}$ could be solved. Within the minimal *Hs* Gemin8$^{190\text{-}230}$:7$^{46\text{-}131}$:6$^{1\text{-}92}$ sub-complex revealed that Gemin7's N-terminal helix (α1) of the Sm-like fold engages directly with the C-terminal helices (α1 and α2) of Gemin8 via polar and salt bridge interactions (Figure 4.11). Notably, Sm-fold domain of Gemin7 interacts with both Gemin8 and Gemin6 via its N- and C-termini respectively. Additionally, the identical biochemical mode of Gemin8:7 interaction in divergent organisms is also reflected in the evolutionary sequence conservation of the interface interacting residues (Figure 5.3).

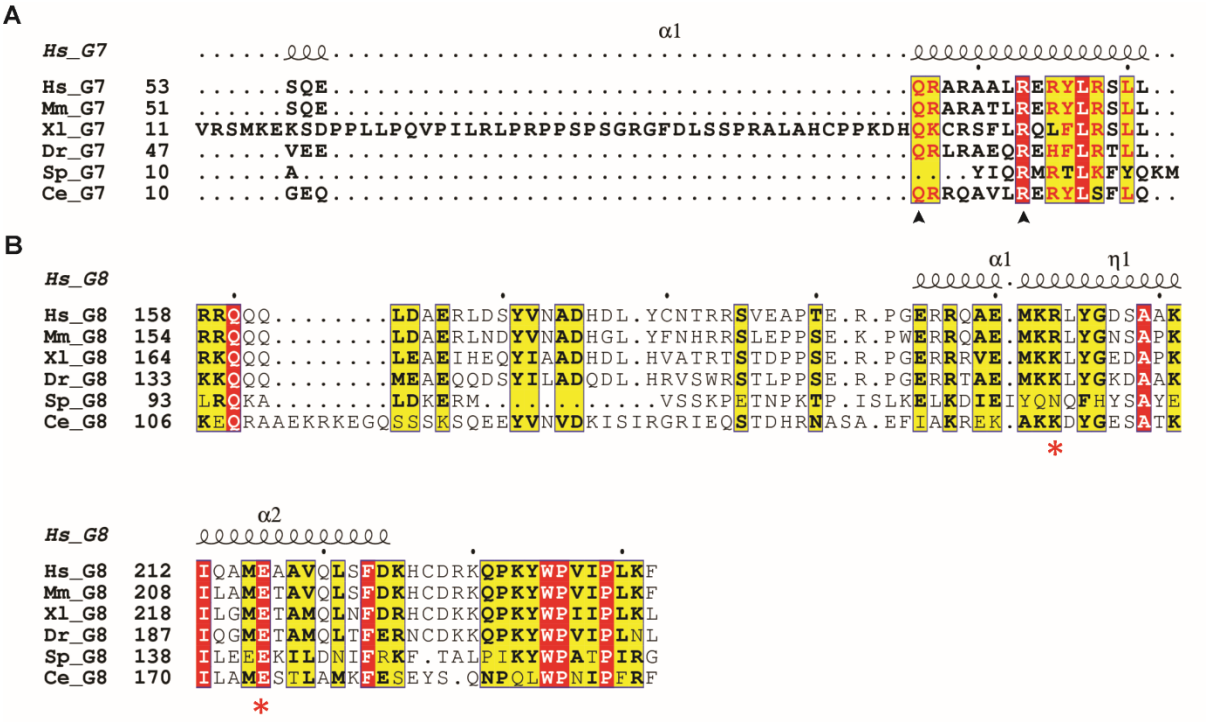


Figure 5.3 Conservation of Gemin8:7 interaction interface in diverse organisms
(A) Multiple sequence alignment of N-terminus of Gemin7 orthologues from diverse organisms. The significantly conserved Gln(Q)56 and Arg(R)63 residues (arrow heads) at the N-terminus α-helix (α1) of Gemin7's Sm-like fold engages with Gemin8's C-terminus (see panel B) via polar and salt bridge interactions.
(B) Multiple sequence alignment of C-terminus of Gemin8 orthologues from diverse organisms. The C-terminus α-helices (α1 and α2*)* of *H. sapiens* Gemin8 (see Section 4.3.3) is displayed above the sequences as squiggles. The highly conserved Arg(R)203 and Glu(E)216 residues (asterisks) at the C-terminus α-helices (α1 and α2) of Gemin8's C-terminus interacts with Gemin7's N-terminus (see panel A) via polar and salt bridge interactions. Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).
Hs: *H. sapiens,* Mm: *M. musculus,* Xl: *X. laevis,* Dr: *D. rerio,* Sp: *S pombe,* Ce: *C. elegans* (see Section 7.5 for UniProtKB codes).

Interestingly, the *Hs* Gemin7 N-terminus region preceding its Sm-like fold domain is longer in comparison to the other metazoan Gemin7 orthologues (Figure 5.2). Apart from Gemins 6 and 8, *Hs* Gemin7 also engages with the WD-repeat protein unrip incorporating the latter into the SMN complex (Figure 1.9) (Otter et al. 2007). Notably *in vitro* binding assays concluded that unrip is tethered to the SMN complex via Gemin7's N-terminal domain (1-56 residues) (Grimmler et al. 2005), the region partly overlapping with the now delineated Gemin8's binding site on Gemin7. In this work, it was assessed that the absence of the largely unstructured *Hs* Gemin7's N-terminus comprising of residues 1-45 abolishes unrip association (Appendix 7.1). Consequently, it could be inferred that the unrip binding domain might be located upstream to Gemin8's Sm-fold domain (the α1 helix) binding domain Gemin7 (Figure 6.1 B). Although, unrip orthologues have so far been identified only in mouse and drosophila, its SMN complex interaction partner Gemin7 is either lacking the delineated N-terminus

unrip binding domain or has not been annotated yet, respectively. This lack of evolutionary conservation is reflected in unrip's diverse functions and its conditional association with the human SMN complex only in the cytoplasm, in contrast to the other components of the SMN complex inventory (Gemins) displaying both cytoplasmic and nuclear localization together with the SMN protein (Grimmler et al. 2005) (Carissimi et al. 2005).

In conclusion, Gemin8 directly binds the Gemin7:6 dimer and, together with unrip, which also binds the Gemin7:6 dimer, forms a hetero-tetrameric subunit within the SMN complex.

5.3 The Gemin7:6 module does not behave as a surrogate Sm-dimer during U snRNP assembly

Although lacking significant sequence similarity to the Sm proteins of U snRNPs, Gemins 6 and 7 display remarkable structural similarity to the canonical Sm-fold domain, the structural hallmark of all Sm proteins (See Section 1.1.2 and Section 4.2.1). Additionally, the dimerization mode within the Gemin7:6 sub-complex is achieved via their Sm-fold domain in a head-to-tail manner, mimicking the Sm-Sm protein mode of interaction (Figure 4.3). Although the relevance of this similarity is unclear, it is tempting to speculate that the Gemin7:6 dimer functions as a Sm protein heterodimer-mimic in assisting the Sm core assembly on U snRNAs (Ma et al. 2005). This proposition was tested by *in vitro* binding assays using the recombinant His_6-GST-*Hs* Gemin8$^{190-230}$:Gemin7:6 module with Sm heteromers, Sm core assembly intermediates and the assembled Sm core (Section 4.4). Interestingly, the Gemin7:6 module does not engage with any of the Sm heterodimers (Sm D1/D2, Sm B/D3 and Sm E/F/G) (Figure 4.13 A). This however contrasts with an earlier report wherein the Gemin7:6 sub-complex could independently engage with individual Sm proteins *in vitro* via their Sm-like folds, albeit with varying affinities (Ma et al. 2005). This variation might be due to the heteromeric status of the Sm dimers masking cognate interaction interfaces for the binding of the dimeric Gemin7:6 module or even the presence of Gemin8. The latter notion might be most probable given that Gemin8 binds to the N-terminus of Gemin7's Sm-fold, probably hindering Sm association for the latter.

Structurally mimicking the Sm-Sm mode of interaction, the assembly chaperone pIcln pre-organizes the Sm proteins in a kinetically trapped assembly intermediate

(6S complex) during the early cytoplasmic phase of the U snRNP assembly pathway (Grimm et al. 2013). Similarly, it was tested whether the Sm-dimer like Gemin7:6 sub-complex could engage with the partially open Sm sub-core (Sm D1/D2/E/F/G) assembly intermediate until the addition of Sm B/D3 for Sm ring closure on the U snRNA is achieved. Refuting this notion, the Gemin7:6 sub-complex failed to associate with the open Sm sub-core under the tested binding conditions (Figure 4.13 B). Probably, a feasible scenario would entail the assistance of additional factors and/or structural re-arrangements within the SMN complex to facilitate and/or disable the Gemin7:6 sub-complex participation during Sm core assembly. However, this scenario is less likely given the mechanistic implications accompanying such a process. Interestingly, it was recently suggested that the narrow Sm sub-core would disable any Sm-dimer to bind, until a cognate RNA engages with the open Sm sub-core and widens the latter for allowing Sm B/D3 binding (Yi et al. 2018). Alternatively, Gemin7:6's structural resemblance to Sm heterodimers might participate during U snRNP maturation in the nuclear Cajal bodies analogous to SMN's Tudor domain (Hebert et al. 2001).

5.4 Conservation of SMN's YG-box domain in humans and nematodes

SMN forms the oligomeric core of the SMN complex contributing to its oligomeric heterogeneity *in vitro* and *in vivo*. SMN's C-terminal YG-box domain harbouring the conserved tyrosine and glycine residues forming the $(YxxG)_3$ motif is primarily responsible for oligomerisation (Martin et al., 2012). In spite of subtle differences in the nematode and human YG box region of SMN (Figure 4.16 A), the novel *Ce* YG box dimer structure depicted in Figure 4.16 reinforces the structural conservation of this self-oligomerization motif along the evolutionary time-line. The MBP-fusion *Ce* YG box protein folds into glycine zipper helical dimer with residues of the $(YxxG)_3$ motif participating directly in the dimer interface, reminiscent of the human and fission yeast YG box structural models (Martin et al. 2012; Gupta et al. 2015).

SMA patient missense mutations mapping to SMN's YG-box have been ascribed to greatly influence the oligomeric status of the SMN protein *in vitro* (Martin et al. 2012; Lorson et al. 1998). Strikingly, all of the amino acids that are target of substitutions in patients with SMA are conserved in the nematode SMN's YG-box region

(with the exception of the conservative *Hs* Gly279 to *Ce* Ala198 variation, see Figure 4.16 A) as well. This is also remarkably reflected in the identical *in vitro* oligomerisation behaviour of the *Hs* YG-box and *Ce* YG-box modules bearing SMA patient mutations (Figure 4.17, Figure 4.18 and Table 4.1). Interestingly, from the alanine scanning mutagenesis of the conserved *Hs* Gly279 residue reported to demonstrating no oligomerisation defect (Martin et al. 2012), thus by extension, the equivalent nematode AIa198 variant residue (to the conserved *Hs* Gly279, see Figure 4.16 A) in *Ce* YG-box WT does not affect the influence *Ce* SMN's oligomerization behaviour or the molecular architecture of the dimeric YG-box model (Figure 4.16 and Figure 4.18).

In conclusion, the structural and biochemical comparison of the nematode and human SMN self-oligomerisation YG-box domain lends justification to the functional conservation of SMN across diverse organisms.

5.5 Influence of SMA patient mutations in Gemin8 incorporation into the SMN complex

The self-interaction scaffold formed by the conserved surface of SMN's C-terminus YG box has been proposed as an interaction site for binding other components within the SMN complex. Gemin8 directly engages with SMN's C-terminus thus incorporating the peripheral Gemin7:6 (and, unrip) into the SMN complex (Figure 1.9) (Otter et al. 2007). Notably, with the recently delineated *S. pombe* SMN complex biochemistry it could demonstrated Gemin8 showed direct association with SMN's C-terminus YG-box domain (Figure 4.22) (also, unpublished data, RG Fischer). With SMA patient mutations clustered in SMN's YG-box domain resulting in a range of oligomeric properties *in vitro*, the ability of these missense mutations and/or resultant oligomeric status of SMN in influencing Gemin8 binding was investigated in this work. Using *in vitro* binding assays (Section 4.5.2), it could be demonstrated that *Ce* Gemin8's N-terminus domain is essential for interacting with *Ce* SMN's YG-box domain, however, specific SMA patient mutations in the YG-box domain impaired this association (Figure 4.19). Strikingly, no distinct feature of SMN's oligomerisation status influenced Gemin8 incorporation. This propelled us to examine the nature of oligomerisation displayed by SMN's YG-box dimers. Recently, it has been postulated that the *S. pombe* SMN's YG-box oligomers do not form a symmetric helical bundle of dimers based on disulphide cross-linking experiments (Gupta et al. 2015). Validating this data, a recent structural model

of a biochemically relevant minimal *S. pombe* SMN (see Figure 4.20 for details) showed that the YG-box dimer forms the fundamental structural unit of SMN with the capacity to self-associate across their N- and C- termini forming higher-order oligomers (Figure 5.4 A, left panel) (unpublished data, RG Fischer). Thus, the YG-box dimer subunit stacks head-to-tail forming an asymmetric bundle of subunits. Consequently, two distinct features of the SMN oligomers can be distinguished: the classical YG-box glycine zipper dimerization interface and the oligomerization interface formed when the YG-box dimers associate (Figure 5.4 A, right panel).

As described earlier (Figure 4.16 A), given the high sequence and structural conservation of SMN's YG-box domain in diverse organisms, the *S. pombe* SMN oligomer structural model was examined for comprehending a plausible Gemin8 binding mechanism to SMN. On this note, the *Ce* SMN S185 residue (in *S. pombe*, A130, see Figure 5.4 B and C) apart from stabilizing the YG-helix participates in the oligomerisation interface, but does not engage with Gemin8 in the presence of the Type II SMA S185P substitution. In stark contrast, the solvent exposed *Ce* SMN H192 residue (in *S. pombe*, Y137, see Figure 5.4 B and C) not involved in neither of the dimeric or oligomeric interface, in the occurrence of a SMA substitution (H192R) is oligomerisation competent but entirely loses its ability to bind Gemin8. On the other hand, SMA substitutions of other YG-box residues (M182R, M182T, Y191C, T193I) participating in either the dimerization and/or oligomerisation interfaces only showed reduced or relatively identical Gemin8 binding like the WT YG-box (Table 5.1). Consequently, it could be assessed that Gemin8's engagement with SMN's YG-box domain is dependent on the accessibility of specific residues within the YG module and, the oligomerisation status of SMN does not directly influence Gemin8 binding (Table 5.1). Iterating this notion, is that for SMA patient substitutions such as H192R and T193I (*Hs* H273R and T274I), it seems unlikely that competent (or moderately competent) oligomerisation *in vivo* is responsible for the SMA phenotype given that these variants demonstrate abrogated or reduced Gemin8 association respectively. Therefore, a potential, albeit speculative, function of the YG-box oligomerisation interface could involve both Gemin8 binding and stabilization within the SMN complex.

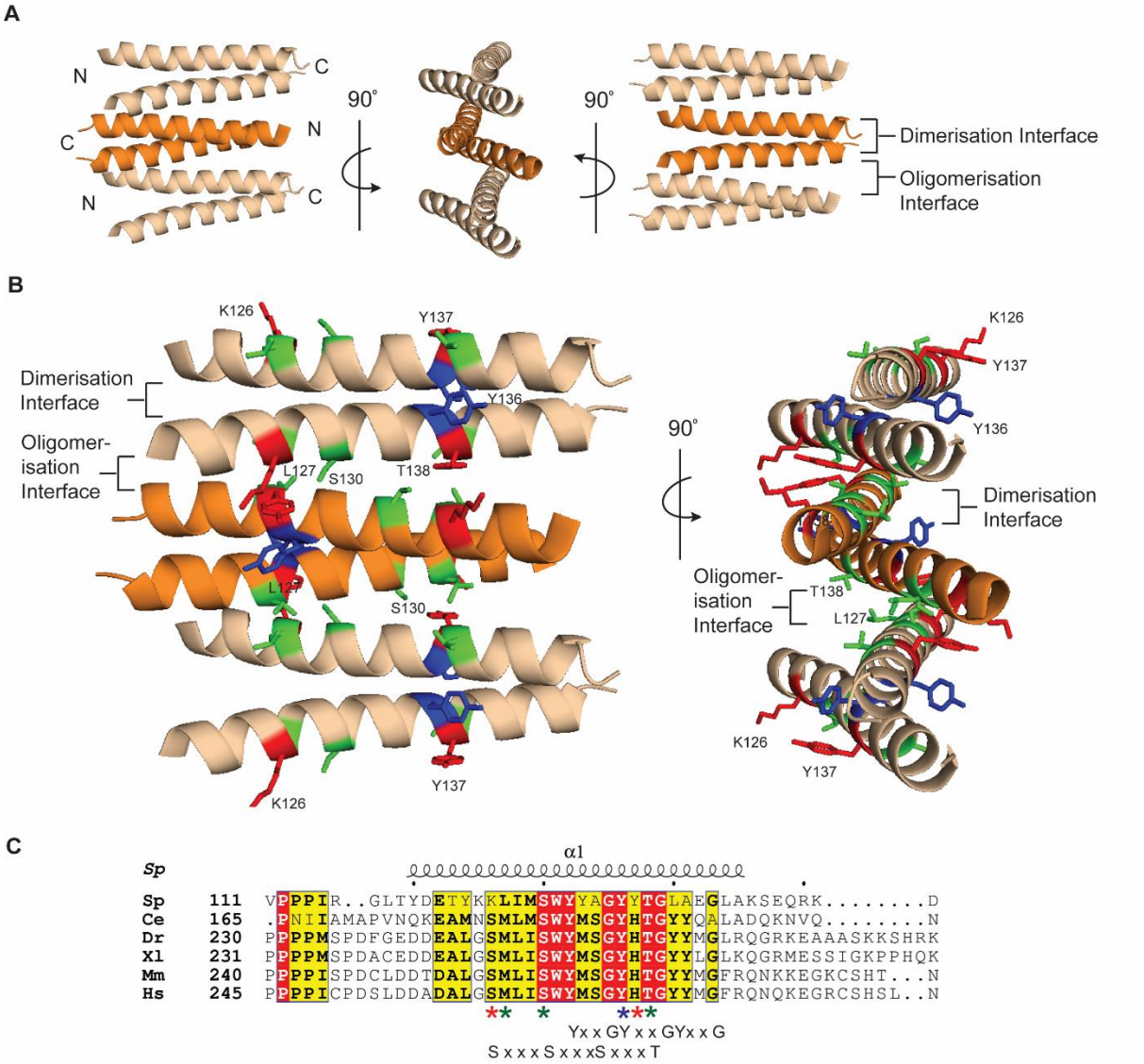


Figure 5.4 SMN Oligomerisation: A structural model

(A) *S. pombe* (*Sp*) SMN's oligomers[++]. Only the YG-box harbouring C-terminus of a minimal *S. pombe* SMN variant (see Figure 4.20) is depicted here. *Left and middle panel*s: The YG-dimer forms the basal structural unit (orange) that assembles in an asymmetric manner (head-to-tail) to form the SMN oligomers (orange and wheat). The N- and C-termini of the YG-dimers protein are indicated. *Right panel*: Two interfaces are thus discerned in this structural model: the dimerization interface involving the classical YxxxYxxxY motif (see panel C) and the oligomerisation interface formed by the stacked YG-dimers.

(B) Dimerization and oligomerisation interfaces in the SMN oligomer. *Left and Right panels*: Only the residues wherein SMA patient missense mutations were studied have been highlighted in the respective interfaces. The dimerization interface includes the conserved tyrosines and glycines of the YxxxYxxxY motif (see Figure 4.14 and Figure 4.16 for details). One such conserved tyrosine residue in this interface, Y136 shown here in blue (Y191 in nematodes) when replaced with a SMA patient mutation shows defective Gemin8 binding. The oligomerisation interface includes the L127 (*Ce* M182), S130 (*Ce* S185) and T138 (*Ce* T193) residues shown in green. The solvent exposed residues that do not participate in either of the interfaces K126 (*Ce* S181) and Y137 (*Ce* H192) are shown in red.*For corresponding amino acid residues in other SMN orthologues, see **(C)** and Table 5.1.

(C) Alignment of SMN's YG-box sequences from diverse organisms. Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. The conserved S, G and Y motifs are indicated. The secondary structure element (α-helix) of the *S. pombe* YG-box from the minimal *S. pombe* SMN variant (see Figure 4.20) is displayed above the sequences as squiggles. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014). The coloured asterisks reflect the residues evaluated for SMA patient substitutions in the context of Gemin8 binding and the colours correspond to the highlighted residues in **(B)**.

Sp: *S pombe*, Ce: *C. elegans*, Dr: *D. rerio,* Xl: *X. laevis*, Mm: *M. musculus*, Hs: *H. sapiens* (also, see Section 7.5).

[++] Unpublished data, RG Fischer.

Table 5.1 SMN's YG-box SMA variants influencing Gemin8 binding

[a]*S. pombe* YG-box residues highlighted in Figure 5.4

[b]Oligomeric status of the MBP-*Ce* YG-box fusions

[c]Interface interactions determined from the *S. pombe* SMN oligomer structural model. Here, only the major roles in contributing to the two interfaces has been taken into account.

[d]Gemin8 binding affinity is based on the quantification of the [^{35}S]-autoradiography signal (Figure 4.19Figure).

[e]SMA severity for each amino acid variant also depends on the *SMN2* gene copy number present in the patient. See Burges and Beattie (2009) for additional information.

[*]Solvent exposed serine residue stabilizes the intra-helical contacts but not directly involved the dimerization interface.

[**]Non-solvent exposed serine residue stabilizes the intra-helical contacts but not directly involved the dimerization interface.

ND: Not defined.

Sp YG-box residues[a]	*Hs* YG-box variants	*Ce* YG-box variants	Motif	Oligomeric Status[b]	Dimerization/ Oligomerisation Interface[c]	Gemin8 binding[d]	SMA Type (*SMN2* copy number)[e]
WT	WT	WT	$(YxxG)_3$ $(SxxxS)_3$	Octamer	Both	++++	Normal
K126	S262G	S181G	$(SxxxS)_3$	Octamer	Neither[*]	++	III (1)
L127	M263R	M182R	Within $(SxxxS)_3$	Tetramer	Oligomeric interface	+++	I (1)
L127	M263T	M182T	Within $(SxxxS)_3$	Tetramer	Oligomeric interface	++++	II (2)
S130	S266P	S185P	$(SxxxS)_3$	Monomer	Oligomeric interface[**]		II (2)
Y136	Y272C	Y191C	$(YxxG)_3$	Monomer	Dimeric interface	++	I (2)
Y137	H273R	H192R	Within $(YxxG)_3$	Octamer	Neither		II (ND)
T138	T274I	T193I	Within $(YxxG)_3$	Monomer-Octamer	Both	++	III (1)

SMA afflicted patients have a homozygous loss of or mutations in the *SMN1* gene locus, but still retaining one or more copies of the near identical homolog, *SMN2*. SMA pathophysiology is therefore due to low levels of functional SMN protein derived from the *SMN2* gene and not its entire absence, consequently resulting in clinical heterogeneity (Lefebvre et al. 1997; Coovert et al. 1997). Notably, no distinct correlation was evident for the currently tested SMA patient substitutions of different clinical SMA forms for Gemin8 binding as well (Table 5.1).

The YG-box domain of *Hs* SMN is coded by the exons 6 and 7 of both the *SMN1* and *SMN2* gene loci (Singh et al. 2017). However the predominant *SMN2* gene product, the truncated SMNΔ7 isoform arising from the exclusion of exon7, includes most of the YG-box residues barring the terminal glycine of the $(YxxG)_3$ motif and the subsequent C-terminal residues (Burghes and Beattie 2009). SMNΔ7, although unstable *in vivo* and fails to compliment SMN in the SMA pathophysiology, partially retains oligomerisation competency and shows Gemin8 association *in vitro* but with abrogated U snRNP assembly activity (Neuenkirchen et al. 2015). Gemin8 association with SMNΔ7 is not therefore surprising since the current data indicates that the inferred Gemin8 binding domain on SMN's YG-box is still retained in the SMNΔ7 truncated isoform.

Apart from the YG-box domain, the conserved regions of homology observed at the functionally relevant N-terminal Gemin2 binding domain and the middle Tudor domain between the nematode and human SMN orthologues is suggestive of an associated functional role conserved throughout evolution (Appendix 7.2). This is indeed corroborated by the co-immunoprecipitations of SMN-Gemin2 (SMI-1, in *C. elegans*) and, SMN RNAi experiments in worms demonstrating embryonic lethality and movement defects similar to SMA phenotypes (Burt, Towers, and Sattelle 2006) (Miguel-Aliaga et al. 1999). Also, the indistinguishable modular composition of the nematode Gemin8:7:6 sub-complex to their human orthologues (Section 4.3.1) further reinforces a structural-functional conservation of the SMN complex across the evolutionary timeline. Therefore, these results although centred on the nematode SMN and Gemin8 orthologues can be extended to their human and fission yeast counterparts as well. Indeed, much like the nematode SMA variant of Y191C demonstrating reduced Gemin8 association (Figure 4.19 B, lane 31), an earlier report assessed that the *Hs* SMN bearing the analogous Y272C SMA mutation also showed defective Gemin8 association (Otter et al. 2007). Likewise, reflecting the human and nematode SMN:

Gemin8 interaction mode, the requirement of *S. pombe* Gemin8's N-terminus for SMN's YG-box domain interaction is apparent (Figure 4.22). In addition, substitution of the A130 residue in *Sp* SMN YG box also abolished Gemin8 binding as evidenced for the SMA substitution of the corresponding residue in the nematode SMN (residue S185) (RG Fischer, unpublished data).

6 Conclusions and Outlook

Despite SMN's oligomerisation behaviour being well-documented, comprehending the influence of SMA patient mutations in ordering the multi-subunit SMN complex architecture has been lacking. In this work, the SMA patient mutations mapping to SMN's YG-box domain dictate Gemin8's incorporation into the SMN complex and, by extension Gemin7:6 sub-complex inclusion. Gemin8, thus moulds the heteromeric SMN complex by bridging the Gemin7:6 (unrip, as well) and SMN:Gemin2 modules (Figure 6.1). Additionally, the biochemical and structural insights gained form employing SMN complex orthologues from divergent organisms (*H. sapiens*, *C. elegans*, *S. pombe*) in this work highlights SMN complex's (and, sub-complexes, thereof) evolutionary conservation in diverse eukaryotes. Nevertheless, the molecular details of the entire SMN complex's architecture still remains elusive. Inherent difficulties in establishing a non-aggregation prone and/or soluble expression of full-length protein modules of the SMN complex inventory from recombinant sources has precluded in successful structural elucidation of the functionally relevant pentameric SMN complex comprising of SMN:Gemin2 with Gemins 6-8. On this note, the recently elucidated *S. pombe* SMN complex (Section 4.5.3 and, unpublished data from RG Fischer) extends a favourable biochemistry enabling prospective structural investigations by Cryo-Electron Microscopy (EM). The present biochemical data and atomic resolution structural models could then easily be integrated into a single Cryo-EM model of the pentameric SMN complex model facilitating a holistic comprehension of the same.

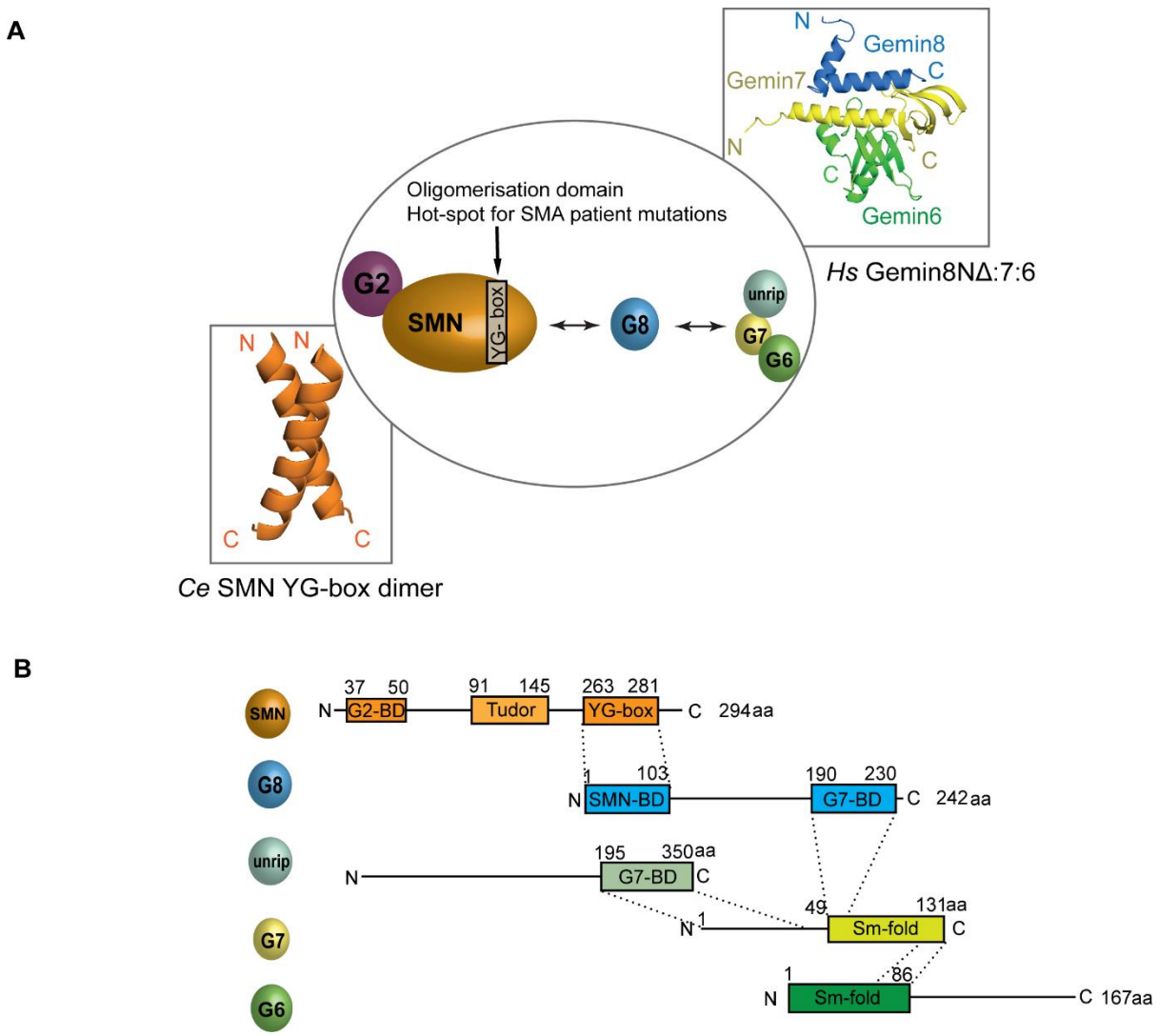


Figure 6.1 Gemin8 sculpts the modular architecture of the multimeric SMN complex
(A) Gemin8 (G8) links the SMN:Gemin2 (G2) and Gemin7(G7):6 (G6) modules building up the heteromeric SMN complex. Gemin8 engages with the Gemin7 via its highly conserved C-terminus (upper right panel). On the other hand, Gemin8 associates with SMN via its N-terminus. SMN's multimerisation YG-box domain, a major hot-spot for SMA patient mutations serves as the binding site for Gemin8. Consequently, SMA patient mutations in the YG-box influence Gemin8 incorporation and, by extension Gemin7:6:unrip association into the SMN complex. The inventory of structural models obtained during this work are also present here. The novel *Hs* Gemin8ΔN7:6 structure (upper right panel) delineates the Gemin8:7 interaction interface. The *Ce* YG-box dimer (lower left panel) adds to the existing inventory of YG-box structural models and reiterates the molecular architecture conservation of this domain along the evolutionary time-line.
(B) Domain mapping of interaction interfaces within the SMN complex. The interaction domains delineated in this study along with earlier reports based on Otter et al., 2000 for SMN, Gemin8, Gemin7, unrip and Gemin6 are depicted here. The interaction interfaces mapping to defined domains are highlighted by the dotted lines. The amino acid residue numbering for the domains are based on the human orthologues.

In light of the current work, it would be interesting to assess whether SMN's YG-box SMA missense mutations affecting Gemin8 binding could alter cellular U snRNP assembly and, subsequently eliciting splicing defects. This speculation lends justification since RNAi mediated Gemin8 depletion or presence of YG-box SMA patient mutations (independently) has been reported to impair SMN complex activity during U snRNP

assembly (Carissimi, Saieva, Baccon, et al. 2006; Pellizzoni, Charroux, and Dreyfuss 1999).

While SMN's oligomerisation status does not distinctly predict Gemin8 association within the SMN complex, relating SMN's oligomerisation to other plausible functional roles *in vivo* is pivotal. The highly conserved SMN' self-interaction YG-box domain presents intriguing possibilities of serving as an interaction platform for other SMN complex components and/or as a signalling hub and/or enabling the assembly of Sm cores on U snRNAs. Indeed, the YG-box module not only serves as an interaction interface for Gemin8, but also for Gemin3, influencing the latter's incorporation into the SMN complex (Charroux et al. 1999). Additionally, SMN's multimerisation was recently shown to dictate the accessibility of a phosphodegron embedded within the YG-box module, thus establishing plausible connections to disease relevant pathways in SMA phenotypes (Gray et al. 2018). However, the intriguing possibility of different oligomeric states of SMN required for the Sm core assembly or in maintaining the integrity of Cajal Bodies given that self-association competency is a shared behaviour of major nuclear body proteins (Hebert and Matera 2000), awaits investigation.

Within the SMN complex inventory, Gemin8 also engages with Gemin4 apart from SMN and Gemin7 (Otter et al. 2007). Gemin4, is believed to serve as a co-factor to the putative RNA helicase Gemin3 of the SMN complex (Charroux et al. 1999; Charroux et al. 2000). Although, Gemins3-4 have been shown to be dispensable for assembly of U snRNPs *in vitro*, their requirement *in vivo* is unclear (Neuenkirchen et al. 2015). Understanding the role of Gemin8 in incorporating Gemin4 into the SMN complex, independent of Gemin3 or otherwise could be a future avenue of study.

7 Appendix

7.1 Unrip engages with Gemin7's N-terminus

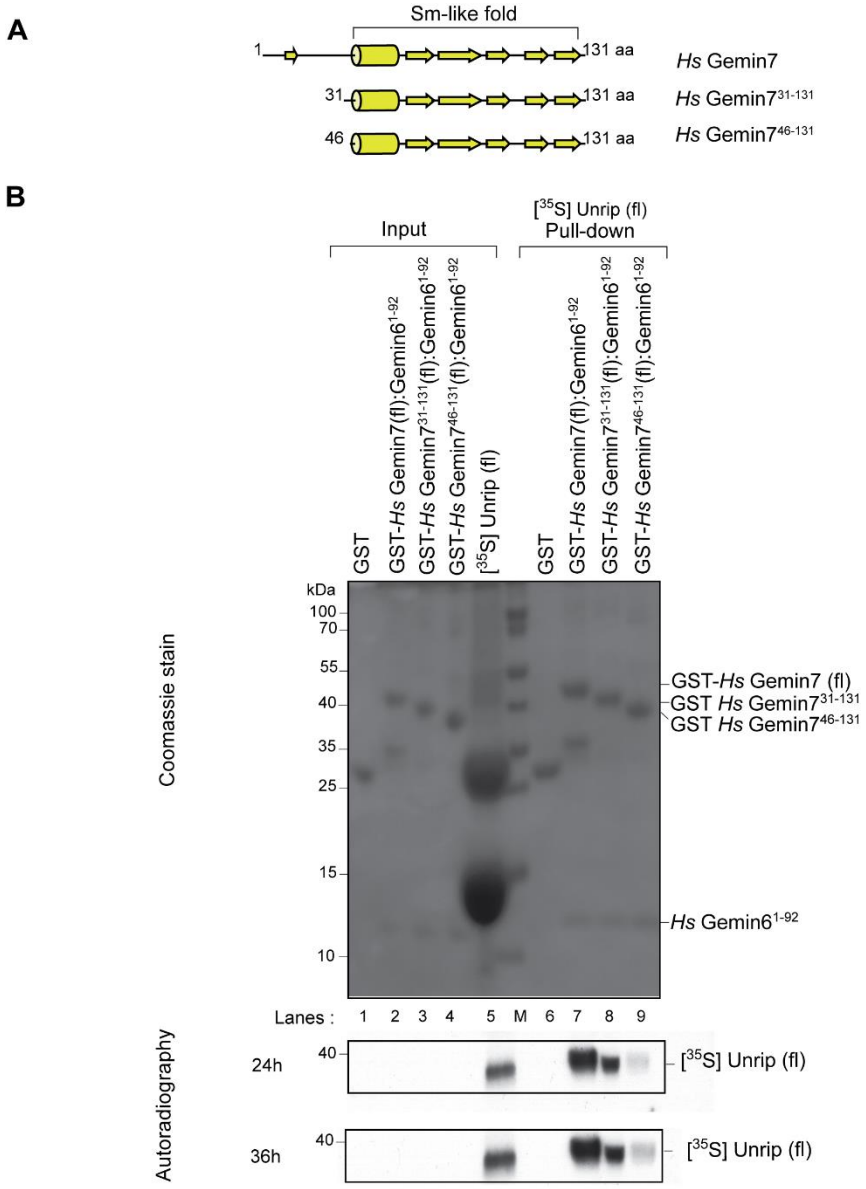


Figure 7.1 Unrip engages with the SMN complex via Gemin7
(A) Schematic representation of *Hs* Gemin7 full-length (fl) and associated truncation mutants secondary structural elements (based on PSIPRED algorithm output). The α-helices, β-sheets and coils are displayed as cylinders, arrows and straight lines respectively.
(B) *In vitro* [^{35}S]-labelled Unrip (fl) (lane 5) was incubated with GSH-immobilized GST-tagged full-length (fl) and truncation variants of Gemin7 in complex with Gemin6$^{1-92}$. After extensive washing, the bound proteins were eluted by boiling the beads in 1X SDS-loading buffer and resolved by SDS-PAGE on a 13.5 % Bis-Tris gel for visualisation by coomassie staining (top panel) and subsequently by autoradiography of the dried gels (middle and lower panels). The dried gels were subjected to 24 h (middle) and 36 h (lower) exposure periods. The total panel shows 5 % of the proteins used for binding. M: PageRuler^{TM} Unstained Protein ladder (200-10 kDa) (Section 3.1.8.4.2).
In contrast to the dimeric Gemin7 (fl): Gemin6$^{1-92}$ complex, Gemin7$^{31-131}$ (N-terminal deletion variant) partially loses unrip binding, while complete removal of the largely unstructured N-terminus of *Hs* Gemin7 abrogated unrip binding. In conjecture with the results from Section 4.3.2, Unrip's binding site on Gemin7 precedes the Sm fold of Gemin7 to which Gemin8 binds. Thus, Gemin7's N-terminus includes exclusive binding sites for Gemin8 and unrip (Figure 6.1 B).

7.2 Sequence and structural alignment of SMN orthologues

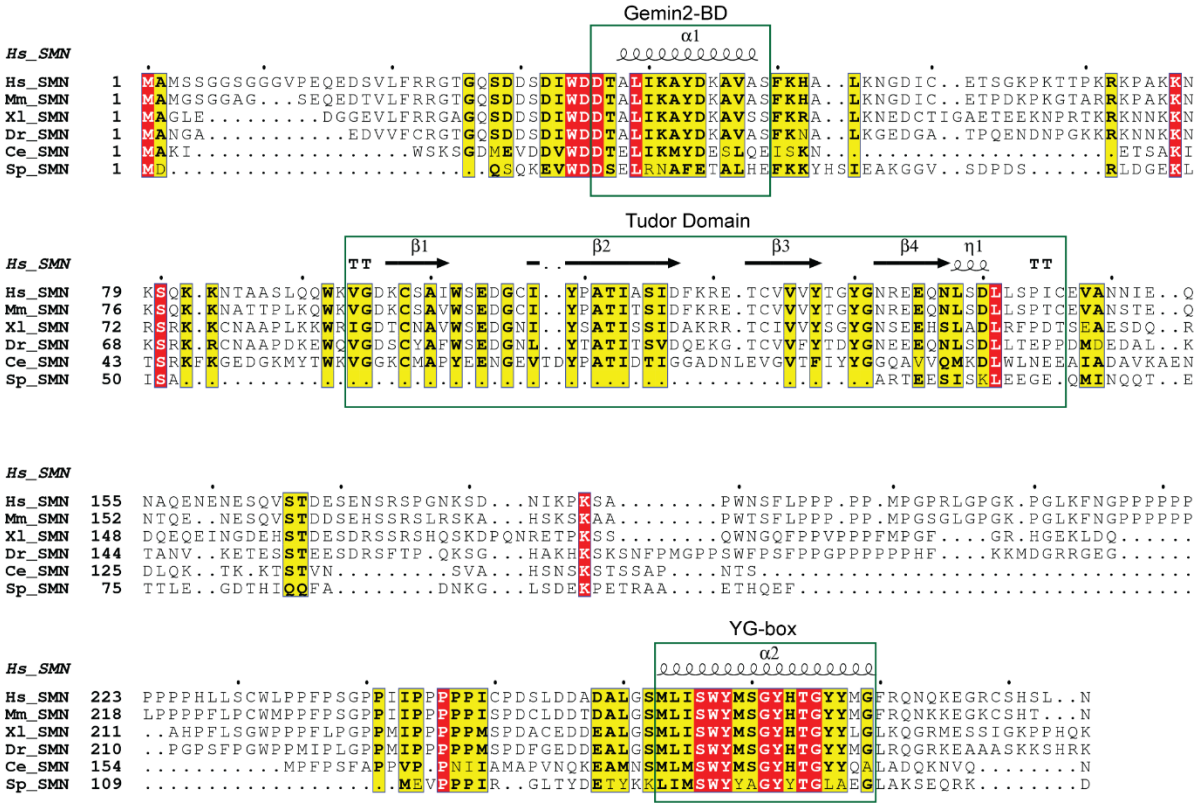


Figure 7.2 Conservation of domains in SMN orthologues
Alignment of SMN orthologues from diverse organisms. The secondary structure elements of the *H. sapiens* SMN protein domains is displayed above the sequences. α-helices and 3_{10}-helices (η) are displayed as squiggles, β-strands are rendered as arrows, strict β-turns as TT letters and strict α-turns as TTT. Domain organisation of the SMN protein includes the highly conserved Gemin2 binding domain (BD) N-terminus helix (α1), a central Tudor domain includes a beta-barrel comprised of five stranded anti-parallel beta sheets (β1 - β5) which is lacking in the *S. pombe* (Sp) SMN orthologue and the C-terminus multimerisation YG-box domain shared across divergent organisms.
Identical and similar residues shared by the orthologues are shown as a red box with white character and a yellow box with black character respectively. This alignment was computed using T-Coffee (Notredame et al., 2000) and visualised using ESPript 3.0 (Robert and Gouet, 2014).
Hs: *H. sapiens,* Mm: *M. musculus,* Xl: *X. laevis,* Dr: *D. rerio,* Sp: *S pombe,* Ce: *C. elegans* (see Section 7.5 for UniProtKB codes).

7.3 Vectors pETM41 and pETM41* variations

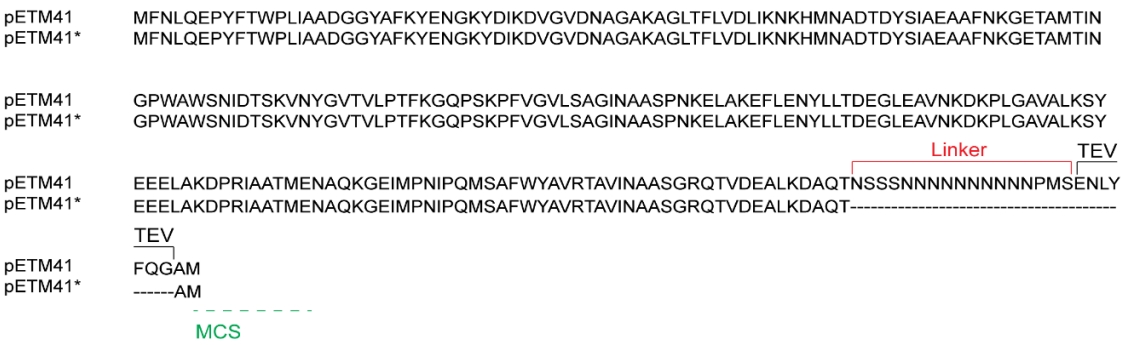


Figure 7.3 Alignment of pETM41 vector variants
The original pETM41 (Section 3.1.5) was modified by Overlap extension PCR (Section 3.2.1.5) to generate the modified pETM41*vector lacking the linker (red) and TEV site (black) preceding the *Nco*I (the terminal Methionine residue depicted here) site of the MCS (green, dashed lines). This modified vector enabled generating SMN YG-box variants wherein the oligomeric status could be controlled (see Section 4.5.1.1 for details).

7.4 PDB codes of atomic resolution structures used for scientific illustrations in this work

PDB Code	Protein Domain/Biomolecule/Complex	Reference
3PGW	U1 snRNP	Weber et al., 2010
4WZJ	U4 snRNP	Leung et al., 2011
4F7U	6S complex	Grimm et al., 2013
5XJL	7S complex	Zhang et al., 2012
4V98	8S complex	Grimm et al., 2013
1Y96	Human Gemin7:6	Ma et al., 2005
1MHN	Human SMN Tudor domain	Sprangers et al., 2003
4GLI	Human SMN YG-box domain	Martin et al., 2012
3B7G	Human Gemin3	Schutz et al., 2010
5HIJ	Human Gemin5	Jin et al., 2016
4RG5	Fission yeast SMN YG-box domain	Gupta et al., 2015

7.5 SMN orthologues UniProtKB accession numbers

*Only the proteins and their orthologues employed in this work have been included here.

Protein*	SMN	Gemin8	Gemin7	Gemin6
Species				
Homo sapiens	Q16637	Q9NWZ8	Q9H840	Q8WXD5
Mus musculus	P97801	Q8BHE1	Q9CWY4	Q9CX53
Xenopus laevis	Q6NU78	A1L2Q9	A0A1L8F860	A0A1L8GIXF
Danio rerio	Q9W6S8	FIQAZ5	F1R879	A0A140LG88
*Drosophila Melanogaster***	Q9VV74	-	-	-
Caenorhabditis elegans	G5EC16	P343999	P343999	O01871
*Schizosaccharomyces pombe****	O42661	N/A	G2TRR1	N/A

** No orthologues for Gemins8, 7 and 6 have yet been annotated or identified in this species.
*** The accession numbers for the *S. Pombe* Gemin8 and Gemin6 orthologues are not available and the sequences were obtained from RG Fischer, University of Würzburg (Germany)

7.6 Symbols, Units, Acronyms and Abbreviations

Symbol	***Definition***
α	Alpha
β	Beta
Δ	Delta (for protein deletion variants)
µ	micro
M_r	Relative Molecular weight
V_E	SEC elution volume
V_o	SEC Void Volume
Units	
Å	Angstrom
bp	Base pairs
°C	Degree Celsius
h	hours
Da/ kDa	Dalton/ Kilo Dalton
min	Minute
M	Molar
L / mL / µL	Litre / Millilitre / Microlitre
M / mM / µM	Molar / Millimolar / Micromolar
s	Seconds
%	Percent
rpm	Revolutions per minute
v/v	volume-volume percentage
w/v	Weight-volume percentage
Acronyms and Abbreviations	***Definition/Expansion***
aa	Amino acid
APS	Ammonium persulfate
Bis-Tris	2-bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol
BSA	Bovine serum albumin
cDNA	Complementary DNA
C-terminal	Carboxyl terminal
Da	Dalton
ddH2O	Double-distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside-5‘-triphosphosphate

ds	Double-stranded
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
fl	Full-length
GDP	Guanosine-5‘-diphosphate
GSH	Glutathione sepharose
GST	Glutathione-S-transferase
GTP	Guanosine-5‘-triphosphate
GOI	Gene of interest
HCl	Hydrochloric acid
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethane sulfonic acid
His_6	His6 Hexahistidine tag
IMAC	IMAC
IP	Immunoprecipitation
IPTG	Isopropyl-s-D-thiogalactoside
LB	Luria Bertani
m_3G	2,2,7 - trimethylguanosine
m_7G	7 - monomethylguanosine
MCS	Multiple cloning site
MEP50	Methylosome protein 50 (= WD45, WDR77)
MES	2-(N-morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
mRNA	Messenger RNA
NaCl	Sodium chloride
NaOH	Sodium hydroxide
N/A	Not Applicable or Not Available
Ni-NTA	Nickel-nitrilotriacetic acid
N-terminal	Amino terminal
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pICln	Integral component of a nucleotide-sensitive chloride channel
PSc	Prescission protease

RanGDP	GDP-bound nuclear protein Ran
RanGTP	GTP-bound nuclear protein Ran
RE	Restriction enzyme
RG	Research group
RMSD	Root mean square deviation
RNA Ribonucleic acid	Ribonucleic acid
RNP	Ribonucleoprotein particle (RNA-protein particle)
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
Sm	Smith
SMA	Spinal muscular atrophy
SMN	Survival motor neuron
TAE	Tris acetate EDTA
TB	Terrific Broth
TBE	Tris borate EDTA
TE	TE Tris/EDTA
TEMED	N,N,N',N',-tetramethylethylenediamine
TEV	Tobacco Etch Virus
U snRNA	Uridine rich small nuclear ribonucleic acid
U snRNP	Uridine rich small nuclear ribonucleoprotein particle
WD45	WD-repeat domain 45 (= MEP50, WDR77)
WT	Wild-type

7.7 Nucleotide bases and Amino acid codes

Nucleotide bases

A	Adenine
C	Cytosine
G	Guanine
T	Thymine (DNA only)
U	Uracil (RNA only)

Amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Iso	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

7.8 Individual contributions of personnel towards the structural work presented in this dissertation

Protein / Sub-complex	Recombinant protein production	Crystallogenesis	Data Collection*	Structure determination and refinement	Structural Analysis**	Corresponding Figure/Section in the dissertation
Hs Gemin7$^{31\text{-}131}$:6 (fl)	AV	AV	CG (75 %) AV (25 %)	CG	AV	Figure 4.3
Hs Gemin8$^{190\text{-}230}$:7$^{46\text{-}131}$:6$^{1\text{-}92}$	AV	AV	CG (75 %) AV (25 %)	CG	AV	Figures 4.8 and 4.9
Sp Gemin8ΔN30ΔL:7:6	AV	AV	CG (100 %)	N/A (AV***)	N/A	Section 4.3.4
Ce MBP*-YG$^{182\text{-}207}$	AV	AV	CG (95 %) AV (5 %)	CG	AV	Figure 4.16

AV: Aravindan Viswanathan
CG: Dr. Clemens Grimm (RG Fischer)
*Data Collection includes crystal picking, freezing and participation during synchrotron beam time.
**With input from CG on analysis.
***Analysis of data set using BALBES (refer to Section 4.3.4 for details)

8 References

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm, and R. Luhrmann. 1999. 'A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro', *EMBO J*, 18: 5789-802.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2010. 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr*, 66: 213-21.

Antonysamy, S., Z. Bonday, R. M. Campbell, B. Doyle, Z. Druzina, T. Gheyi, B. Han, L. N. Jungheim, Y. Qian, C. Rauch, M. Russell, J. M. Sauder, S. R. Wasserman, K. Weichert, F. S. Willard, A. Zhang, and S. Emtage. 2012. 'Crystal structure of the human PRMT5:MEP50 complex', *Proc Natl Acad Sci U S A*, 109: 17960-5.

Baillat, D., M. A. Hakimi, A. M. Naar, A. Shilatifard, N. Cooch, and R. Shiekhattar. 2005. 'Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II', *Cell*, 123: 265-76.

Battle, D. J., M. Kasim, J. Wang, and G. Dreyfuss. 2007. 'SMN-independent subunits of the SMN complex. Identification of a small nuclear ribonucleoprotein assembly intermediate', *J Biol Chem*, 282: 27953-9.

Berg, O. G., and P. H. von Hippel. 1985. 'Diffusion-controlled macromolecular interactions', *Annu Rev Biophys Biophys Chem*, 14: 131-60.

Borg, R., and R. J. Cauchi. 2013. 'The Gemin associates of survival motor neuron are required for motor function in Drosophila', *PLoS One*, 8: e83878.

Bradford, M. M. 1976. 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Anal Biochem*, 72: 248-54.

Brahms, H., L. Meheus, V. de Brabandere, U. Fischer, and R. Luhrmann. 2001. 'Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein', *RNA*, 7: 1531-42.

Buhler, D., V. Raker, R. Luhrmann, and U. Fischer. 1999. 'Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy', *Hum Mol Genet*, 8: 2351-7.

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

Burt, E. C., P. R. Towers, and D. B. Sattelle. 2006. 'Caenorhabditis elegans in the study of SMN-interacting proteins: a role for SMI-1, an orthologue of human Gemin2 and the identification of novel components of the SMN complex', *Invert Neurosci*, 6: 145-59.

Camasses, A., E. Bragado-Nilsson, R. Martin, B. Seraphin, and R. Bordonne. 1998. 'Interactions within the yeast Sm core complex: from proteins to amino acids', *Mol Cell Biol*, 18: 1956-66.

Carissimi, C., J. Baccon, M. Straccia, P. Chiarella, A. Maiolica, A. Sawyer, J. Rappsilber, and L. Pellizzoni. 2005. 'Unrip is a component of SMN complexes active in snRNP assembly', *FEBS Lett*, 579: 2348-54.

Carissimi, C., L. Saieva, J. Baccon, P. Chiarella, A. Maiolica, A. Sawyer, J. Rappsilber, and L. Pellizzoni. 2006. 'Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly', *J Biol Chem*, 281: 8126-34.

Carissimi, C., L. Saieva, F. Gabanella, and L. Pellizzoni. 2006. 'Gemin8 is required for the architecture and function of the survival motor neuron complex', *J Biol Chem*, 281: 37009-16.

Carrillo Oesterreich, F., N. Bieberstein, and K. M. Neugebauer. 2011. 'Pause locally, splice globally', *Trends Cell Biol*, 21: 328-35.

Chari, A., and U. Fischer. 2010. 'Cellular strategies for the assembly of molecular machines', *Trends Biochem Sci*, 35: 676-83.

Chari, A., M. M. Golas, M. Klingenhager, N. Neuenkirchen, B. Sander, C. Englbrecht, A. Sickmann, H. Stark, and U. Fischer. 2008. 'An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs', *Cell*, 135: 497-509.

Charroux, B., L. Pellizzoni, R. A. Perkinson, A. Shevchenko, M. Mann, and G. Dreyfuss. 1999. 'Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems', *J Cell Biol*, 147: 1181-94.

Charroux, B., L. Pellizzoni, R. A. Perkinson, J. Yong, A. Shevchenko, M. Mann, and G. Dreyfuss. 2000. 'Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli', *J Cell Biol*, 148: 1177-86.

Coady, T. H., and C. L. Lorson. 2011. 'SMN in spinal muscular atrophy and snRNP biogenesis', *Wiley Interdiscip Rev RNA*, 2: 546-64.

Coovert, D. D., T. T. Le, P. E. McAndrew, J. Strasswimmer, T. O. Crawford, J. R. Mendell, S. E. Coulson, E. J. Androphy, T. W. Prior, and A. H. Burghes. 1997. 'The survival motor neuron protein in spinal muscular atrophy', *Hum Mol Genet*, 6: 1205-14.

Cordin, O., D. Hahn, and J. D. Beggs. 2012. 'Structure, function and regulation of spliceosomal RNA helicases', *Curr Opin Cell Biol*, 24: 431-8.

Cusco, I., M. J. Barcelo, E. del Rio, M. Baiget, and E. F. Tizzano. 2004. 'Detection of novel mutations in the SMN Tudor domain in type I SMA patients', *Neurology*, 63: 146-9.

Egloff, S., D. O'Reilly, and S. Murphy. 2008. 'Expression of human snRNA genes from beginning to end', *Biochem Soc Trans*, 36: 590-4.

Ellis, R. J., and A. P. Minton. 2006. 'Protein aggregation in crowded environments', *Biol Chem*, 387: 485-97.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, and J. A. Piccirilli. 2013. 'RNA catalyses nuclear pre-mRNA splicing', *Nature*, 503: 229-34.

Fischer, U., J. Heinrich, K. van Zee, E. Fanning, and R. Luhrmann. 1994. 'Nuclear transport of U1 snRNP in somatic cells: differences in signal requirement compared with Xenopus laevis oocytes', *J Cell Biol*, 125: 971-80.

Fornerod, M., M. Ohno, M. Yoshida, and I. W. Mattaj. 1997. 'CRM1 is an export receptor for leucine-rich nuclear export signals', *Cell*, 90: 1051-60.

Friesen, W. J., and G. Dreyfuss. 2000. 'Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN)', *J Biol Chem*, 275: 26370-5.

Friesen, W. J., S. Massenet, S. Paushkin, A. Wyce, and G. Dreyfuss. 2001. 'SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets', *Mol Cell*, 7: 1111-7.

Friesen, W. J., S. Paushkin, A. Wyce, S. Massenet, G. S. Pesiridis, G. Van Duyne, J. Rappsilber, M. Mann, and G. Dreyfuss. 2001. 'The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins', *Mol Cell Biol*, 21: 8289-300.

Friesen, W. J., A. Wyce, S. Paushkin, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss. 2002. 'A novel WD repeat protein component of the methylosome binds Sm proteins', *J Biol Chem*, 277: 8243-7.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch. 2003. 'ExPASy: The proteomics server for in-depth protein knowledge and analysis', *Nucleic Acids Res*, 31: 3784-8.

Gavin, A. C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga. 2006. 'Proteome survey reveals modularity of the yeast cell machinery', *Nature*, 440: 631-6.

Gilbert, W. 1978. 'Why genes in pieces?', *Nature*, 271: 501.

Gray, K. M., K. A. Kaifer, D. Baillat, Y. Wen, T. R. Bonacci, A. D. Ebert, A. C. Raimer, A. M. Spring, S. T. Have, J. J. Glascock, K. Gupta, G. D. Van Duyne, M. J. Emanuele, A. I. Lamond, E. J. Wagner, C. L. Lorson, and A. G. Matera. 2018. 'Self-oligomerization regulates stability of survival motor neuron protein isoforms by sequestering an SCF(Slmb) degron', *Mol Biol Cell*, 29: 96-110.

Grimm, C., A. Chari, J. P. Pelz, J. Kuper, C. Kisker, K. Diederichs, H. Stark, H. Schindelin, and U. Fischer. 2013. 'Structural basis of assembly chaperone- mediated snRNP formation', *Mol Cell*, 49: 692-703.

Grimm, C., A. Chari, K. Reuter, and U. Fischer. 2010. 'A crystallization screen based on alternative polymeric precipitants', *Acta Crystallogr D Biol Crystallogr*, 66: 685-97.

Grimmler, M., S. Otter, C. Peter, F. Muller, A. Chari, and U. Fischer. 2005. 'Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization', *Hum Mol Genet*, 14: 3099-111.

Gruss, O. J., R. Meduri, M. Schilling, and U. Fischer. 2017. 'UsnRNP biogenesis: mechanisms and regulation', *Chromosoma*, 126: 577-93.

Gupta, K., R. Martin, R. Sharp, K. L. Sarachan, N. S. Ninan, and G. D. Van Duyne. 2015. 'Oligomeric Properties of Survival Motor Neuron.Gemin2 Complexes', *J Biol Chem*, 290: 20185-99.

Hallais, M., F. Pontvianne, P. R. Andersen, M. Clerici, D. Lener, H. Benbahouche Nel, T. Gostan, F. Vandermoere, M. C. Robert, S. Cusack, C. Verheggen, T. H. Jensen, and E. Bertrand. 2013. 'CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing', *Nat Struct Mol Biol*, 20: 1358-66.

Hamm, J., E. Darzynkiewicz, S. M. Tahara, and I. W. Mattaj. 1990. 'The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal', *Cell*, 62: 569-77.

Hannus, S., D. Buhler, M. Romano, B. Seraphin, and U. Fischer. 2000. 'The Schizosaccharomyces pombe protein Yab8p and a novel factor, Yip1p, share structural and functional similarity with the spinal muscular atrophy-associated proteins SMN and SIP1', *Hum Mol Genet*, 9: 663-74.

Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. 'From molecular to modular cell biology', *Nature*, 402: C47-52.

Hebert, M. D., and A. G. Matera. 2000. 'Self-association of coilin reveals a common theme in nuclear body localization', *Mol Biol Cell*, 11: 4159-71.

Hebert, M. D., P. W. Szymczyk, K. B. Shpargel, and A. G. Matera. 2001. 'Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein', *Genes Dev*, 15: 2720-9.

Hermann, H., P. Fabrizio, V. A. Raker, K. Foulaki, H. Hornig, H. Brahms, and R. Luhrmann. 1995. 'snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions', *EMBO J*, 14: 2076-88.

Huber, J., U. Cronshagen, M. Kadokura, C. Marshallsay, T. Wada, M. Sekine, and R. Luhrmann. 1998. 'Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure', *EMBO J*, 17: 4114-26.

Hunt, S. L., J. J. Hsuan, N. Totty, and R. J. Jackson. 1999. 'unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA', *Genes Dev*, 13: 437-48.

Izaurralde, E., J. Lewis, C. Gamberi, A. Jarmolowski, C. McGuigan, and I. W. Mattaj. 1995. 'A cap-binding protein complex mediating U snRNA export', *Nature*, 376: 709-12.

Jablonka, S., B. Holtmann, G. Meister, M. Bandilla, W. Rossoll, U. Fischer, and M. Sendtner. 2002. 'Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death', *Proc Natl Acad Sci U S A*, 99: 10126-31.

Jin, W., Y. Wang, C. P. Liu, N. Yang, M. Jin, Y. Cong, M. Wang, and R. M. Xu. 2016. 'Structural basis for snRNA recognition by the double-WD40 repeat domain of Gemin5', *Genes Dev*, 30: 2391-403.

Jones, D. T. 1999. 'Protein secondary structure prediction based on position-specific scoring matrices', *J Mol Biol*, 292: 195-202.
Kabsch, W. 2010. 'Xds', *Acta Crystallogr D Biol Crystallogr*, 66: 125-32.
Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li, and K. Nagai. 1999. 'Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs', *Cell*, 96: 375-87.
Karijolich, J., and Y. T. Yu. 2010. 'Spliceosomal snRNA modifications and their function', *RNA Biol*, 7: 192-204.
Kashima, T., and J. L. Manley. 2003. 'A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy', *Nat Genet*, 34: 460-3.
Kroiss, M., J. Schultz, J. Wiesner, A. Chari, A. Sickmann, and U. Fischer. 2008. 'Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster', *Proc Natl Acad Sci U S A*, 105: 10045-50.
Kufareva, I., and R. Abagyan. 2012. 'Methods of protein structure comparison', *Methods Mol Biol*, 857: 231-57.
Lefebvre, S., L. Burglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani, and et al. 1995. 'Identification and characterization of a spinal muscular atrophy-determining gene', *Cell*, 80: 155-65.
Lefebvre, S., P. Burlet, Q. Liu, S. Bertrandy, O. Clermont, A. Munnich, G. Dreyfuss, and J. Melki. 1997. 'Correlation between severity and SMN protein level in spinal muscular atrophy', *Nature Genetics*, 16: 265-69.
Lerner, M. R., and J. A. Steitz. 1979. 'Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus', *Proc Natl Acad Sci U S A*, 76: 5495-9.
Leung, A. K., K. Nagai, and J. Li. 2011. 'Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis', *Nature*, 473: 536-9.
Liu, Q., U. Fischer, F. Wang, and G. Dreyfuss. 1997. 'The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins', *Cell*, 90: 1013-21.
Long, F., A. A. Vagin, P. Young, and G. N. Murshudov. 2008. 'BALBES: a molecular-replacement pipeline', *Acta Crystallogr D Biol Crystallogr*, 64: 125-32.
Lorson, C. L., and E. J. Androphy. 1998. 'The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding', *Hum Mol Genet*, 7: 1269-75.
Lorson, C. L., J. Strasswimmer, J. M. Yao, J. D. Baleja, E. Hahnen, B. Wirth, T. Le, A. H. Burghes, and E. J. Androphy. 1998. 'SMN oligomerization defect correlates with spinal muscular atrophy severity', *Nat Genet*, 19: 63-6.
Ma, Y., J. Dostie, G. Dreyfuss, and G. D. Van Duyne. 2005. 'The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure', *Structure*, 13: 883-92.
Martin, R., K. Gupta, N. S. Ninan, K. Perry, and G. D. Van Duyne. 2012. 'The survival motor neuron protein forms soluble glycine zipper oligomers', *Structure*, 20: 1929-39.
Massenet, S., L. Pellizzoni, S. Paushkin, I. W. Mattaj, and G. Dreyfuss. 2002. 'The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway', *Mol Cell Biol*, 22: 6533-41.
Matera, A. G., and Z. Wang. 2014. 'A day in the life of the spliceosome', *Nat Rev Mol Cell Biol*, 15: 108-21.
Mattaj, I. W., and R. Zeller. 1983. 'Xenopus laevis U2 snRNA genes: tandemly repeated transcription units sharing 5' and 3' flanking homology with other RNA polymerase II transcribed genes', *EMBO J*, 2: 1883-91.
McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. 'Phaser crystallographic software', *J Appl Crystallogr*, 40: 658-74.
Meier, I. D., M. P. Walker, and A. G. Matera. 2018. 'Gemin4 is an essential gene in mice, and its overexpression in human cells causes relocalization of the SMN complex to the nucleoplasm', *Biol Open*, 7.

Meister, G., C. Eggert, D. Buhler, H. Brahms, C. Kambach, and U. Fischer. 2001. 'Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln', *Curr Biol*, 11: 1990-4.

Meister, G., and U. Fischer. 2002. 'Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs', *EMBO J*, 21: 5853-63.

Miguel-Aliaga, I., Y. B. Chan, K. E. Davies, and M. van den Heuvel. 2000. 'Disruption of SMN function by ectopic expression of the human SMN gene in Drosophila', *FEBS Lett*, 486: 99-102.

Miguel-Aliaga, I., E. Culetto, D. S. Walker, H. A. Baylis, D. B. Sattelle, and K. E. Davies. 1999. 'The Caenorhabditis elegans orthologue of the human gene responsible for spinal muscular atrophy is a maternal product critical for germline maturation and embryonic viability', *Hum Mol Genet*, 8: 2133-43.

Mitchell, S. A., K. A. Spriggs, M. J. Coldwell, R. J. Jackson, and A. E. Willis. 2003. 'The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr', *Mol Cell*, 11: 757-71.

Mowry, K. L., and J. A. Steitz. 1987. 'Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's', *Science*, 238: 1682-7.

Neuenkirchen, N., C. Englbrecht, J. Ohmer, T. Ziegenhals, A. Chari, and U. Fischer. 2015. 'Reconstitution of the human U snRNP assembly machinery reveals stepwise Sm protein organization', *EMBO J*, 34: 1925-41.

Noble, S. M., and C. Guthrie. 1996. 'Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations', *Genetics*, 143: 67-80.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. 'T-Coffee: A novel method for fast and accurate multiple sequence alignment', *J Mol Biol*, 302: 205-17.

Ogawa, C., K. Usui, M. Aoki, F. Ito, M. Itoh, C. Kai, M. Kanamori-Katayama, Y. Hayashizaki, and H. Suzuki. 2007. 'Gemin2 plays an important role in stabilizing the survival of motor neuron complex', *J Biol Chem*, 282: 11122-34.

Ogawa, C., K. Usui, F. Ito, M. Itoh, Y. Hayashizaki, and H. Suzuki. 2009. 'Role of survival motor neuron complex components in small nuclear ribonucleoprotein assembly', *J Biol Chem*, 284: 14609-17.

Ohno, M., A. Segref, A. Bachi, M. Wilm, and I. W. Mattaj. 2000. 'PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation', *Cell*, 101: 187-98.

Otter, S., M. Grimmler, N. Neuenkirchen, A. Chari, A. Sickmann, and U. Fischer. 2007. 'A comprehensive interaction map of the human survival of motor neuron (SMN) complex', *J Biol Chem*, 282: 5825-33.

Paknia, E., A. Chari, H. Stark, and U. Fischer. 2016. 'The Ribosome Cooperates with the Assembly Chaperone pICln to Initiate Formation of snRNPs', *Cell Rep*, 16: 3103-12.

Palacios, I., M. Hetzer, S. A. Adam, and I. W. Mattaj. 1997. 'Nuclear import of U snRNPs requires importin beta', *EMBO J*, 16: 6783-92.

Pannone, B. K., S. D. Kim, D. A. Noe, and S. L. Wolin. 2001. 'Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein', *Genetics*, 158: 187-96.

Papasaikas, P., and J. Valcarcel. 2016. 'The Spliceosome: The Ultimate RNA Chaperone and Sculptor', *Trends Biochem Sci*, 41: 33-45.

Patel, A. A., and J. A. Steitz. 2003. 'Splicing double: insights from the second spliceosome', *Nat Rev Mol Cell Biol*, 4: 960-70.

Paushkin, S., A. K. Gubitz, S. Massenet, and G. Dreyfuss. 2002. 'The SMN complex, an assemblyosome of ribonucleoproteins', *Curr Opin Cell Biol*, 14: 305-12.

Pellizzoni, L., B. Charroux, and G. Dreyfuss. 1999. 'SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins', *Proc Natl Acad Sci U S A*, 96: 11167-72.

Pillai, R. S., M. Grimmler, G. Meister, C. L. Will, R. Luhrmann, U. Fischer, and D. Schumperli. 2003. 'Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN

complex and the role of a new component, Lsm11, in histone RNA processing', *Genes Dev*, 17: 2321-33.

Plessel, G., U. Fischer, and R. Luhrmann. 1994. 'm3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain', *Mol Cell Biol*, 14: 4160-72.

Plessel, G., R. Luhrmann, and B. Kastner. 1997. 'Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (E.F.G) protein heteromer and the intact snRNP core', *J Mol Biol*, 265: 87-94.

Pomeranz Krummel, D. A., C. Oubridge, A. K. Leung, J. Li, and K. Nagai. 2009. 'Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution', *Nature*, 458: 475-80.

Prusty, A. B., R. Meduri, B. K. Prusty, J. Vanselow, A. Schlosser, and U. Fischer. 2017. 'Impaired spliceosomal UsnRNP assembly leads to Sm mRNA down-regulation and Sm protein degradation', *J Cell Biol*, 216: 2391-407.

Raker, V. A., K. Hartmuth, B. Kastner, and R. Luhrmann. 1999. 'Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner', *Mol Cell Biol*, 19: 6554-65.

Raker, V. A., G. Plessel, and R. Luhrmann. 1996. 'The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro', *EMBO J*, 15: 2256-69.

Reddy, R., D. Henning, G. Das, M. Harless, and D. Wright. 1987. 'The capped U6 small nuclear RNA is transcribed by RNA polymerase III', *J Biol Chem*, 262: 75-81.

Rino, J., and M. Carmo-Fonseca. 2009. 'The spliceosome: a self-organized macromolecular machine in the nucleus?', *Trends Cell Biol*, 19: 375-84.

Robert, X., and P. Gouet. 2014. 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Res*, 42: W320-4.

Sarachan, K. L., K. G. Valentine, K. Gupta, V. R. Moorman, J. M. Gledhill, Jr., M. Bernens, C. Tommos, A. J. Wand, and G. D. Van Duyne. 2012. 'Solution structure of the core SMN-Gemin2 complex', *Biochem J*, 445: 361-70.

Schutz, P., T. Karlberg, S. van den Berg, R. Collins, L. Lehtio, M. Hogbom, L. Holmberg-Schiavone, W. Tempel, H. W. Park, M. Hammarstrom, M. Moche, A. G. Thorsell, and H. Schuler. 2010. 'Comparative structural analysis of human DEAD-box RNA helicases', *PLoS One*, 5.

Selenko, P., R. Sprangers, G. Stier, D. Buhler, U. Fischer, and M. Sattler. 2001. 'SMN tudor domain structure and its interaction with the Sm proteins', *Nat Struct Biol*, 8: 27-31.

Seraphin, B. 1995. 'Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs', *EMBO J*, 14: 2089-98.

Sharma, A., A. Lambrechts, T. Hao le, T. T. Le, C. A. Sewry, C. Ampe, A. H. Burghes, and G. E. Morris. 2005. 'A role for complexes of survival of motor neurons (SMN) protein with gemins and profilin in neurite-like cytoplasmic extensions of cultured nerve cells', *Exp Cell Res*, 309: 185-97.

Sharp, P. A., and C. B. Burge. 1997. 'Classification of introns: U2-type or U12-type', *Cell*, 91: 875-9.

Singh, R. N., M. D. Howell, E. W. Ottesen, and N. N. Singh. 2017. 'Diverse role of survival motor neuron protein', *Biochim Biophys Acta*, 1860: 299-315.

Singh, R., and R. Reddy. 1989. 'Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA', *Proc Natl Acad Sci U S A*, 86: 8280-3.

Sleeman, J. E., and A. I. Lamond. 1999. 'Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway', *Curr Biol*, 9: 1065-74.

Sprangers, R., M. R. Groves, I. Sinning, and M. Sattler. 2003. 'High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues', *J Mol Biol*, 327: 507-20.

Staley, J. P., and C. Guthrie. 1998. 'Mechanical devices of the spliceosome: motors, clocks, springs, and things', *Cell*, 92: 315-26.

Stark, H., P. Dube, R. Luhrmann, and B. Kastner. 2001. 'Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle', *Nature*, 409: 539-42.

Sun, Y., M. Grimmler, V. Schwarzer, F. Schoenen, U. Fischer, and B. Wirth. 2005. 'Molecular and functional analysis of intragenic SMN1 mutations in patients with spinal muscular atrophy', *Hum Mutat*, 25: 64-71.

Tang, X., S. R. Bharath, S. Piao, V. Q. Tan, M. W. Bowler, and H. Song. 2016. 'Structural basis for specific recognition of pre-snRNA by Gemin5', *Cell Res*, 26: 1353-56.

Tapia, O., V. Lafarga, R. Bengoechea, A. Palanca, M. Lafarga, and M. T. Berciano. 2014. 'The SMN Tudor SIM-like domain is key to SmD1 and coilin interactions and to Cajal body biogenesis', *J Cell Sci*, 127: 939-46.

Thore, S., C. Mayer, C. Sauter, S. Weeks, and D. Suck. 2003. 'Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya', *J Biol Chem*, 278: 1239-47.

Turunen, J. J., E. H. Niemela, B. Verma, and M. J. Frilander. 2013. 'The significant other: splicing by the minor spliceosome', *Wiley Interdiscip Rev RNA*, 4: 61-76.

UniProt, Consortium. 2019. 'UniProt: a worldwide hub of protein knowledge', *Nucleic Acids Res*, 47: D506-D15.

Urlaub, H., V. A. Raker, S. Kostka, and R. Luhrmann. 2001. 'Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure', *EMBO J*, 20: 187-96.

Vallejo, A. N., R. J. Pogulis, and L. R. Pease. 2008. 'PCR Mutagenesis by Overlap Extension and Gene SOE', *CSH Protoc*, 2008: pdb prot4861.

Wahl, M. C., C. L. Will, and R. Luhrmann. 2009. 'The spliceosome: design principles of a dynamic RNP machine', *Cell*, 136: 701-18.

Weber, G., S. Trowitzsch, B. Kastner, R. Luhrmann, and M. C. Wahl. 2010. 'Functional organization of the Sm core in the crystal structure of human U1 snRNP', *EMBO J*, 29: 4172-84.

Will, C. L., and R. Luhrmann. 2001. 'Spliceosomal UsnRNP biogenesis, structure and function', *Curr Opin Cell Biol*, 13: 290-301.

Wolin, S. L., and T. Cedervall. 2002. 'The La protein', *Annu Rev Biochem*, 71: 375-403.

Xu, C., H. Ishikawa, K. Izumikawa, L. Li, H. He, Y. Nobe, Y. Yamauchi, H. M. Shahjee, X. H. Wu, Y. T. Yu, T. Isobe, N. Takahashi, and J. Min. 2016. 'Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly', *Genes Dev*, 30: 2376-90.

Yi, Hongfei, Li Mu, Congcong Shen, Xi Kong, Yingzhi Wang, Yan Hou, and Rundong Zhang. 2018. 'Negative Cooperativity between Gemin2 and RNA Determines RNA Selection and Release of the SMN Complex in snRNP Assembly', *bioRxiv*: 312124.

Yong, J., M. Kasim, J. L. Bachorik, L. Wan, and G. Dreyfuss. 2010. 'Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis', *Mol Cell*, 38: 551-62.

Zhang, R., B. R. So, P. Li, J. Yong, T. Glisovic, L. Wan, and G. Dreyfuss. 2011. 'Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly', *Cell*, 146: 384-95.

Zimmerman, S. B., and A. P. Minton. 1993. 'Macromolecular crowding: biochemical, biophysical, and physiological consequences', *Annu Rev Biophys Biomol Struct*, 22: 27-65.

9 List of publications

1. Aravindan Viswanathan, Clemens Grimm and Utz Fischer (2019). *Structural characterisation of the Gemin8:7:6 module and its engagement within the SMN complex* (manuscript to be submitted)

10 Curriculum Vitae

Aravindan Viswanathan
July 2019, Würzburg

11 Declaration

Affidavit

I hereby confirm that my thesis entitled **'Biochemical and structural characterisation of modules within the SMN complex'** is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg,

Signature:

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation **'Biochemische und strukturelle Charakterisierung von Modulen des SMN-Komplexes'** eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg,

Unterschrift: