Structural Insights into Negative Cooperativity

between Gemin2 and RNA in Sm-class snRNP Assembly

Rundong Zhang

Department of Ophthalmology,

State Key Laboratory of Biotherapy,

West China Hospital,

Sichuan University and Collaborative Innovation Center,

Chengdu 610041, P. R. China

Phone: +86-18200541808

Email: rundongzhang@scu.edu.cn

ABSTRACT

Sm-class ribonucleoprotein particles (RNPs) are ring-shaped structures (Sm cores) formed by Sm hetero-heptamer around a segment of RNA, containing a nonameric oligoribonucleotide, PuAUUUNUGPu, followed by a stem-loop, and are basic structural modules critical for stability and functions of spliceosomal, telemorase and U7 RNPs. In Sm-class RNP assembly, assembly factor Gemin2 not only binds SmD1/D2/F/E/G (5Sm), but also serves as a checkpoint via a negative cooperativity mechanism: Gemin2 constricts the horseshoe-shaped 5Sm in a narrow conformation from outside, preventing non-cognate RNA and SmD3/B from joining; only cognate RNA can bind inside 5Sm and widen 5Sm, dissociating Gemin2 from 5Sm and recruiting SmD3/B. However, the structural mechanics is unknown. Here I describe a coordinate-improved structure of 5Sm bound by Gemin2/SMN. Moreover, via new analysis, comparison of this structure with those of newly coordinate-improved mature U1 and U4 Sm cores reveals how RNAs are selected, 5Sm conformation is changed, and Gemin2 is dissociated from 5Sm. Based on structure-guided sequence alignments, this assembly model is proposed to be conserved for all Sm cores in all eukaryotes. Finally, evolution of the assembly machinery is proposed and implications in spinal muscular atrophy are discussed.

INTRODUCTION

Sm-class ribonucleoprotein particles (RNPs) are a major class of non-coding RNA-protein complexes in eukaryotes that form basic modules for more complicated RNA-protein complexes and play key roles in gene posttranscriptional expression, including pre-messenger RNA splicing (mediated by major spliceosomes U1, U2, U4, U5 and minor spliceosomes U11, U12 and U4atac small nuclear (sn)RNPs) [1-3], suppression of premature cleavage (U1 snRNPs) [4] and histone RNA 3'-end processing (U7 snRNPs) [5-7], and in maintenance of chromosome ends (telomerase RNA subunit) [8, 9]. They are also found in viral RNAs and play important roles in viral functions [10, 11]. The first and also most studied Sm-class RNPs are spliceosomal snRNPs which have a core structure in common, termed as snRNP core or Sm core. In the Sm core, seven Sm proteins, D1, D2, F, E, G, D3, and B in the order, form a ring-like structure around a special segment of snRNA, termed as snRNP code, consisting of a nonameric oligoribonucleotide (Sm site), usually PuAUUUNUGPu, plus a 3'-adjacent stem-loop (SL), which usually contains at least 6-7 base pairs[12]. U7 snRNPs are special Sm-class RNPs, because SmD1/D2 is replaced by heterodimeric Sm-like protein Lsm10/11 and a special Sm site is used, which is AAUUUGU<u>CUAG</u> (the difference from the Sm site above is underlined)[5, 6, 13]. However, U7 snRNPs follow the same maturation pathway as the spliceosomal Sm-class snRNPs [6, 7, 14, 15]. Therefore I will discuss their assembly together with the canonical Sm core.

The assembly of Sm cores is a key step in snRNP biogenesis. Sm core assembly takes a stepwise fashion as has been demonstrated by *in vitro* study from three heteromeric Sm subcomplexes, SmD1/D2, SmF/E/G, and SmD3/B[16, 17]. At first, SmD1/D2 and SmF/E/G associate to form a metastable intermediate complex, which then binds RNA to form a stable Sm subcore. At last, SmD3/B joins to form a highly stable Sm core. The reaction *in vitro* is a

spontaneous process and Sm cores can form on any RNA containing just the nonameric Sm site [16, 17]. However, in eukaryotic cells (especially more complexed metazoans), two complexes, the PMRT5 and SMN complexes, are sequentially involved in the Sm core assembly and they help enhance the assembly specificity of Sm proteins, allowing them only to assemble on the cognate RNAs, mostly snRNAs, which contain not only the Sm site, but a 3'-adjacent stem-loop as well.

The PRMT5 complex is comprised of PRMT5, WD45 and pICln. PRMT5 and WD45 form a hetero-octamer and methylate the C-terminal arginine residues of SmD3, SmB and SmD1, which is thought to enhance the interactions between Sm proteins and SMN [18, 19]. pICln binds SmD1/D2 and SmF/E/G into a ring-shaped 6S complex, stabilizing these two Sm subcomplexes in the finally assembled order and simultaneously preventing the entry of any RNAs to the RNA-binding pocket [20]. In addition, pICln also binds SmD3/B [20].

The SMN complex consists of 9 proteins, SMN, Gemin2-8 and unrip, in higher eukaryotes and functions in the later phase. It accepts SmD1/D2/F/E/G (5Sm) and SmD3/B and releases pICln [20], and also ensures Sm proteins assemble only on cognate snRNAs [12, 21]. Gemin2 is the most conserved component in the SMN complex as only its orthologue is found in all eukaryotes, including single eukaryotes, like S. cerevisiae. Gemin2 is the acceptor of 5Sm [22, 23], which binds the horseshoe-shaped SmD1/D2/F/E/G from the peripheral side, with its N- and C-terminal domains (NTD and CTD) binding F/E and D1/D2 respectively, and leaves the RNA-binding channel of 5Sm open for RNA to enter [22]. SMN is the second conserved component and seems to serve as a scaffold protein. It tightly binds Gemin2 by its N-terminal Gemin2-binding domain (Ge2BD, residues 26-51) [22] [24] and also interacts with Gemin8 by its C-terminal self-oligomized YG box [25]. In vertebrates, either Smn or Gemin2 gene knockout

causes early embryonic death, indicating their essential roles in eukaryotic cells [26, 27]. Moreover, the deficiency of SMN causes neurodegenerative disease spinal muscular atrophy (SMA), emphasizing the medical relevance of the Sm-core assembly pathway [28-30]. Gemin8 further binds Gemin6/7 and Unrip, but their roles are poorly understood [25, 31, 32]. Gemin3, a putative RNA helicase [33], associates with Gemin4 [34], and both are required for Sm core assembly in higher eukaryotes, but their specific functions are unknown. Gemin5 is the component to initially bind precursor (pre)-snRNAs and deliver them to the rest of the SMN complex for assembly into the Sm core [35]. Although it had long been considered as the protein conferring the RNA assembly specificity by direct recognition of the snRNP code [36-41], our recent study indicates that it is Gemin2 that plays the role of enhancing RNA assembly specificity by an unusual way, via a negative cooperativity with RNA in binding to 5Sm (https://doi.org/10.1101/312124).

Gemin2, independent of its N-terminal tail (residues 1-39), constricts the horseshoe-shaped 5Sm in a narrow conformation by binding outside of it (https://doi.org/10.1101/312124). The nonameric Sm site RNA, AAUUUUUGA, which can form Sm subcore if incubated with SmD1/D2 and SmF/E/G alone, cannot bind 5Sm again, even if additional RNA sequence is at its 5'-side. Only the RNAs with a 3'-SL immediately following the Sm site can preferentially get into the RNA-binding pocket inside the horseshoe-shaped 5Sm, but only to a small extent. The binding of RNAs obviously widens the opening between SmD1 and SmG, which makes space for SmD3/B to bind to form the final Sm core. Moreover, the widening of SmD1-SmG also causes Gemin2 to detach from 5Sm (Figure 1). There is no spatial clash between Gemin2 and RNA's binding to 5Sm, but each's binding to 5Sm inhibits the other's. It is the phenomenon of negative cooperativity. This also provides answers to these two basic questions: how the SMN

complex enhances RNA assembly specificity and how it releases. However, in terms of structural point of view, these questions are not answered yet: How does Gemin2's constriction of 5Sm in a narrow conformation enhance RNA assembly specificity? How does RNA binding cause 5Sm conformation to change? And how does the conformational change of 5Sm cause Gemin2 to release from the Sm subcore?

Here I describe a re-refined and coordinate-improved structure of the key intermediate of the SMN complex we had previously determined [22], the 7S complex, which contains Gemin2/SMN (residues 26-62)/SmD1/D2/F/E/G. In addition, using a method different from our previous analysis [22] to make a new comparison and analysis of the reprocessed structure of the 7S complex and the better structures of two mature snRNP cores, U1, which was recently determined at relatively high resolution of 3.3 Å [42], and U4, the coordinate of which was recently improved [43], I propose a model of the conformational changes of the 7S complex upon RNA binding, RNA selection and Gemin2 release at detailed structural level. Based on the model, the evolution of Sm core assembly machinery is also discussed. This model will help understand the fundamental mechanics in Sm core assembly at structural level and facilitate further studies on the roles of other chaperon proteins in this basis biochemical process as well as on pathogenesis of SMA.

MATERIALS AND METHODS

Re-refinement of crystal structure

The previously determined crystal structure of 7S complex [22] is re-refined with reference to two related complex structures accessible in recent years, the NMR structure of the complex Gemin2 (residues 95-280)/SMN (residues 26-51) (PDB code 2LEH) and the 8S complex (PDB code 4V98), containing human SmD1/D2/F/E/G and Drosophila melanogaster pICln (residues 1-180), SMN (residues 1-122) and full-length Gemin2 (residues 1-245) [23, 24] by cycles of manual rebuilding in Coot [44] and REFMAC refinement [45] in CCP4 suite [46]. The final structure has improved quality as indicated by the reduction of R and Rfree from 25.3% and 33.2% to 21.7% and 29.5% respectively (Supplementary Table S1). The final model (PDB code 5XJL) contains SmD1 (residues 1-81), SmD2 (residues 21-77 and 89-117), SmF (residues 3-76), SmE (residues 14-90), SmG (residues 8-52 and 55-72), Gemin2 (residues 22-31, 48-73, 79-124, 135-151, and 174-278), and SMN (residues 35-51). Ramachandran plot shows 95.8% of the dihedral angles in favored region, 2.6% in additional allowed region, and 1.6 (9 out of 576) in disallowed region (Supplementary Table S1). All the 9 outliers are located in the loop regions with relatively poor electron density. Only the regions supported by high quality electron density maps are presented and discussed in detail (Figures 2 & 4).

Building of Sm site RNA model in 7S complex

The first 7 nucleotides of the Sm site in U4 snRNP (PDB code 4WZJ) were individually saved together with their interacting Sm proteins. Each of the coordinates was then aligned with its corresponding Sm protein in 7S complex. The 7 nucleotides were linked in Coot [44] and followed by a relaxing of conformational constrains.

Structure alignment, analysis and preparation of images

PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger) was used for structure alignment and analysis as well as preparation of images.

RESULTS AND DISCUSSION

The refined structure of 5Sm-Gemin2-SMN (residues 26-62) complex

We previously resolved the crystal structure of the key intermediate of the SMN complex, the 7S complex, to a resolution at 2.5 Å at one direction [22]. Although majority of the structure is accurate, due to its anisotropic feature at high resolution, some parts of the complex structure were not precisely determined, which hinders analysis and interpretation of mechanics of conformational changes of the 7S complex, especially on Gemin2 releasing caused by RNA-dependent 5Sm splay. In addition, in recent years, two related complex structures, the NMR structure of the complex Gemin2 (residues 95-280)/SMN (residues 26-51) and the 8S complex, containing human SmD1/D2/F/E/G and *Drosophila melanogaster* pICln (residues 1-180), SMN (residues 1-122) and full-length Gemin2 (residues 1-245), were determined [23, 24], which makes it possible to improve the 7S coordinate with reference to them. In this study, I reprocessed the 7S complex data and gained a better quality of the 7S coordinate as indicated by the reduction of R and R_{free} (Supplementary Table S1). In the new structure, there is little change in the 5 Sm proteins and the SMN segment, but a few subtle yet significant improvements in Gemin2, which are critical for interpretation of the structural mechanism of the negative cooperativity between Gemin2 and RNA in binding to 5Sm. One significant place is in the N-terminal E/F binding domain of Gemin2, where the linker (residues 63-65) between the α1 and β1 is improved (details are in below). The second place is the Gemin2 C-terminal tail (residues 270-280), where a 3_{10}-helix is rebuilt (details are in below). Both these improvements are critical to the explanation of Gemin2-Sm protein interactions and Gemin2 release from Sm core upon RNA binding. Another significant plaFigurece is α3 (residues 139-151), where the register error

is corrected, which contributes to the reduction of R values, although this correction plays little role in Gemin2-Sm protein interactions.

The architecture of the assembly intermediate complex, 7S, is the 5 Sm proteins, D1, D2, F, E and G, form a horseshoe-shaped heteropentamer, with Gemin2 bound on the periphery of 5Sm (Figure 1A). Gemin2 has two functional domains. Its NTD consists of a single α-helix and a single β-strand connected by a short linker. The NTD binds SmE/F on the helix face where the helices of Sm proteins are located. Its CTD is a helical bundle, which binds to SmD1/D2 on the peripheral face. On the opposite side of the CTD is bound by the single SMN helix. Connecting NTD and CTD is a long loop, which is flexible as indicated by the significant loss of electron density. In the previous study, the N-terminal tail of Gemin2 resides inside the RNA-binding pocket and was thought to play an inhibitory role in RNA binding. As in our recent characterization, the N-tail can automatically flip out of the binding pocket and seems to play a less significant role in RNA selection (https://doi.org/10.1101/312124), and therefore it is not drawn in Figure 1 and will not be discussed below.

At the interface of Gemin2-SmE/F (Figure 2A-B), most of the interactions have been described in our previous study [22]. However, new interactions of the linker (residues 63-65) between α1 and β1 of Gemin2 to SmF are found (Figure 2B). The last 2 residues in the 3-residue linker provide two carbonyl oxygens from the main chain to form a hydrogen bonding network with the side chains of Lys8 and Asn12 on the N-terminal helix of SmF. In this way, the tight hydrogen bonding interactions are continuous from the main chain of the residue Pro64 to Ala69 of Gemin2, like a zipper, covering last two thirds of the linker and the following β1. Gemin2 α1 has two types of interactions to SmE/F: the first one is on the N-terminal part of α1, where the hydrophobic interactions dominate between Pro49, Tyr52, Leu53 and Val56 from Gemin2, and

Ile18, Phe22, Leu25, and Phe50 from SmE. These interactions are kind of "ridges-into-grooves" pattern. The second type is hydrogen bonding interactions from three residues, Tyr52, Gln57 and Glu59 from the middle part of Gemin2 α1, to the main chain amide of Phe50 and the side chain of Glu52 from SmE, and to the side chain of Asn6 from SmF, respectively. Overall, Gemin2 NTD is like two sticks, one being α1(residues 49-62) and the other being the zipper (residues 64-69), connected by a short joint (residue 63). Roughly each "stick" interacts with one helix and one strand from each Sm protein.

The CTD of Gemin2 is a helical bundle consisting of 7 α-helices (α2-α8). In the new structure, there is also a 3_{10}-helix after α8 (Figures 1A & 2C-D), which is a short, but conserved structure (Figure 3). The QxDL(272-275) motif is highly conserved sequence. Q272 and D274 form salt bridges with R235 from α7. Interestingly, these three residues are 100% conserved among the orthologs of various eukaryotic species (Figure 3). In addition, Q272 also hydrogen bonds with the main chain amide nitrogen from F270 at the C-terminal end of α8 to maintain the 3_{10}-helix conformation. Leu275 faces the interface between α7-8. The hydrophobicity of this residue is conserved (Figure 3). D277 form salt bridges with R245 from α7. These salt bridges are also highly conserved (Figure 3). These further stabilize the 3_{10}-helix interaction with its neighboring α7-8. Recently, the importance of the QxDL motif was signified in the Gemin2 orthorlog protein Brr1 in *S. cerevisiae* by a genetic approach. The truncated *BRR1-CΔ6* allele encoding Brr1(1-335), in which the last hexapeptide QKDLIE(336-341) was deleted, and the Brr1 double mutant Q336A/D338A failed to act as the wild-type Brr1 in complementing *brr1Δ/SMF-F33A* and *brr1Δ/SME-K83A*[47].

The interactions of Gemin2 CTD with SmD1 are the same as described in our previous study [22]. Briefly, the N-terminal parts of α5 and α7 and the loop between α6-α7 of Gemin2

make contacts with β2 near loop 2 and β3-β4 near loop4 of SmD1, and most of the contacts are hydrogen bonding (Figure 2E). However, particular attention should be paid to the interaction pattern: the interacting atoms of Gemin2 are mostly from the main chain. This explains why the sequence of the loop between α6-α7 in Gemin2 is not conserved but its length indeed is, indicating a less flexibility on this loop (Figure 3). On the other hand, the interacting residues in SmD1 are mostly located near loops 2 and 4, where variations are possible.

The interactions of Gemin2 CTD with SmD2, however, are more than described previously [22]. The main interacting parts are α7 and 3_{10}-helix from Gemin2 and β3-β4 from SmD2 (Figure 3F). From Gemin2 α7, R235 and R239 make salt bridges with D93.SmD2, and H231 makes hydrogen bond with R94.SmD2, as described previously. In addition, S232 hydrogen bonds the amide nitrogen of D93.SmD2 and the side chain of E229 hydrogen bonds with the side chain of W74.SmD2. From Gemin2 3_{10}-helix, Q272 forms hydrogen bond with the amide nitrogen of Y95.SmD2, and R273 forms a hydrogen bond network with the side chains of N69 and Y95 from SmD2. Overall, the interactions between Gemin2 CTD and SmD2 are mostly polar and are intensive. Because of the above described features, the interacting surface of Gemin2 to SmD1/D2 is highly rigid and can be viewed as a solid rock surface.

In metazoans, the assembly of U7 snRNP takes the same way as spliceosomal snRNP cores do, and is also mediated by the SMN-Gemin complexes [6, 7, 14, 15]. U7 Sm core shares five Sm proteins with spliceosomal snRNP cores, but has Lsm10/11 heterodimer in the place of SmD1/D2. The alignments of Lsm10 versus SmD1 and Lsm11 versus SmD2 reveal that majority of the residues to interact with Gemin2 CTD are maintained or replaced by similar residues (Supplementary Figure S1). So, we can predict that Gemin2 binds Lsm10/11 and SmF/E/G in a similarly narrow conformation and applies a similar constraint role on U7 snRNAs.

RNA-dependent widening of 5Sm and RNA selection

RNA binding is the cause of the splay of 5Sm from the 7S intermediate to the Sm subcore, which should have the conformations of both RNA and 5Sm very similar to these in the mature Sm core (Figure 1B-C), although the structure of Sm subcore has not been experimentally determined. At first, the conformation of the Sm site RNA inside Sm subcores and cores is unusual because the RNA backbone, which is negatively charged, forms a ring inside a very narrow space with the bases protruding outside to contact Sm proteins (Figures 1B-C & 4A). There must be very strong repulse among the negative backbone charges. This conformation of RNA would hardly exist in itself. The existence of such a highly tensive ring of the Sm site is only possible when there are very intensive specific interactions between the 9 Sm site bases and at least 5 Sm proteins (Sm subcore). The recent U1 and U4 crystal structures describe the detailed interactions between the Sm site bases and 7 Sm proteins [42, 43]. These interactions form the primary structural basis for snRNA binding specificity in Sm core assembly.

When superimposing SmD2 of the snRNP cores and 7S, there are clearly clashes of SmE and SmG with the Sm site RNA and of SmE with the 3'-SL RNA (Figure 4A-B), indicating that the RNA conformation in its initial binding to 7S is different from that in mature Sm cores. To make a Sm site model in the 7S Sm heteropentamer, the first 7 bases of the Sm site in U4 snRNP were individually saved and linked again after individual Sm protein is aligned with the corresponding one in 7S, followed by a relaxing of conformational constrains. In contrast to a circle seen inside U4 snRNP, the Sm site RNA inside 7S is ellipse with two bases (Ura4 and Ura5) bulging out at the SmD1-SmG opening (Figure 4C-D). The negative charges at the opposite sides of the minor axis of the ellipsoid Sm site RNA backbone would repulse each other much more than those at the opposite sides of the major axis. In other words, the evenly

distributed negative charges on the backbone tend to splay the ellipsoid Sm site RNA into a circle, which provides the transition force to widen 5Sm.

A nonameric Sm site RNA oligonucleotide (9nt), AAUUUUUGA, is able to form a stable Sm subcore with SmD1/D2 and SmF/E/G alone [17], however it cannot bind to 7S (https://doi.org/10.1101/312124). The formation of Sm subcore from SmD1/D2 and SmF/E/G alone should have a transient 5Sm complex first before the binding of 9nt. Although we don't know whether there is a narrow conformation of 5Sm complex, if there is, the conformation of 5Sm would be quite flexible to be able to allow the binding of 9nt and an easy splay of 5Sm. The intensive specific interactions between 9nt and 5Sm seem sufficient to provide the energy to keep the Sm site RNA stably binding to 5Sm and very reasonably in a circular conformation as it is in the mature Sm core. However, in 7S, because the binding of Gemin2 from the periphery of 5Sm constricts them in a stable, narrow conformation, the binding energy provided solely from the interactions between 9nt and 5Sm seem not enough for 9nt RNA to bind to 5Sm in 7S. We can image that a very transient complex would have the 9nt Sm site RNA bound to 5Sm but staying in a highly tensive ellipsoid shape and therefore the RNA would be immediately repealed out of 5Sm. To bind to 7S, there must be an additional RNA at the 3' side of the Sm site, which can provide additional interactions with 5Sm and additional energy. This explains why a U4 RNA derivative keeping 5' side of Sm site intact but deleting 3' side is unable to bind to 7S (https://doi.org/10.1101/312124).

The best candidate at the 3' side of the Sm site is a sequence-independent stem-loop, which has a minimal 5-6 base pairs, immediately following the Sm site. These features together with the Sm site termed snRNP code and are characteristic features of snRNAs [12]. A detailed analysis of RNA-Sm protein interactions in U4 snRNP (PDB code 4WZJ) explains why it is the

case. There are many positively charged residues on the inner side of 5Sm to interact with both strands of the 3'-SL (Figure 5A-B & G). From the downward strand of 3'-SL, R47 and N49 from loop 2 in SmD2 interact with phosphate backbone at the 1st and 3rd base pairs (Figure 5D&G), R66 from β5 and K48 from loop 2 in SmD1 contact the phosphates at the 2nd and 3rd base pairs (Figure 5C&G). Going further down, there are a group of positively charged residues, K79, K82 and K85 in loop 4 from SmD2 contacting the phosphates at the 5-6th base pairs. In addition, the RNA upward strand also contacts Sm proteins. R50 in loop4 of SmD1 forms salt bridges with the backbone between the 8th and 9th base pairs. K84 in loop4 of SmD2 forms salt bridges with the backbone between the 10th and 11th base pairs. K22 at β1 near loop2 of SmF contacts the phosphate at 3'-end of RNA (Figure 5E). SmE makes more polar contacts with the RNA upward strand (Figure 5F). K67 at loop4 contacts the phosphate between the 2nd and 3rd base pairs, N40 at loop2 contacts the phosphate between the 1st and 2nd base pairs, Q38 at loop2 contacts the phosphate at free RNA 3'-end, and H65 at loop4 hydrogen bonds the 2' oxygen of the last ribose. These charged interactions provide additional binding energy for RNA to bind into the central channel of 5Sm and for the Sm site to splay into a circle, which widen the opening between SmD1 and SmG. Almost all these interactions are between the RNA backbone and the Sm proteins. This explains the sequence independence of the 3'-SL in RNA binding[12]. The size of 3'-SL had been studied on U7 snRNA to test its influence on U7 Sm core assembly and it was found that less than 5 base pairs in the 3'-SL impair U7 Sm core assembly [13]. It is consistent with the fact that majority of the interactions of 3'-SL with Sm proteins are located within the first 5 base pairs. As there are substantial interactions between the upward backbone of the first 3 base pairs of the 3'-SL and SmE/F, especially SmE, a SL immediately adjacent to the 3' end of the Sm site would be preferred for Sm core assembly. This gives a good explanation of an in

vivo selection experimental result [48]: a library of 20 random nucleotides (N20) inserted into the position between two SLs of a U1 snRNA derivative, replacing the original Sm site, was injected into oocytes, and the assembled Sm cores in the nucleus were enriched and selected by 12 rounds of Sm-specific Y12 antibody precipitation, PCR amplification and injection. The selected RNAs assembled into Sm cores all contain AAUUUUUGG at the 3' end of N20, immediately adjacent to a 3'-SL.

For biochemical characterization, a single-stranded RNA 3'-downstream of the Sm site seems sufficient to allow RNA to bind into the binding site and to widen the SmD1-SmG opening (https://doi.org/10.1101/312124). An *in vivo* experiment in which injection of de-stem U4 RNA into oocytes could be assembled into Sm core albeit with low efficiency also demonstrates this [12]. A single-stranded RNA would keep most of the polar interactions with SmD1 and D2, as the downward strand of the SL does. But SL structure is more efficient, as it not only provides additional interactions to SmF and SmE, but also offers a more rigid double-stranded 3'-RNA for better interactions with Sm proteins. This feature explains why snRNAs having a 3'-SL immediately downstream of the Sm site dominate the Sm core assembly inside cells [21].

The interactions between Sm proteins and RNA are also highly conserved among eukaryotes. For example, R47 and N49 in SmD2, R66 in SmD1 and K22 in SmF are 100% conserved (Supplementary Figure S1). K48 and R50 in SmD1 and K79 and K85 are highly conserved, the changes being mostly from K to R or R to K. A special case is *Saccharomyces cerevisiae* in which Loop 4 of SmD2 is shorter than those of all other organisms. The loop 4 of SmD2 contains several RNA-interacting residues, which should be lost in *S. cerevisiae* SmD2. However, the loop 4 of SmD1 is 27 residues longer than those of other organisms and it contains

fully or partially positively charged residues, K, N and Q, which would potentially interact with 3'SL and compensate for the loss of interaction of Loop 4 of SmD2 with RNA.

As in the assembly of U7 snRNP core Lsm10/11 replaces SmD1/D2, comparison of the SL-interacting residues of Lsm10 versus SmD1 and Lsm11 versus SmD2 also gives rise to a high conservation (Supplementary Figure S1). K20, K48 and R66 in SmD1 have R34, R62 and R80 in their equivalent positions in Lsm10. R47 and K82 in SmD2 have the same residues in Lsm11 in their equivalent positions. This high conservation explains similar interactions between Lsm proteins and 3'SL and a similar requirement of 3'SL of U7 snRNA in U7 Sm core assembly.

Conformational changes of 5Sm

In the previous analysis by comparison of 5Sm in 7S and U1 snRNP, it was suggested that the interfaces within each of the Sm sub-complexes, SmD1/D2 and SmF/E/G, are rigid, but the SmD2-SmF interface is widened in U1 snRNP [22]. As Gemin2's NTD and CTD bind to SmF/E and SmD1/D2 respectively, and only the flexible connecting loop between NTD and CTD corresponds to the SmD2-SmF interface, this suggestion could not explain both RNA binding selection and Gemin2 release. In the previous analysis, full length sequences of Sm proteins were used for superimposition, which might shield potential subtle changes of their N-terminal helices and the following β sheets. To make a better comparison, in this study, 48 residues of each Sm protein, which are from the β sheet and are less varied, are used (Figure 6A). At first, the recent U1 and U4 snRNP cores, which have the highest resolutions of 3.3 and 3.6 Å respectively, are examined (PDB code 4PJO and 4WZJ)[42, 43]. There are 4 U1 snRNP molecules and 12 U4 snRNP molecules in one asymmetric unit (ASU) respectively. Do they have similar conformations? Which one should be used for comparison? Superposition of 4 U1 snRNP

molecules gives rise to almost identical conformation of the Sm proteins (Supplementary Figure S2A), indicating that the conformation of 4 different molecules in each ASU are very similar. Superposition of 12 U4 snRNP core molecules also produces a similar conclusion. Then one representative U1 snRNP Sm core is compared with one U4 snRNP Sm core. When each pair of the corresponding Sm proteins is superposed, although the root mean square deviations (RMSDs) are larger than those of U1 or U4 core molecules in ASU, the neighboring Sm protein pairs are also superposed well (Supplementary Figure S2, B-D), indicating that the conformation of Sm proteins and their relative positions in U1 and U4 snRNP core do not vary much. This conclusion indicates that using either U1 or U4 snRNP core is proper for comparison of 7S and snRNP core and analysis of conformational changes from the 7S assembly intermediate state to mature snRNP core.

When any pair of the 5 corresponding Sm proteins in 7S and U1 snRNP is superposed, the widening of the SmD1-SmG opening in the mature U1 snRNP is apparent (Figure 4A). To quantify the variations of each pair of Sm proteins, the RMSDs of all 5 pairs of Sm proteins were calculated when each pair is superposed. The RMSDs of SmD1 and SmD2 pairs are relatively small (less than 0.8 Å for 192 main chain atom pairs each) when either adjacent pair is superposed (Figure 6B). However, the RMSD of SmF pair is relatively large (1.82 Å for 192 main chain atom pairs) when SmD2 pair is superposed and vice versa. This indicates that the conformational changes between SmD1 and SmD2 are smaller than those between SmD2 and SmF. This conclusion is consistent with our previous study. However, when each pair of SmF, SmE and SmG is superposed respectively, the RMSDs of their adjacent Sm protein pairs are also large. For example, the RMSDs of SmD2 and SmE pairs are 1.64 and 1.69 Å respectively when SmF pair is superposed. This indicates that the relative position between SmF and SmE, as well

as between SmE and SmG are also altered significantly upon snRNA binding. This conclusion has not been noticed before.

To make more detailed analysis of the conformational changes between Sm proteins, two parameters are used (Figure 6C): the distance between the Cα of the 48 residues of each pair of adjacent Sm proteins, $d_n(Sm_1\text{-}Sm_2)$, and the angle of the Cα of the 48 resides from 3 adjacent Sm proteins, $\angle_n(Sm_1\text{-}Sm_2\text{-}Sm_3)$. For each adjacent Sm pair, either in 7S or in U1 snRNP, distances are small at the loop regions, which face the central RNA-binding channel, and large at the perimeter (Figure 6E & 6F). The distance changes, Δd_n ($=d_m\text{-}d_i$), (m, mature Sm core, and i, intermediate 7S), are mainly positive at each loop and negative at the perimeter (Figure 6G), indicating a splay of the crescent 5Sm. Large distance increases occur at two loop regions, loops 2 and 4, both of which are located at the lower part of 5Sm interacting with 3'-SL RNA. The average Δd_n are between -0.8 and 0.8 Å, indicating a small adjustment upon snRNA binding to 5Sm. However, when examining individual adjacent Sm pairs, the changes are not equal at their corresponding positions. The distance changes are relatively small for SmD1-SmD2 and SmE-SmG, while larger changes occur for SmD2-SmF and SmF-SmE. This is consistent with the larger base insertions, A1 into between SmD2 and SmF, and A2 and A9 into between SmF and SmE.

The average angle is 124.27° for 7S and 128.80° for U1 snRNP (Figure 6H & 6I). Considering a regular hexagon with inner angle of 120° and a regular heptagon of 128.57° (Figure 6D), the angle for U1 snRNP is very close to the standard value of a heptamer, but the Sm pentamer occupy 5 positions of a rough 6.5-Sm-forming ring. Interestingly, the average angles of D1-D2-F, D2-F-E, and F-E-G are not similar. They are 124.46°, 120.45° and 127.92° in 7S, and 128.49°, 126.13° and 131.79° in U1 snRNP. This indicates that the angle from D2-F-E

is smaller than those of other adjacent Sm proteins and is close to the standard angle of a hexamer in 7S. In addition, although the average angle change, $\Delta\angle$ ($\angle_m$ - $\angle_i$), is 4.53°, the average angle change for each set of adjacent Sm proteins varied (Figure 6J). They are 4.03°, 5.68° and 3.87° for D1-D2-F, D2-F-E and F-E-G. This indicates that the angle from D2-F-E is widened more upon RNA binding and Sm core formation.

In addition, the angle changes over the 48 positions vary very much. At the loop3, $\Delta\angle_{18\text{-}22}$ (D2-F-E) increase dramatically, reaching a value of about 15°, three-fold the average change. This is caused by the insertions of two adenines into D2-F and F-E, consistent with the distance changes at this location stated above. Correspondingly, the $\Delta\angle_{18\text{-}22}$ (D1-D2-F) and the $\Delta\angle_{18\text{-}22}$ (F-E-G) become negative. At the loops 2 and 4, $\Delta\angle$ (F-E-G) are larger than the average while $\Delta\angle$ (D2-F-E) are close to zero and even becomes negative. This is caused by the 3'-end of RNA passing between the loops 4 of SmF and SmE. In summary, the conformations of individual Sm proteins and the interfaces between adjacent Sm proteins are changed, but not evenly, which are affected by the local environment of the bound RNA.

Structural basis for Gemin2 release

The binding of the snRNA to the Sm pentamer in 7S complex triggers the release of Gemin2. The widening of 5Sm accompanies the splay of the Sm site RNA is the major reason. As analyzed above, the changes occur all over the Sm proteins, although most of them are quite subtle. Because there is no apparent constraint exerted on the interface between SmD2 and SmF by Gemin2, the subtle changes of the interactions between Gemin2 NTD and SmF/E and between Gemin2 CTD and SmD1/D2 are the major source of Gemin2 release.

Superposing SmE or SmF can generate a slight twist and location shift. When superposing SmE based on the 48 residues, for example, the distance between the Gemin2-interacting residues are slightly altered (Supplementary Figure S3A). The Cα-Cα distance between I18 and E52 decreases 0.4 Å, while that between L25 and F50 increases only 0.2 Å. When superposing SmF on the basis of the 48 residues, a rigid-body shift of the helix 1 toward periphery is observed (Supplementary Figure S3B). The Cα-Cα distance between N12 and G38 increases 1.7 Å, while that between N12 and V34 decreases 1.1 Å. The distance between the helix 1 of SmE and β1 of SmF, the two structural parts marking the edges of the Gemin2 NTD-SmE/F interaction set, is reduced (Figure 7A & 7B). The Cα-Cα distance between I18.SmE and G38.SmF decreases 2.3 Å, from 22.9 to 20.6 Å. It is just about 10% reduction. How can such a small change of the SmE/F make Gemin2 NTD get away? One might think that such a small change can be accommodated by protein adjustment as we all know proteins are usually plastic. However, the key point here is that Gemin2 NTD is like two sticks linked by a joint. When helix 1 of SmE moves towards SmF, the "ridges into grooves" interaction between the N-terminal α1 of Gemin2 and helix 1 of SmE makes the entire Gemin2 α1 move towards SmF too , the extra 2.3 Å spatial tension needs the 3-residue linker to release (Figure 7A). There are only two ways to do so: (1) popping the linker inward to the central RNA-binding channel; (2) popping the linker outward to the periphery (Figure 7C). In the case (1), the linker moves in the direction opposite to that of its interacting residues K8 and N12 of SmF, and the hydrogen bond network breaks. In addition, as the connection between α1 and the linker is covalent, the C-terminal part of α1 would also move in the direction of the linker. This would increase the distance between Y52, E59 in Gemin2 and N6 in SmF, and the hydrogen bonds between them would break too. This way, multiple hydrogen bonds would be lost. In the case (2), the linker moves in the same

direction as that of its interacting residues K8 and N12 of SmF, the hydrogen bond interactions mentioned above would be kept. However, the extra 2.3 Å length tension would be relayed to β1 of Gemin2. As the distance between two adjacent Cαs in β strand is about 3.8 Å, the 2.3 Å extra length is too short to shift the register of the interacting residues by one to make hydrogen bonding again. Instead, the distance between adjacent hydrogen bonds from antiparallel β strands is about 2.4 Å, the 2.3 Å relay would actually move H-donor nitrogen on Gemin2 β1 to face H-donor nitrogen from β strand of SmF. All the hydrogen bonds between the antiparallel strands would not only be lost, but the elements on the strands otherwise hydrogen bonding would be in a repulse mode. This way, multiple hydrogen bonds would lose too. In either way, the "stick-joint-stick" of Gemin2 NTD would no longer fit into the binding sets of SmF/E after the splay of 5Sm.

The superposition of SmD2 does not make many shifts of the interacting residues from SmD2, as the Cαs are only moved within short distance of 0.1-0.2 Å (Figure 7D). However, loop 2 and loop 4 in SmD1 move closer to 3'-SL RNA, because of the positively charged residues at the loops making polar contacts with the RNA backbone. As the contact residues of SmD1 with Gemin2 CTD are located in loops 2 and 4, the flipping of the loops of SmD1 towards the bound RNA increases the distance between the interacting surfaces. And as the SmD1/D2-interacting surface of Gemin2 CTD is quite rigid like a rock surface, it would hardly make a plastic adjustment and therefore, the contact between Gemin2 CTD and SmD1 would be lost (Figure 7D-E).

In summary, in the 7S intermediate state of Sm core assembly, Gemin2 constrains 5Sm in a narrow conformation mostly by its NTD and CTD, which makes the splay of a nonameric Sm site RNA oligonucleotide hard to achieve unless a 3'-SL immediately follows the Sm site RNA,

therefore determining RNA specificity. The splay of the Sm site causes the conformational change of 5Sm, including the splay of 5Sm and the loops of SmD1 moving towards 3'-SL RNA, which reduces the affinity of both NTD and CTD of Gemin2 to 5Sm and causes Gemin2 to release. In addition, the opening of 5Sm is also required for the binding of SmD3/B, which stabilizes the widened conformation of 5Sm and moves Sm core assembly to a completion. From the above structural analysis, this mechanism should be conserved in all eukaryotes and apply to all kinds of Sm core assembly mediated by the SMN-Gemin2 complex, including canonical spliceosomal snRNP cores and special U7 snRNP cores.

Evolution of Sm core assembly machinery

Gemin2 binds 5Sm, specifies RNA together with the bound 5Sm, and releases Sm core upon finishing the assembly. Gemin2 interacts with Sm proteins with highly conserved residues and is the most conserved protein of the SMN complex. From evolutionary point of view, Gemin2 is the central protein of the assembly machinery of Sm core and could have been the first protein evolved in this pathway. It is consistent with the observation that there is only Brr1, the ortholog of Gemin2 of the 9-component human SMN complex, is found in the simplest eukaryote, *S. cerevisiae* [49, 50]. The join of SMN protein might help to replace pICln more efficiently as both pICln and SMN proteins are present in *Schizosaccharomyces pombe* [51-54], another single cell eukaryote with more requirement of snRNP core formation as most of its genes contain introns. Gemin2 plus the N-terminal part of SMN (residues 1-122), which contains the Gemin2 binding domain, binds the 6S complex, which contains 5Sm and pICln, but stalks in the intermediate 8S state, where pICln blocks the entry of RNAs to the binding pocket [20, 23]. The C-terminal part of SMN protein may play an important role in releasing pICln from 5Sm. In Drosophila, the SMN and Gemin2 are sufficient to assemble Sm core efficiently

[55]. Therefore, Gemin2 and SMN are the minimal assembly machinery in most eukaryotes with less complexity. Gemin5 may have joined in this pathway accidentally by mutation of an ancient WD40 domain-containing protein to gain its RNA binding capability, therefore interfering with Sm core assembly. To cope with this accident, cells may have evolved a RNA helicase-containing protein, Gemin3 [33], to take off RNAs from Gemin5, and other proteins, like Gemin4, Gemin6/7/8 and unrip, to aid in transferring Gemin5 to the vicinity of the SMN-Gemin2 assembly line to allow the RNA take-off at site. In vertebrates, the assembly of Sm core requires ATP hydrolysis [56], and this might be the driving force for Gemin3 to take RNA out of Gemin5.

The Sm cores evolved as the assembly machinery, the SMN-Gemin2 complex, became more complex. In simple eukaryotes there is no Lsm10/11. But in metazoans, Lsm10/11 appears and assembles into a different Sm core by replacing SmD1/D2 and accepting a slightly different U7 snRNA, which has a different Sm site but still keeps a 3' adjacent SL. More importantly, U7 Sm core assembly can still use the same assembly machinery.

Implications in SMA

The illumination of the basic mechanism of Sm core assembly also has important implications in pathogenesis of SMA. Since Gemin2 plays most of the basic roles in Sm core assembly and there might be a way to assemble Sm cores in any eukaryotes without the need of the SMN protein, like in S. cerevisiae, where only the ortholog of Gemin2, Brr1, is important in snRNP core assembly[49]. The way to assemble Sm core without SMN at least could be used to dissect the pathogenesis of SMA. The SMA-causing protein, SMN, might only play a role in efficiently releasing pICln to make the Gemin2-bound 5Sm accessible for RNAs. The unstable,

dysfunctional SMNΔexon7 product, might bind to most of Gemin2 using its N-terminal Gemin2 binding domain, and subject both Gemin2 and itself to degradation, and therefore cause low efficiency of snRNP core assembly. A stable protein which can bind to Gemin2 in competition with SMNΔexon7 might avoid this passive loss of Gemin2 in cells in studies of SMA therapy.

ACKNOWLEDGEMENTS

This work was supported by National Key R&D programs (No. 2017YFA0504300 and 2017YFA0505900) and National Natural Science Foundation of China (No.31570720 and 81441109). Structure factors and atomic coordinate have been deposited in PDB (code 5XJL). The author declares no competing financial interests.

FIGURES

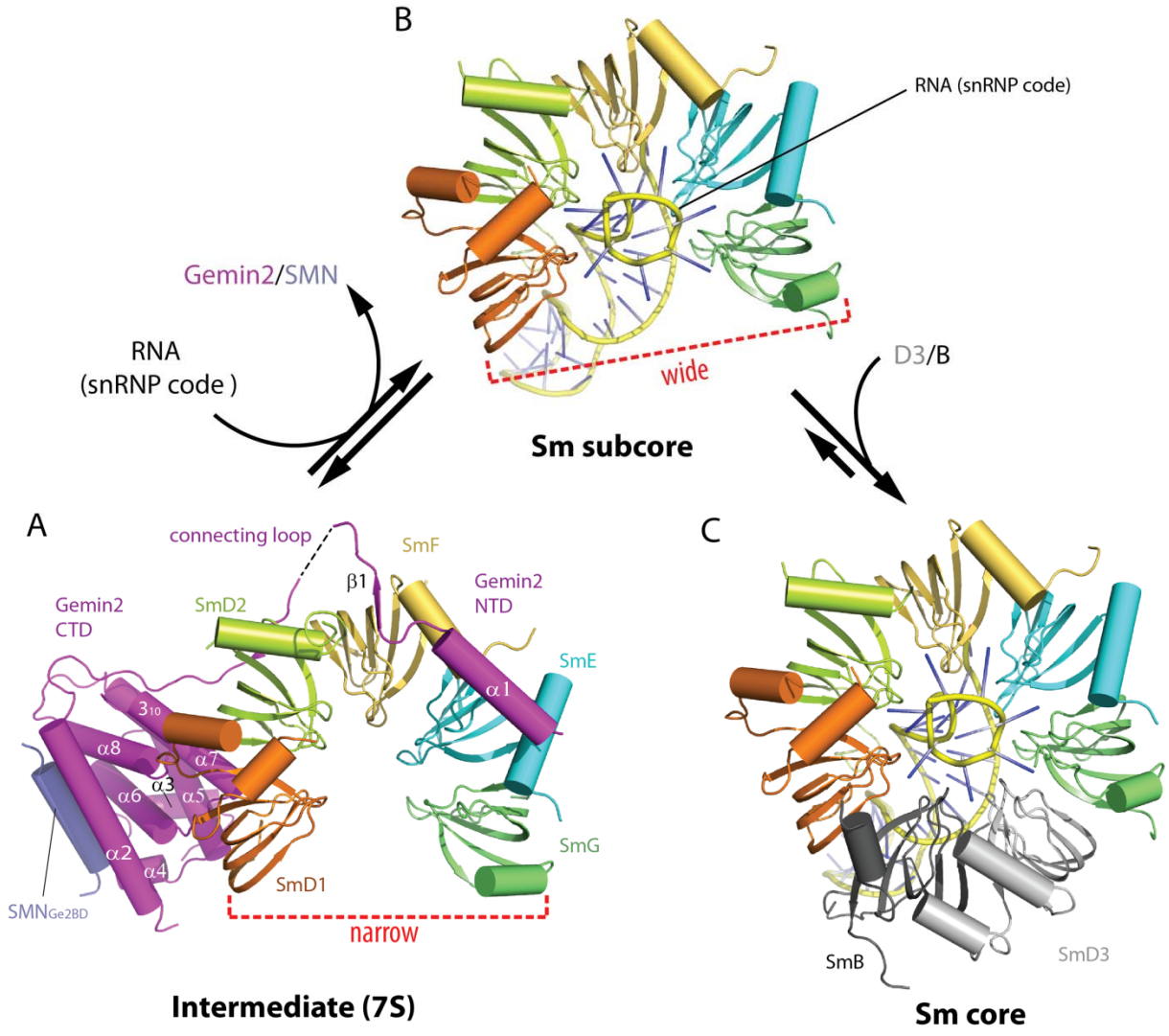


Figure 1. Structures of the 7S intermediate, Sm subcore and Sm core and the transition during assembly. (A) 5Sm bound by Gemin2/SMN_{Ge2BD} is in a narrow conformation (PDB code 5XJL). The N-tail of Gemin2 (residues 1-47, only 22-31 is visible in crystal structure) is not shown. (B) Sm subcore, which has not been determined, but should be very similar to Sm core (PDB code 4WZJ) without SmD3/B. (C) Mature Sm core represented by U4 snRNP (PDB code 4WZJ). There are two steps, RNA binding coupled with Gemin2/SMN release to form Sm subcore, and SmD3/B joining to form Sm core, in the process of Sm core assembly. The 7 Sm proteins, D1, D2, F, E, G, D3 and B, are colored in orange, lemon, yellow, cyan, green, light grey and dark grey respectively. Gemin2 and SMN_{Ge2BD} are colored in purple and blue respectively.

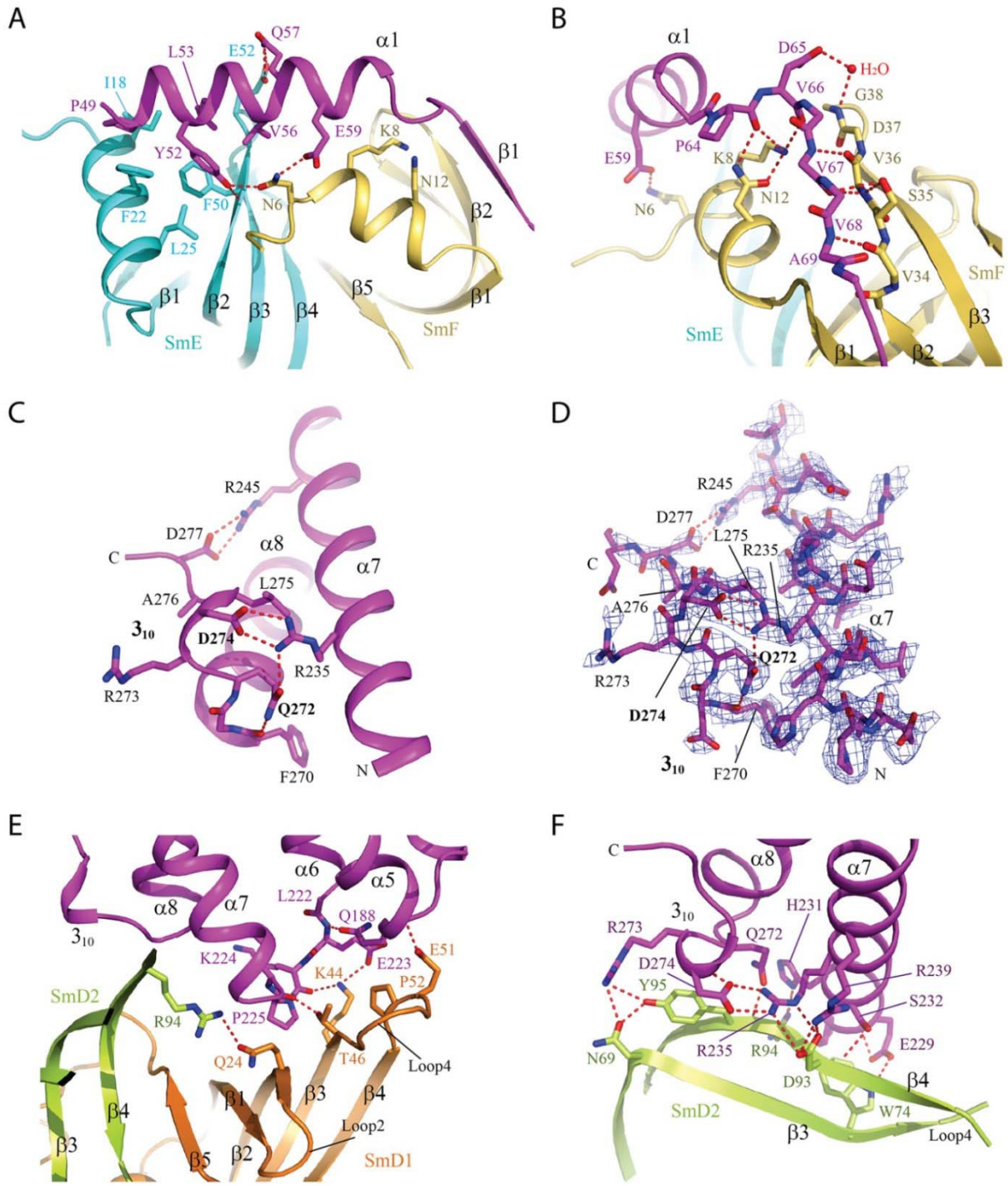


Figure 2. Detailed views of interactions between Gemin2 and Sm proteins. (A-B) Gemin2 NTD interacts with SmF/E. (C-D) the C-terminal 3_{10} helix of Gemin2 and its interactions with helices α7-8. SigmaA-weighted 2Fo-Fc electron density maps (blue meshes) are contoured at 1.1σ in (D). (E-F) Gemin2 CTD interacts with SmD1/D2. Hydrogen bonds and salt bridges are shown as red dashed lines. Color scheme is the same as in Figure 1. Oxygen and nitrogen are colored in red and blue respectively. Water is shown in red sphere. See main text for detailed description.

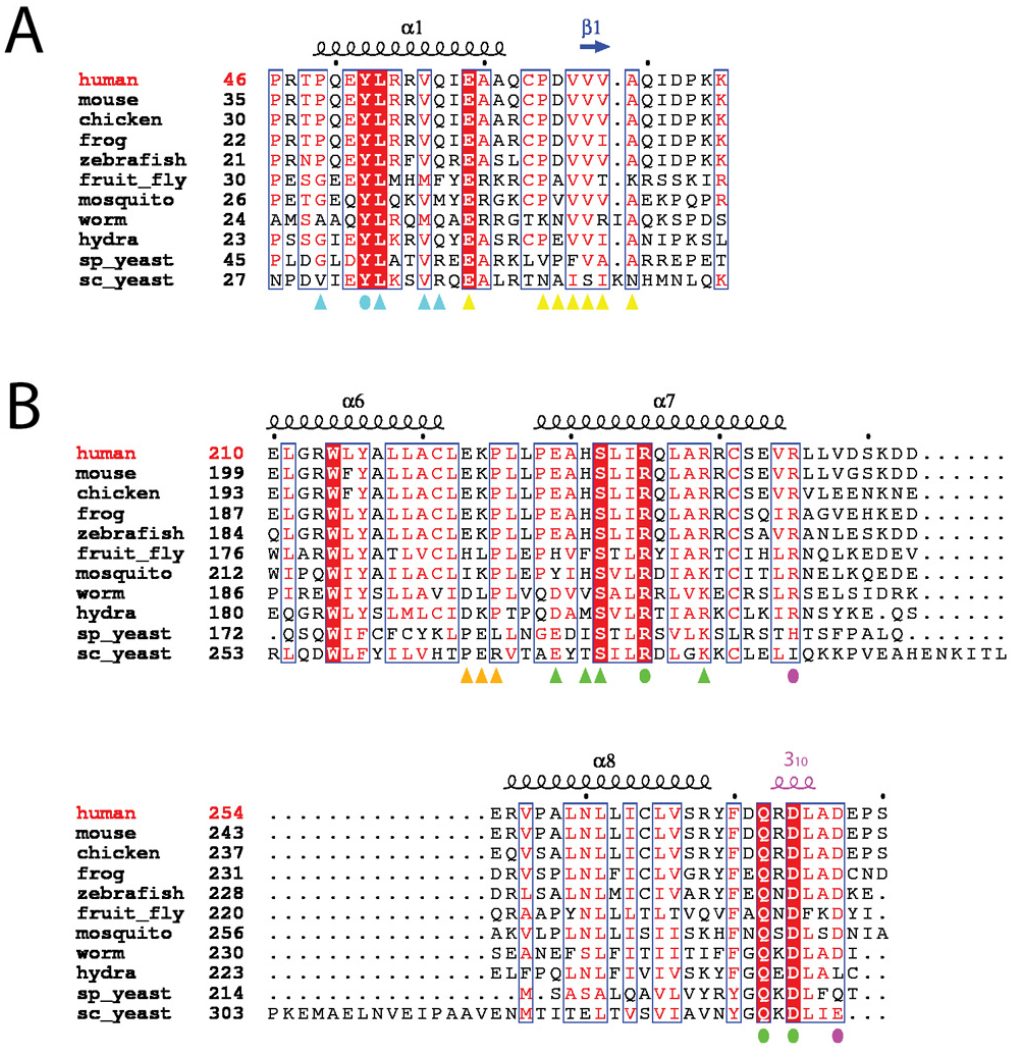


Figure 3. The residues of Gemin2 orthologs contacting Sm proteins in diverse organisms are conserved. Multiple sequence alignment was performed by Clustal Omega[57] and the figure was prepared by ESPript [58]. (A) Gemin2 NTD. (B) Part of Gemin2 CTD interacting with SmD1/D2. Human, *Homo sapiens* (accession number NP_003607); mouse, *Mus musculus* (NP_079932); chicken, *Gallus gallus* (NP_989530); frog, *Xenopus laevis* (NP_001087945.2); zebrafish, *Danio rerio* (NP_001017608); fruit_fly, *Drosophia melanogaster* (NP_649092.1); mosquito, *Anopheles gambiae* (XP_001688660.1); worm, *Caenorhabditis elegans* (NP_001022847.1); hydra, *Hydra magnipapillata* (XP_002160350); sp_yeast, *Schizosaccharomyces pombe* (CAB88094); sc_yeast, *Saccharomyces cerevisiae* (NP_015382.1). Absolute identical residues (100%) are in white font with red background. Highly conserved residues (over 70%) are in red font and boxed in blue. The secondary structure of human Gemin2 is shown on the top. Cyan, yellow, orange and green triangles indicate the residues contacting SmE, SmF, SmD1 and SmD2 respectively. Cyan circle indicates the residue contacting both SmE and SmF. Green circles indicate the residues contacting each other as well as SmD2. Pink circles indicate the residues contacting each other.

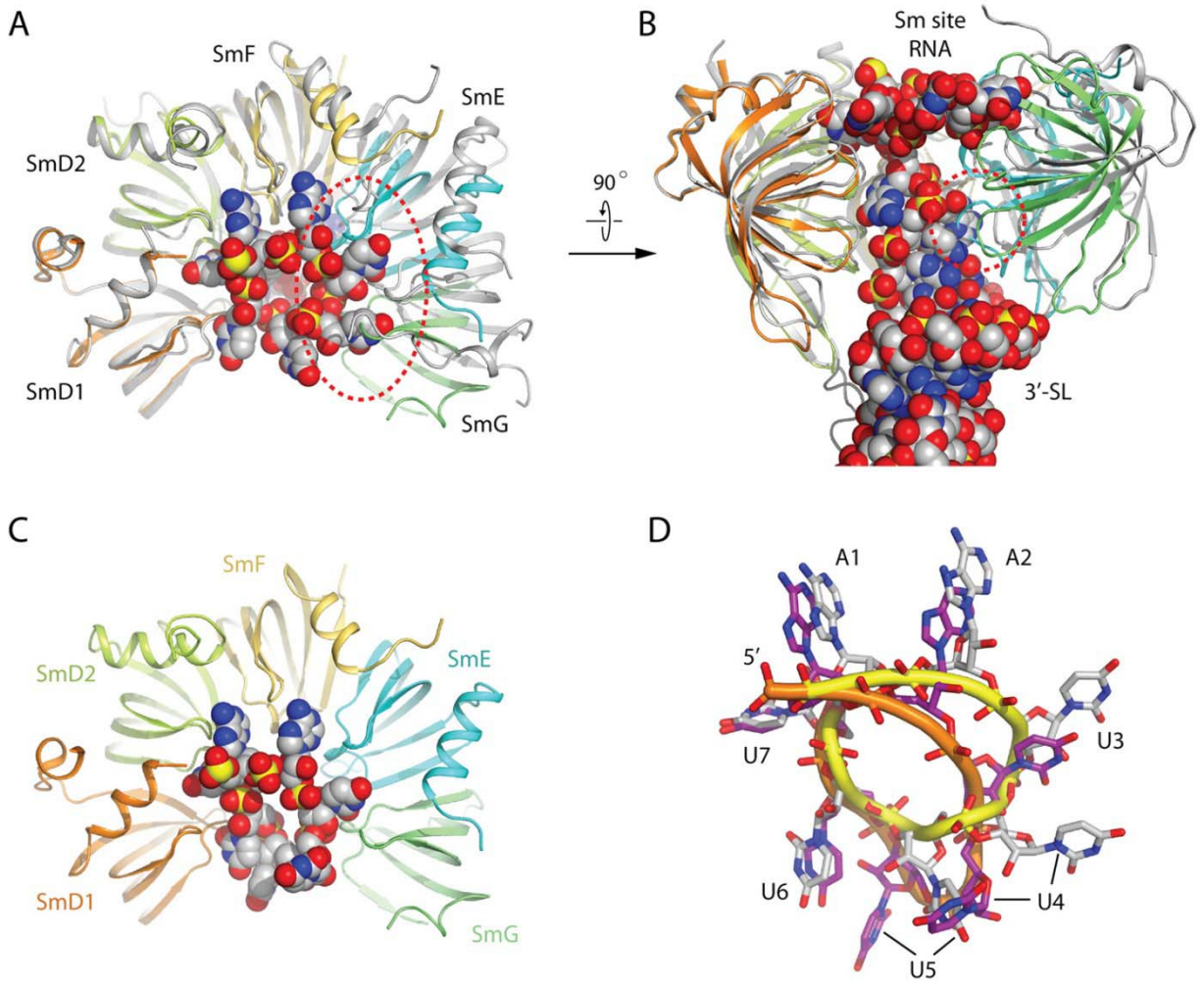


Figure 4. RNA-dependent widening of 5Sm. (A-B) Comparison of 5Sm in 7S intermediate (PDB code 5XJL) and Sm core (represented by U4 snRNP core, PDB code 4WZJ). SmD2 is used for superposition. The 5 Sm proteins in 7S are colored the same as in Figure 1. They are colored in light grey in Sm core. The RNA in Sm core is showed in sphere. Oxygen, nitrogen and phosphor are colored in red, blue and yellow respectively. The collided portions between the RNA in Sm core and the 5Sm in 7S are indicated by red dashed circles. (C) RNA Sm site model in 7S. (D) Comparison of RNA Sm site model in 7S (carbon is colored in pink and phosphate backbone in orange) with the Sm site in Sm core (carbon is colored in light grey and phosphate backbone in yellow). Numbering starts from the first nucleotide in Sm site.

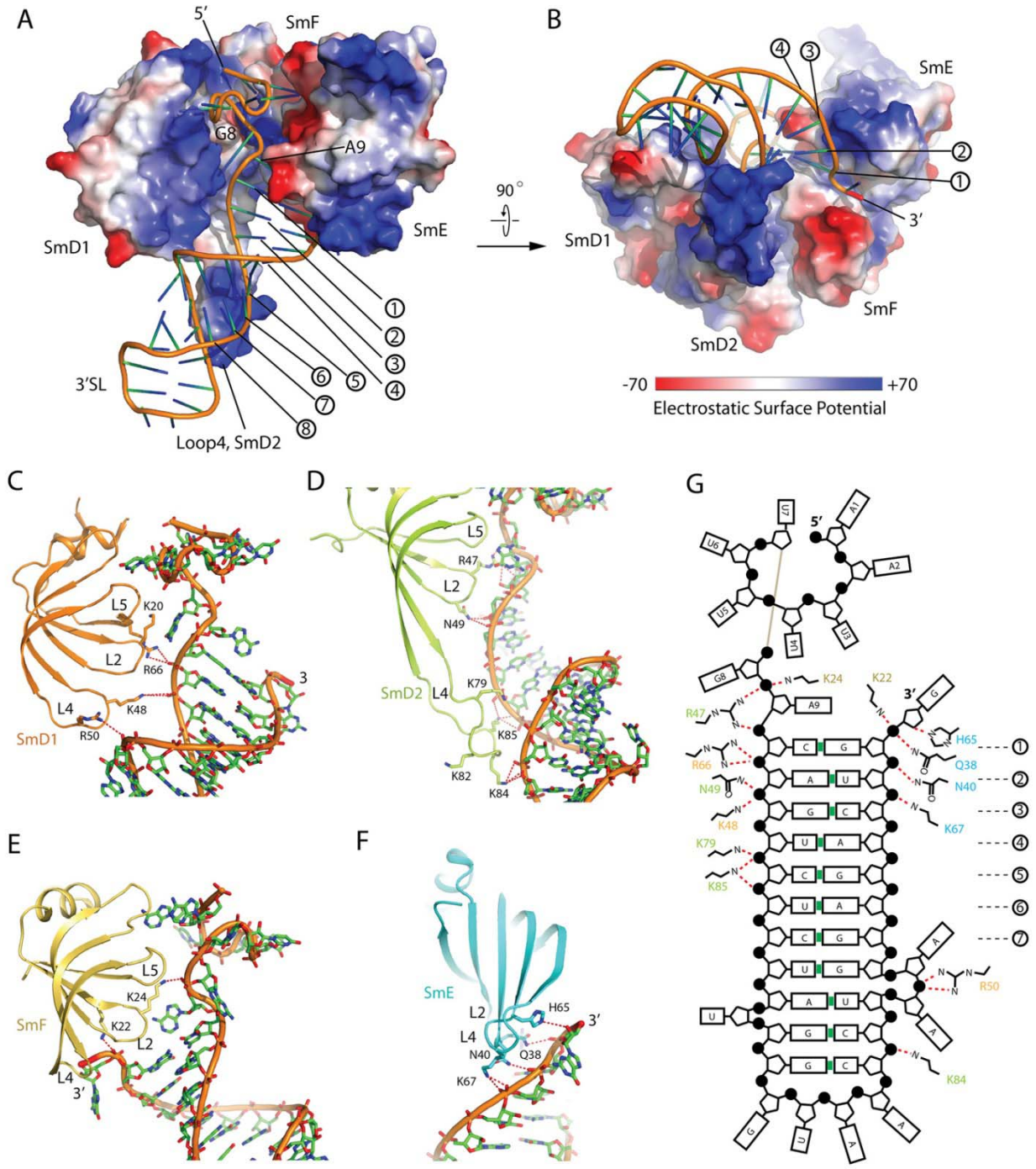
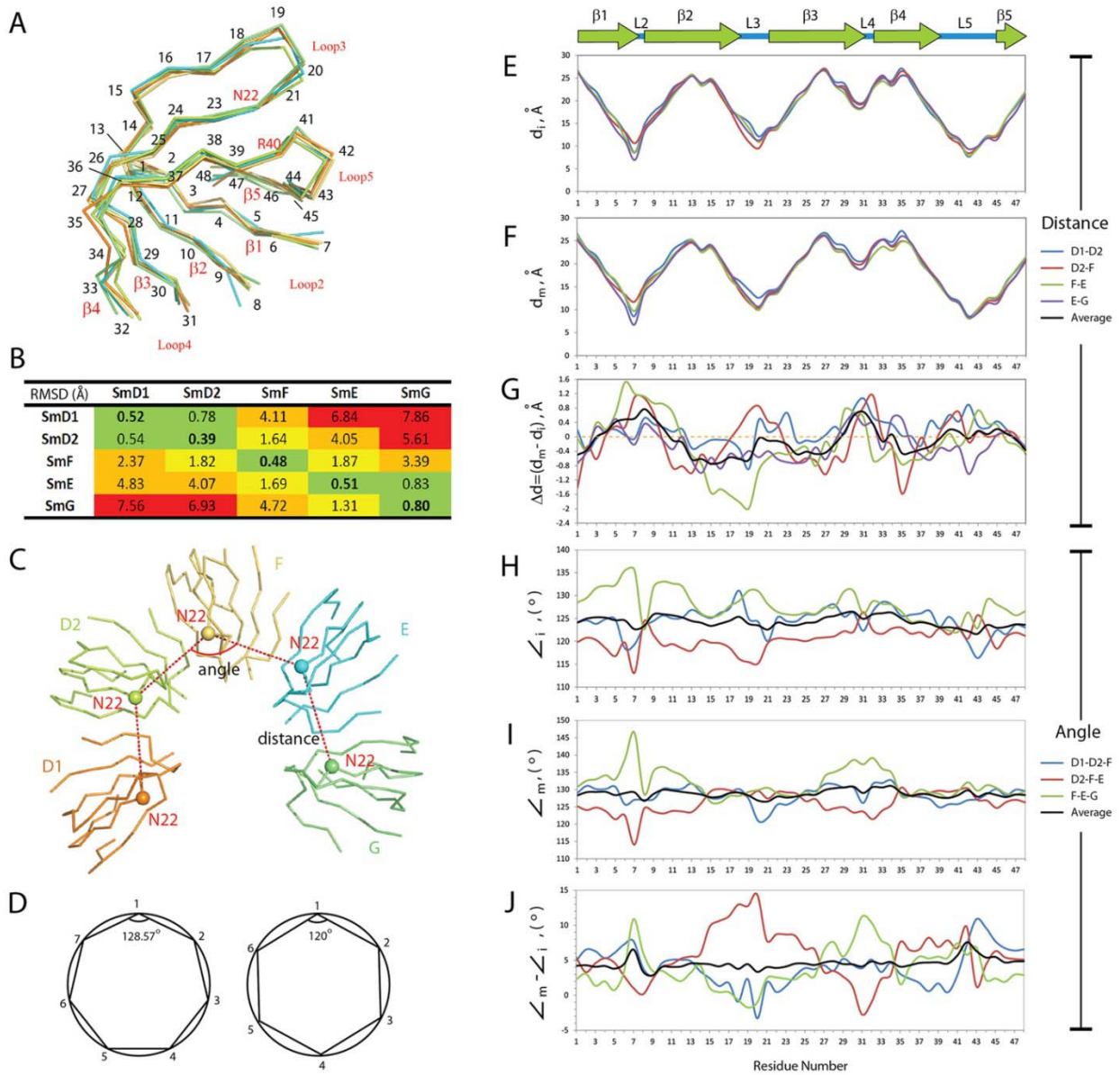


Figure 5. Interactions between Sm proteins and RNA 3'-SL are critical for RNA binding selectivity. (A-B) interactions of snRNP code with SmD1/D2/F/E in Sm core (PDB code 4wzj). Because of its little contact with 3'SL of RNA, SmG is not shown for clarify. Sm proteins are shown in surface view and electrostatic surface potential is shown. RNA is shown in cartoon. The base pairs directly after Sm site RNA are numbered. (C-F) Detailed interactions between each Sm protein and RNA 3'-SL. (G) Schematic summary of the interactions between Sm proteins and RNA 3'-SL.



RMSD (Å)	SmD1	SmD2	SmF	SmE	SmG
SmD1	**0.52**	0.78	4.11	6.84	7.86
SmD2	0.54	**0.39**	1.64	4.05	5.61
SmF	2.37	1.82	**0.48**	1.87	3.39
SmE	4.83	4.07	1.69	**0.51**	0.83
SmG	7.56	6.93	4.72	1.31	**0.80**

Figure 6. Conformational changes of 5Sm from the 7S intermediate state to assembled Sm cores. (A) The Sm fold defined by 48 residues of each Sm protein used for superposition and analysis. Numbering residues in this system is shown at each position. (B) Position differences of each pair of Sm proteins in 7S and U1 Sm core when one pair is superposed using the 48-residue Sm fold system. (C) Definition of the distance and angle used in analysis. (D) Standard angles for heptamer and hexamer. Distances between each pair of positions in adjacent Sm proteins in 7S, d_i (E), and in U1 Sm core, d_m (F), are plotted against the 48 positions. (G) Distance changes from 7S to Sm core, $\Delta d=d_m-d_i$, are plotted against the 48 positions. Angles of three adjacent Sm proteins in 7S, $\angle_i$ (H), and in U1 Sm core, $\angle_m$ (I), are plotted against the 48 positions. (J) Angle changes from 7S to Sm core, $\angle_m-\angle_i$, are plotted against the 48 positions. The secondary structures of the 48 positions are shown on the top of the plots.

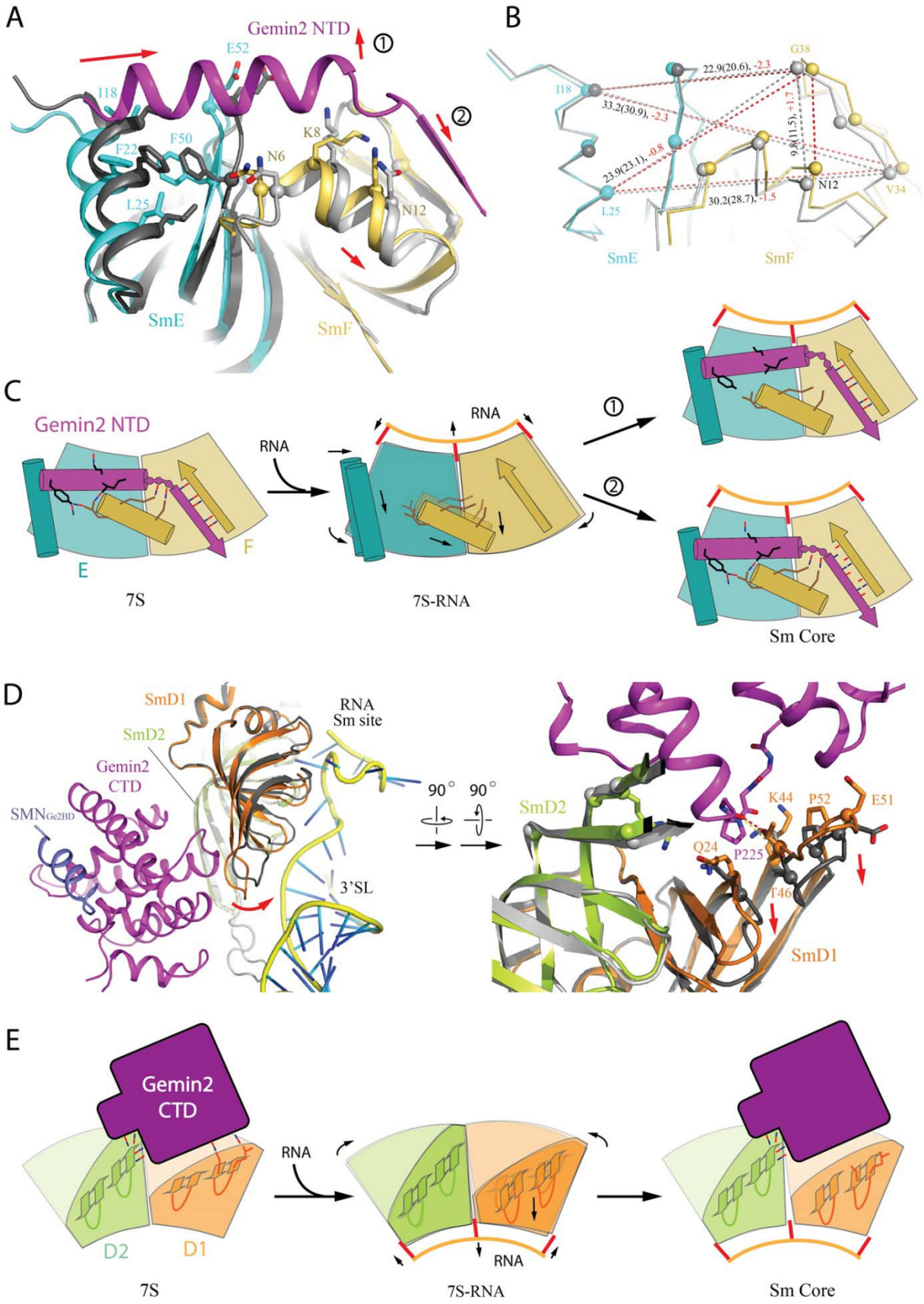

A
Gemin2 NTD
E52
I18
F22
F50
K8
N6
L25
N12
SmE
SmF
B
G38
I18
22.9(20.6), -2.3
33.2(30.9), -2.3
23.9(23.1), -0.8
9.8(11.5), +1.7
30.2(28.7), -1.5
N12
V34
L25
SmE
SmF
C
Gemin2 NTD
RNA
E
F
7S
7S-RNA
Sm Core
D
SmD1
SmD2
Gemin2 CTD
SMNGe2BD
RNA Sm site
3'SL
90°
90°
SmD2
K44
P52
E51
Q24
P225
T46
SmD1
E
Gemin2 CTD
RNA
D2
D1
7S
7S-RNA
Sm Core

Figure 7. Structural basis for Gemin2 release coupled with RNA-dependent 5Sm conformational changes. (A) Potential conformational change of Gemin2 NTD upon Sm core assembly. SmFs in 7S and Sm core are superposed based on the 48 positions defined previously. Gemin2 α1 would move towards SmF following the N-terminal helix of SmE as indicated by the red arrow. The second half of Gemin2 NTD would have two possible movements as indicated by red arrows labeled with ⍰and ⍰. (B) Distance changes between the indicated positions on Sm proteins from 7S to Sm core. SmEs in 7S and Sm core are superimposed based on the 48 positions defined early. Distances (Å) in 7S and in Sm core are shown in black with the latter in brackets, and distance changes are shown in red. (C) Schematic model of Gemin2 NTD reducing contacts with SmE/F. (D) Gemin2 CTD interactions change upon Sm core assembly. Superimposition of SmD2s of 7S and Sm core reveals a flipping of SmD1 loops 2 and 4 towards RNA 3'-SL (red arrows), which causes Gemin2 CTD to lose most contacts with SmD1. (E) Schematic model of Gemin2 CTD reducing contacts with SmD1/D2 upon Sm core assembly. Hydrogen bonds are shown in red (H-acceptor) and blue (H-donor) bar pairs in schematic models (C) and (E).

REFERENCES

1. Leung, A.K., K. Nagai, and J. Li, *Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis.* Nature, 2011. **473**(7348): p. 536-9.
2. Weber, G., et al., *Functional organization of the Sm core in the crystal structure of human U1 snRNP.* EMBO J, 2010. **29**(24): p. 4172-84.
3. Pomeranz Krummel, D.A., et al., *Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution.* Nature, 2009. **458**(7237): p. 475-80.
4. Kaida, D., et al., *U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation.* Nature, 2010. **468**(7324): p. 664-8.
5. Marzluff, W.F., E.J. Wagner, and R.J. Duronio, *Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail.* Nat Rev Genet, 2008. **9**(11): p. 843-54.
6. Azzouz, T.N., et al., *Toward an assembly line for U7 snRNPs: interactions of U7-specific Lsm proteins with PRMT5 and SMN complexes.* J Biol Chem, 2005. **280**(41): p. 34435-40.
7. Schumperli, D. and R.S. Pillai, *The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein.* Cell Mol Life Sci, 2004. **61**(19-20): p. 2560-70.
8. Tang, W., et al., *Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes.* Nature, 2012. **484**(7393): p. 260-4.
9. Seto, A.G., et al., *Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle.* Nature, 1999. **401**(6749): p. 177-80.
10. Lee, S.I. and J.A. Steitz, *Herpesvirus saimiri U RNAs are expressed and assembled into ribonucleoprotein particles in the absence of other viral genes.* J Virol, 1990. **64**(8): p. 3905-15.
11. Cazalla, D., T. Yario, and J.A. Steitz, *Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA.* Science, 2010. **328**(5985): p. 1563-6.
12. Golembe, T.J., J. Yong, and G. Dreyfuss, *Specific sequence features, recognized by the SMN complex, identify snRNAs and determine their fate as snRNPs.* Mol Cell Biol, 2005. **25**(24): p. 10989-1004.
13. Kolev, N.G. and J.A. Steitz, *In vivo assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site.* Nat Struct Mol Biol, 2006. **13**(4): p. 347-53.
14. Pillai, R.S., et al., *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, 2003. **17**(18): p. 2321-33.
15. Tisdale, S., et al., *SMN is essential for the biogenesis of U7 small nuclear ribonucleoprotein and 3'-end formation of histone mRNAs.* Cell Rep, 2013. **5**(5): p. 1187-95.
16. Raker, V.A., G. Plessel, and R. Luhrmann, *The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro.* EMBO J, 1996. **15**(9): p. 2256-69.
17. Raker, V.A., et al., *Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner.* Mol Cell Biol, 1999. **19**(10): p. 6554-65.
18. Friesen, W.J., et al., *The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins.* Mol Cell Biol, 2001. **21**(24): p. 8289-300.
19. Brahms, H., et al., *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, 2001. **7**(11): p. 1531-42.
20. Chari, A., et al., *An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs.* Cell, 2008. **135**(3): p. 497-509.

21. Pellizzoni, L., J. Yong, and G. Dreyfuss, *Essential role for the SMN complex in the specificity of snRNP assembly.* Science, 2002. **298**(5599): p. 1775-9.
22. Zhang, R., et al., *Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly.* Cell, 2011. **146**(3): p. 384-95.
23. Grimm, C., et al., *Structural basis of assembly chaperone- mediated snRNP formation.* Mol Cell, 2013. **49**(4): p. 692-703.
24. Sarachan, K.L., et al., *Solution structure of the core SMN-Gemin2 complex.* Biochem J, 2012. **445**(3): p. 361-70.
25. Otter, S., et al., *A comprehensive interaction map of the human survival of motor neuron (SMN) complex.* J Biol Chem, 2007. **282**(8): p. 5825-33.
26. Schrank, B., et al., *Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9920-5.
27. Jablonka, S., et al., *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, 2002. **99**(15): p. 10126-31.
28. Lefebvre, S., et al., *Identification and characterization of a spinal muscular atrophy-determining gene.* Cell, 1995. **80**(1): p. 155-65.
29. Fischer, U., Q. Liu, and G. Dreyfuss, *The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis.* Cell, 1997. **90**(6): p. 1023-9.
30. Burghes, A.H. and C.E. Beattie, *Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?* Nat Rev Neurosci, 2009. **10**(8): p. 597-609.
31. Carissimi, C., et al., *Gemin8 is required for the architecture and function of the survival motor neuron complex.* J Biol Chem, 2006. **281**(48): p. 37009-16.
32. Carissimi, C., et al., *Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly.* J Biol Chem, 2006. **281**(12): p. 8126-34.
33. Charroux, B., et al., *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, 1999. **147**(6): p. 1181-94.
34. Charroux, B., et al., *Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli.* J Cell Biol, 2000. **148**(6): p. 1177-86.
35. Yong, J., et al., *Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis.* Mol Cell, 2010. **38**(4): p. 551-62.
36. Battle, D.J., et al., *The Gemin5 protein of the SMN complex identifies snRNAs.* Mol Cell, 2006. **23**(2): p. 273-9.
37. Lau, C.K., J.L. Bachorik, and G. Dreyfuss, *Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains.* Nat Struct Mol Biol, 2009. **16**(5): p. 486-91.
38. Xu, C., et al., *Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly.* Genes Dev, 2016. **30**(21): p. 2376-2390.
39. Tang, X., et al., *Structural basis for specific recognition of pre-snRNA by Gemin5.* Cell Res, 2016. **26**(12): p. 1353-1356.
40. Jin, W., et al., *Structural basis for snRNA recognition by the double-WD40 repeat domain of Gemin5.* Genes Dev, 2016. **30**(21): p. 2391-2403.
41. Wahl, M.C. and U. Fischer, *The right pick: structural basis of snRNA selection by Gemin5.* Genes Dev, 2016. **30**(21): p. 2341-2344.
42. Kondo, Y., et al., *Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition.* Elife, 2015. **4**.
43. Li, J., et al., *Re-refinement of the spliceosomal U4 snRNP core-domain structure.* Acta Crystallogr D Struct Biol, 2016. **72**(Pt 1): p. 131-46.
44. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

45. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 355-67.
46. Potterton, E., et al., *A graphical user interface to the CCP4 program suite.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 7): p. 1131-7.
47. Schwer, B., A.J. Roth, and S. Shuman, *Will the circle be unbroken: specific mutations in the yeast Sm protein ring expose a requirement for assembly factor Brr1, a homolog of Gemin2.* RNA, 2017. **23**(3): p. 420-430.
48. Grimm, C., E. Lund, and J.E. Dahlberg, *In vivo selection of RNAs that localize in the nucleus.* EMBO J, 1997. **16**(4): p. 793-806.
49. Noble, S.M. and C. Guthrie, *Transcriptional pulse-chase analysis reveals a role for a novel snRNP-associated protein in the manufacture of spliceosomal snRNPs.* EMBO J, 1996. **15**(16): p. 4368-79.
50. Liu, Q., et al., *The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins.* Cell, 1997. **90**(6): p. 1013-21.
51. Barbarossa, A., et al., *Characterization and in vivo functional analysis of the Schizosaccharomyces pombe ICLN gene.* Mol Cell Biol, 2014. **34**(4): p. 595-605.
52. Hannus, S., et al., *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, 2000. **9**(5): p. 663-74.
53. Owen, N., et al., *Characterization of the Schizosaccharomyces pombe orthologue of the human survival motor neuron (SMN) protein.* Hum Mol Genet, 2000. **9**(5): p. 675-84.
54. Paushkin, S., et al., *The survival motor neuron protein of Schizosacharomyces pombe. Conservation of survival motor neuron interaction domains in divergent organisms.* J Biol Chem, 2000. **275**(31): p. 23841-6.
55. Kroiss, M., et al., *Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster.* Proc Natl Acad Sci U S A, 2008. **105**(29): p. 10045-50.
56. Meister, G., et al., *A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs.* Nat Cell Biol, 2001. **3**(11): p. 945-9.
57. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
58. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W320-4.