Nucleic Acids Research, 2025, **53**, gkaf451
https://doi.org/10.1093/nar/gkaf451
RNA and RNA-protein complexes

Interconversion and mechanisms between Lsm-type and Sm-type heteroheptameric rings: implications for spliceosome evolution and RNA metabolism

Li Mu[1,2,†], Yan Hou[1,2,†], Yan Hu[1,2], Yingzhi Wang[1,2], Congcong Shen[1,2], Yongbo Luo[2], Dan Su[2], Rundong Zhang[1,2,*]

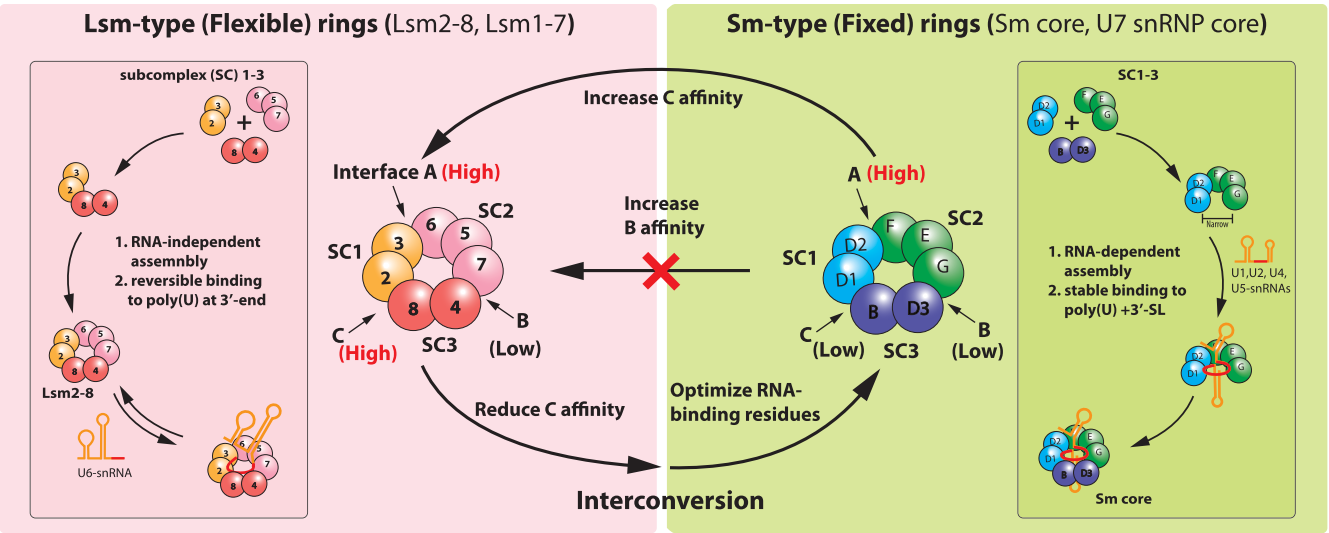
[1]Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu 610041, P.R. China
[2]State Key Laboratory of Biotherapy and Collaborative Innovation Center for Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, P.R. China

[*]To whom correspondence should be addressed. Email: rundongzhang@scu.edu.cn
[†]These authors contributed equally.

Abstract

Eukaryotes harbor both Sm-type and Lsm-type heteroheptameric rings, which are essential in RNA metabolism. Despite their similar subunits and evolutionary ties, they interact with RNA in distinct ways, functioning as scaffolds and chaperones, respectively. However, the mechanistic basis of their evolutionary divergence remains unclear. Using the Sm ring (D1-D2-F-E-G-D3-B) and the Lsm2-8 ring, both of which form the cores of distinct spliceosomal snRNPs, as model systems, we investigated the feasibility and mechanisms of their interconversion. We found that the interactions among subcomplexes (SCs) 1–3 in the Sm ring (D1/D2, F/E/G, and D3/B) differ from those in Lsm2–8 (Lsm2/3, Lsm6/5/7, and Lsm8/4), implying the formation of distinct assembly intermediates. By strengthening the SC1–SC3 interaction, we achieved the conversion of the Sm ring into an Lsm-type ring. Conversely, increasing the SC2–SC3 affinity did not yield a successful conversion. Furthermore, by weakening the SC1–SC3 interaction and introducing mutations in the RNA-binding regions of SC1 and SC2, we converted Lsm2-8 into a Sm-type ring. These findings provide mechanistic insights into how similar protein components can assemble into functionally distinct heteroheptameric rings, a principle likely applicable to Lsm1–7 and the U7 snRNP core, and offer deep insights into spliceosome and eukaryotic evolution.

Graphical abstract



Introduction

Investigating the formation of intricate protein complexes from homologous proteins that have diverged from common ancestral proteins through gene duplication and mutation is pivotal in understanding molecular evolution. Hemoglobin, microtubules, and the 20S core proteasome (CP) [1] are

Received: August 13, 2024. **Revised:** April 4, 2025. **Editorial Decision:** April 24, 2025. **Accepted:** May 17, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

representative examples of globular, fibrous, and hollow cylindrical complexes, respectively. Protein complexes formed from Sm or like-Sm (Lsm) proteins are also remarkable examples [2–4].

These proteins share an evolutionarily conserved Sm fold, characterized by an N-terminal α-helix and five highly bent antiparallel β-strands. Adjacent β4 and β5 strands from different (L)Sm proteins interact, constructing homo- or heteropolymeric ring complexes that play a vital role in RNA metabolism across all domains of life [5–13]. These (L)Sm rings bind to the U-rich sequence of RNA within their central channel on one face, with each protomer specifically interacting with each uridine nucleotide. Each (L)Sm protein employs two conserved motifs—the DXΦXN (Φ refers to an aromatic residue) motif in loop 3 and the RG motif in loop 5—to recognize uracil. Residues Φ and R sandwich uracil through π–π and cation–π interactions, respectively; the invariant residue N provides specificity by forming hydrogen bonds with the uracil [7, 8, 11, 14–18].

Bacteria possess a single Lsm protein, Hfq, forming homohexameric rings [12]. In archaea, typically one or two Lsm proteins assemble into homohexameric or homoheptameric rings [13]. Eukaryotes exhibit much more complexity, with at least 14 distinct (L)Sm proteins arranging into two different heteroheptameric rings. The canonical Sm ring (Fig. 1A), composed of D1-D2-F-E-G-D3-B, is integral to the structure of U1, U2, U4, and U5-snRNPs in the spliceosome, while the Lsm2–8 ring (Fig. 1A) associates with the 3′ end of U6 snRNA. Significant in spliceosome function, these rings contribute to pre-mRNA processing. Many eukaryotes, especially metazoans, possess two additional heteroheptameric rings: the Lsm1–7 (Fig. 1A), which is involved in cytoplasmic RNA degradation [19, 20], and the U7 snRNP core (Fig. 1A), which is crucial for histone pre-mRNA 3′ processing [21, 22]. Additionally, Sm/Lsm rings contribute to telomere maintenance [23, 24].

These rings can be categorized into two types based on their RNA-binding properties and assembly requirements: Lsm-type (flexible) and Sm-type (fixed) rings [2]. Lsm-type rings engage in a dynamic association with RNA containing U-rich sequences at the 3′ end, with the 3′ U-rich segment being encircled by one face of the ring's central channel [8, 15]. They can readily dissociate and function as RNA chaperones (Fig. 1B). Notably, Lsm-type rings assemble independently of RNA and other factors, and they include Lsm2–8, Lsm1–7, and all Lsm homomeric rings in bacteria and archaea. In contrast, Sm-type rings bind RNA containing the snRNP code—a conserved U-rich segment termed the Sm site (typically PuAUUUNUGPu, where Pu is purine) followed by a stem-loop structure [7]—and the bound RNA passes through the central channel [6, 7, 16–18]. They maintain this interaction and serve as a scaffold for RNA (Fig. 1C). The assembly of Sm-type rings is RNA-dependent and requires the assistance of chaperones such as the PRMT5 and SMN–Gemins complexes in almost all eukaryotes [21, 25–31]. Sm-type rings include the canonical spliceosomal Sm ring and the U7-snRNP core.

Phylogenetic analyses indicate that all (L)Sm rings share a common ancestor, and it is hypothesized that eukaryotic diversification involved two principal "duplication and mutation" events among (L)Sm proteins [2–4]. The initial event spawned a flexible heteroheptameric ring (Lsm2–8) from a homooligomeric Lsm ring, followed by a subsequent event transforming this into a fixed ring—forming the canonical Sm spliceosomal ring. These rings form the core of spliceosomal snRNPs, playing a pivotal role in early spliceosome assembly and function during eukaryotic evolution [32]. While the Lsm2–8 ring still retains the chaperone activity as do the homooligomeric Lsm rings in bacteria and archaea, the emergence of the Sm core is significantly distinct from Lsm rings, for it acts as a scaffold for snRNAs, stabilizing snRNAs [33], facilitating the recruitment of additional proteins (i.e. U1-70K and U1-C in U1-snRNP) [6, 17], and shaping the modern spliceosome's architecture.

Remarkably, flexible and fixed heteroheptameric rings share similar components yet possess divergent assemblies and functionalities. Comparative phylogenetic and structural analyses have shown that the seven Lsm proteins correspond precisely to the seven Sm proteins, in the order of Lsm 2–3-6–5-7–4-8 matching Sm D1-D2-F-E-G-D3-B [2–4, 6–8]. Moreover, each set of these proteins forms three distinct subcomplexes (SCs) (Fig. 1A), which exhibit greater stability than individual subunits [34, 35]. The assembly of the Sm ring can occur spontaneously *in vitro* and follows a specific pathway, even *in vivo*, where assembly is generally chaperone-dependent: first, SmD1/D2 and SmF/E/G form a metastable pentamer, then snRNA incorporation leads to a stable subcore, and ultimately SmD3/B joins to complete assembly (Fig. 1C) [28, 34, 36]. Although similar pathways (Lsm2/3 and Lsm6/5/7 form an intermediate complex first) are proposed for Lsm2–8 and Lsm1–7 complexes [37, 38], this raises questions about how such structurally similar proteins can produce distinct functional complexes upon assembly. Therefore, the assembly mechanisms of these heteroheptameric rings remained obscure. Moreover, experimental conversions between Lsm-type and Sm-type rings have not been demonstrated previously.

In this study, we investigated SC interactions and postulated alternate assembly intermediates for Lsm2–8 and the Sm ring. Successful interconversion of the Sm ring to an Lsm-type and Lsm2-8 to a Sm-type ring was achieved, unveiling underlying principles guiding the formation of these rings. These findings suggest a model for the evolutionary transition from an Lsm-type to a Sm-type ring, contributing to our understanding of spliceosome evolution and eukaryogenesis. Moreover, they offer insights into the evolution of other heteromeric complexes and the potential design of protein complexes.

Materials and methods

Plasmid constructions for protein expression in *E. coli*

All of the plasmids used in this study contain human complementary DNAs (cDNAs). (L)Sm proteins in the same dimeric SC were constructed in a single pCDFDuet (or pETDuet or pRSFDuet) vector (Novagen), while in the cases of trimeric SC (SmF/E/G and Lsm6/5/7), two (L)Sm proteins (in the middle and at one side) were constructed in a single pCDFDuet vector and the third one in a separate vector. Expression plasmids containing full-length SmD1 and a N-terminal His_6 tag followed by a TEV cleavage site fused to SmD2 (pCDFDuet-HT-D2-D1, HT: His_6-TEV tag), and the Sm fold portions of SmD3 (residues 1–75) and SmB (residues 1–91) [pCDFDuet-HT-B(1–91)-D3(1–75)] were constructed before [27]. The full-length SmF, SmE, and HT-tagged SmG were constructed in the single pRSF vector (pRSF-F-E-HT-G). Expression plasmids containing full-length Lsm2 and Lsm3 (pETDuet-HT-Lsm3-Lsm2), full-length Lsm4 and Lsm8

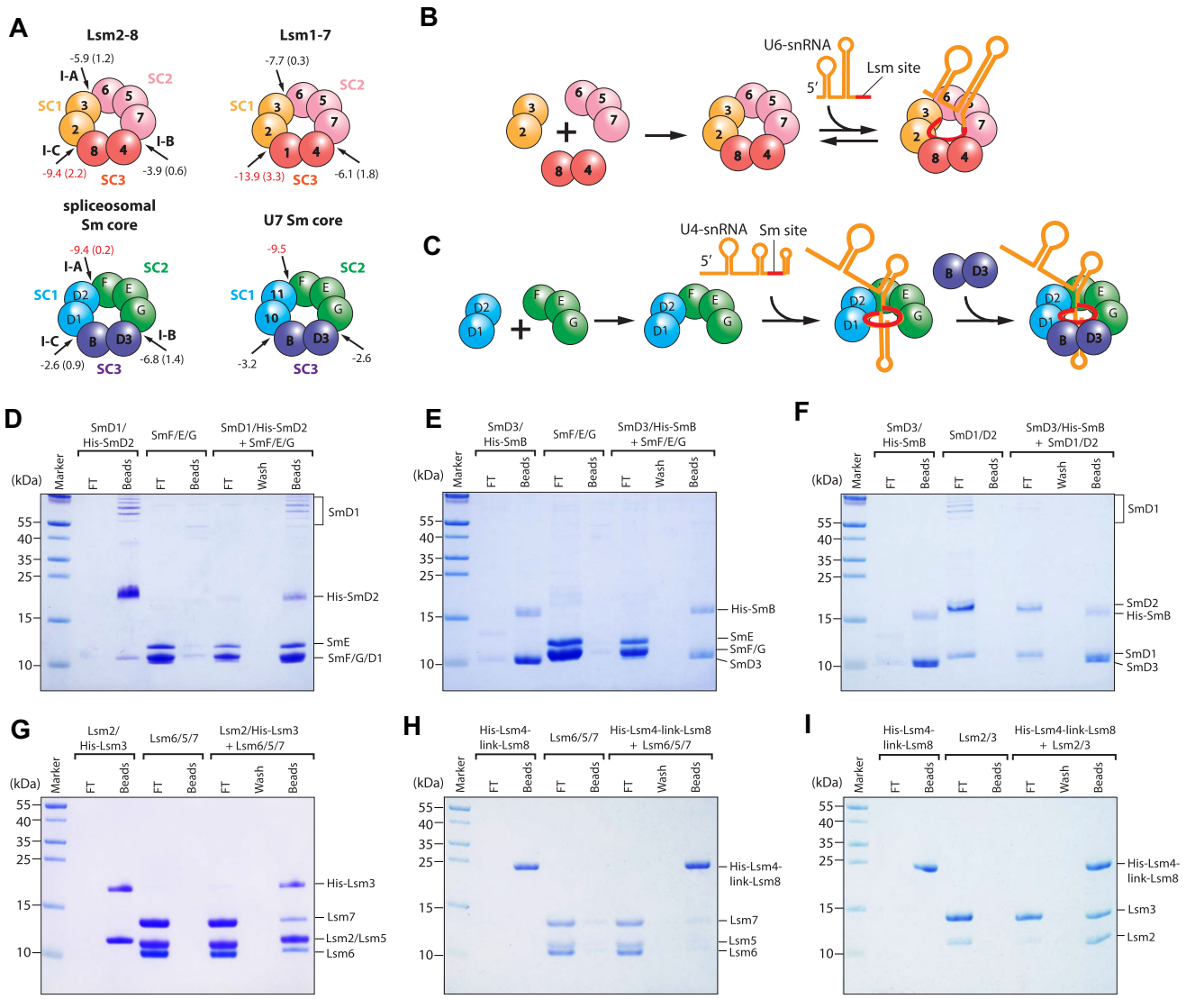


Figure 1. Two types of (L)Sm heteroheptameric rings exhibit distinct interactions between SCs. (**A**) Four (L)Sm rings can be grouped into two types: Lsm-type (three SCs are colored in orange, pink, and red) and Sm-type (three SCs are colored in cyan, green, and purple). Binding energies between SCs are labeled (with standard deviations in brackets) (see more in Supplementary Table S1). (**B**) Assembly of Lsm2–8 and its interaction with U6-snRNA. (**C**) Assembly pathway of spliceosomal Sm core with U4-snRNA. (**D–F**) Interactions between each pair of the 3 SCs in the Sm ring were tested using pull-down assay, in which a His-tagged SC, a tag-removed SC or their mixture was incubated with Ni-beads, washed, and analyzed by SDS–PAGE staining. (**G–I**) Interactions between each pair of the 3 SCs in Lsm2–8 were tested using the same pull-down assay as described above. Lsm4/8 was replaced by Lsm4-link-Lsm8 for better expression and purification (Supplementary Fig. S1). For each in (D–I), one representative result of at least two independent experiments is shown. The results of the inverse pull-down order of (D–I) are in Supplementary Fig. S2. FT: Flow-through.

(pETDuet-HT-Lsm8-Lsm4), full-length Lsm7 and Lsm5 (pETDuet-HT-Lsm7-Lsm5), and full-length Lsm6 (pCDF-HT-Lsm6) were constructed in the way similar to the above Sm proteins. Because the initial expression and purification trial of pETDuet-HT-Lsm8-Lsm4 generated very unbalanced ratio of Lsm8:Lsm4, a new set of four plasmids containing Lsm4 and Lsm8 fused in a tail-to-head way [pCDFDuet-HT-Lsm4(1–87)-link-Lsm8(1–96), pCDFDuet-HT-Lsm4(1–96)-link-Lsm8(1–96), pCDFDuet-HT-Lsm4(1–87)-link-Lsm8(1–80), and pCDFDuet-HT-Lsm4(1–96)-link-Lsm8(1–80)] were constructed and tested. The first of the four plasmids was finally used to express the fused form of Lsm4-link-Lsm8.

All the (L)Sm mutants described in the manuscript, including $SmD1_H/D2$, $SmD3/B_{H1}$, $SmD3/B_{H2}$, $SmF/E/G_H$, $SmD3_H/B$, $Lsm2_L/Lsm3$, $Lsm4/Lsm8_M$, $SmD3_M/B$, $SmF/E/G_M$, $Lsm2/Lsm3_M$, $Lsm6_M$, $Lsm5_{M1}/Lsm7$, and $Lsm5_{M2}/Lsm7_M$, were constructed similar to their corresponding wild-types. Gemin2ΔN39 (pCDF-HT-Gemin2ΔN39) were constructed before [27]. SMN_{Ge2BD}, containing SMN residues 26–62 (pET21-HMT-SMN_{Ge2BD}), was fused with an N-terminal His_6-tag followed by maltose-binding protein (MBP) tag and TEV cleavage site in pET21 vector (Novagen). All the constructs were verified by DNA sequencing.

Protein expression and purification

All the proteins and protein complexes were expressed using BL21(DE3) competent cells under the induction of isopropyl β-D-thiogalactoside (IPTG). For SmF/E/G and Lsm6/5/7 and their mutants, SmF/E and SmG (or Lsm5/7 and Lsm6, or their mutants) were co-expressed. The purification procedure

followed a very similar way to our previous methods [28]. Briefly, in general, proteins or protein complexes were purified by Ni-column first, ion-exchange chromatography second and gel filtration chromatography (GFC) at last. If the His_6-TEV-tag needed to remove, TEV protease cleavage was performed after Ni-column purification, and the tag-removed proteins were collected by passing Ni-column again at low concentration of imidazole in the buffer. Throughout the purification process, the quality and quantity of the proteins were monitored by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) and Coomassie brilliant blue (CBB) staining. Finally, the purified proteins and protein complexes were either directly used or snap-frozen in liquid nitrogen and stored at -80°C for future use.

Protein–protein interaction by pull-down assay

In some cases such as for the interactions between each pair of Sm SCs, Ni-NTA agarose beads were washed by ultrapure water and equilibrated with Buffer A (50 mM Tris–HCl, pH 8.0, 250 mM NaCl, 20 mM imidazole, and 5% glycerol). The above beads were aliquoted 150 μl each into Eppendorf tubes, and each tube was incubated with 9 nmol of His-tagged Sm SC, tag-removed Sm SC or both in Buffer A for 1 h at 4°C. The supernatants were collected after centrifugation (500 × *g*, 3 min), and the beads were then washed with 300 μl Buffer A, five times. After washing, the proteins bound on beads were resolved by SDS–PAGE and CBB staining. For the interaction between each pair of Lsm SCs, the procedure was similar to Sm SCs except that Lsm SCs were pre-incubated with Buffer A plus 5 M urea at 16°C overnight before incubation with Ni-beads to break the potential stable high-order oligomeric state of each SC formed by self-interaction as described before [35]. In other cases in which one protein complex was tested to interact with several other protein complexes, such as for interactions between HT-tagged Gemin2/MBP-SMN(26–62) with SmF/E/G, SmD1/D2 and SmF/E/G, Sm core, and so on, the first step took a variation while the remaining steps kept the same. The variation is, Ni-NTA agarose beads were incubated with HT-tagged protein first, and then aliquoted into different tubes, in which different tag-removed protein complexes were added for interaction.

Protein complexation by GFC

To estimate the oligomeric state of a protein complex, analytical GFC was used. The purified proteins or protein complexes were mixed together at an equimolar ratio and incubated at 4°C for a minimum of 1 h to allow complex formation. After the incubation, the mixture was centrifuged at 13 000 × *g* at 4°C for 15 min to remove any debris and collect the supernatant. The supernatant was then passed through a GFC column called Superdex 200 Increase 10/300 GL (GE Healthcare, USA) with the running buffer (25 mM Tris–HCl, 250 mM NaCl, and 1 mM EDTA, pH 7.5). The eluted fractions were collected at intervals of 0.5 ml and analyzed by SDS–PAGE and CBB or silver staining. The apparent molecular weights (aMWs) of the protein complexes were calculated on the basis of the GFC protein standards (Bio-Rad, USA) as a control using the equation: $aMW = 10^{(-0.1976*Vol +4.7009)}$ [$R^2 = 0.999$, aMW(kDa), Vol (ml)]. The theoretical MWs and aMWs of the proteins and complexes in this study can be found in Supplementary Table S2.

In vitro RNA production and purification

The 30-nt RNAs (U4-5′Δ and U4-5′Δ-ΔSm) were chemically synthesized by Takara. The long RNAs, including U1, U2, U4, U5, and U6 snRNAs, were produced using *in vitro* transcription with the T7 High Yield RNA Transcription Kit (Vazyme, China). The process was similar to that described before [28]. The DNA templates were made by polymerase chain reaction (PCR). Transcribed RNAs were purified by GFC (Superdex 200 Increase 10/300 GL) in the running buffer containing 20 mM TrisvHCl, 250 mM NaCl, 2 mM $MgCl_2$, and 1 mM EDTA, pH 8.0. The qualities of these RNAs were further checked on agarose gel electrophoresis. The sequences of the RNAs used in this study can be found in Supplementary Table S3.

In vitro RNA–protein complex assembly assay

RNA–protein complex assembly assays were performed by incubating protein complexes with RNA in final volume of 400 μl in assembly buffer containing 20 mM Tris–HCl (pH 7.5), 250 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, and 1 mM DTT, with equal molar ratio. RNAs were pre-incubated at 65°C for 10 min followed by cool-down at room temperature before mixing with proteins. After incubation at 37°C for 40 min, the samples were collected at 15 000 rpm for 5 min in a table centrifuge and applied into Superdex 200 Increase 10/300 GL (GE Healthcare, USA) GFC via a 500 μl of sample loop. The elution was monitored by OD_{280nm} and OD_{260nm}. The eluted fractions were collected each 0.5 ml, resolved by SDS–PAGE, and visualized by silver staining. In certain situations, SDS–PAGE was also stained by nucleic acids dye, Super GelRed (US Everbrigh), and visualized under UV to identify the positions of RNAs.

Electron microscopy analysis

For electron microscopy (EM) analysis of GFC-purified protein complex 7Sm-LLH or RNA–Lsm complex, negative staining with uranyl formate was performed according to a previous protocol [39]. Preparations were examined with a JEM-1400Flash electron microscope operating with 120 kV. Electron micrographs were taken at a magnification of 150 000. See more details in the Supplementary File.

Results

Distinct subcomplex interactions in Sm and Lsm rings

To elucidate the unique assembly of seven structurally similar components into two distinct ring formations, we investigated the binding affinities and assembly sequences of the SCs, which have not been systematically studied before. We took two different approaches. Initially, we utilized the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/) to assess binding energies at the interfaces of SCs, mapped from crystallographic data. The interfaces of the Sm core, denoted as A-C (I-A, I-B, and I-C) between SCs, exhibited binding energies of -9.4 ± 0.2, -6.8 ± 1.4, and -2.9 ± 0.9 kcal/mol, respectively, suggesting an affinity hierarchy of I-A > I-B > I-C (Fig. 1A and Supplementary Table S1). Conversely, the Lsm2–8 interfaces demonstrated a different affinity order of I-C > I-A > I-B, with energies of -5.9 ± 1.2, -3.9 ± 0.6, and -9.4 ± 2.2 kcal/mol, respectively. This suggests the existence of distinct assembly intermediates, specifically SC1/SC2

(comprising SmD1/D2/F/E/G) during Sm core formation and SC1/SC3 (consisting of Lsm4/8/2/3) during Lsm2–8 assembly (Fig. 1A and Supplementary Table S1). The affinity order in Lsm1−7 (I-C > I-A > I-B) is similar to that in Lsm2-8 (Fig. 1A and Supplementary Table S1).

Next, considering the potential inaccuracy of theoretical computations, we expressed and purified the six SCs from both the human Sm core and Lsm2–8 for binding testing. The formation of these six individual SCs has been demonstrated in previous studies [5, 26, 27, 35]. However, in our initial trial, we encountered a significant imbalance in the co-expression of Lsm4 and Lsm8. To improve expression and facilitate biochemical testing, we therefore used a linked Lsm4/8 construct, a strategy similar to a previous study of Lsm4/1, Lsm2/3, and Lsm6/5 [37]. As demonstrated in Supplementary Fig. S1, this construct had no adverse effect on Lsm2–8 assembly. There are two reasons for choosing human versions: (i) individual Sm and Lsm proteins are highly conserved across various species (i.e. seven Sm proteins share 48%–79% identity and 70%–88% similarity between humans and *Schizosaccharomyces pombe*, and Lsm2–Lsm8 share 46%–66% identity and 73%–90% similarity between humans and *S. pombe*); (ii) human Sm and Lsm proteins have been successfully expressed in *Escherichia coli* and biochemically characterized to be functional [5, 26, 27, 35].

Using pull-down assays, we detected a tight interaction between SmD1/D2 and SmF/E/G (Fig. 1D), but little interaction between SmF/E/G and SmD3/B (Fig. 1E) or between SmD1/D2 and SmD3/B (Fig. 1F). This is consistent with the above affinity analysis of Sm core and the assembly order studied previously in humans and other eukaryotes [28, 34, 36]. We detected a strong interaction between Lsm 2/3 and Lsm6/5/7 (Fig. 1G), which was not well predicted by the above affinity analysis but is consistent with the experimental results obtained with yeast homologs [37, 38]. We observed little interaction between Lsm4/8 and Lsm6/5/7 (Fig. 1H), and a strong interaction between Lsm2/3 and Lsm4/8 (Fig. 1I), as predicted by the above affinity analysis. However, the strong interaction between Lsm2/3 and Lsm4/8 had not been studied biochemically before. Moreover, pull-down assays conducted in a reversed order resulted in the same conclusions (Supplementary Fig. S2). These results indicate that the assembly orders of Sm and Lsm rings differ.

Sm core converts into an Lsm-type ring by enhancing the SC1–SC3 interaction

Considering that Lsm-type rings may form due to high-affinity interactions at two of the three interfaces, we postulated that enhancing either interface B or C could convert the Sm ring to an Lsm-type ring. We first attempted to increase the affinity of interface C by engineering SmB and SmD1. We made two SmB mutants, SmB_{H1} and SmB_{H2} (chimeras with Lsm8 by replacing the N- and C-termini, the subscript H means high affinity) and a SmD1 mutant, $SmD1_H$ (a chimera with Lsm2 by replacing the middle segment) (Fig. 2A and Supplementary Fig. S3). All three mutant SCs, $SmD1_H$/D2, SmD3/B_{H1}, and SmD3/B_{H2}, were expressed and behaved well (Fig. 2B). As expected, in pull-down assays, His-tagged $SmD1_H$/D2 retained both SmD3/B_{H1} and SmD3/B_{H2} whereas His-tagged wild-type SmD1/D2 did not retain SmD3/B (Fig. 2B).

To assess if an Lsm-type ring forms, we used several approaches, including pull-down assays, GFC, dynamic light scattering (DLS), and mass spectrometry (MS). His-tagged $SmD1_H$/D2 pulled down SmF/E/G and SmD3/B_{H1} together, or SmF/E/G and SmD3/B_{H2} together (Fig. 2C); however, it only pulled down SmF/E/G and not wild-type SmD3/B in the control (Fig. 2C). The mixture of SmD3/B_{H1}, $SmD1_H$/D2, and Sm F/E/G eluted from GFC as a single peak at 14.47 ml (Fig. 2D), corresponding to an aMW of 69.4 kDa, which is close to the MW of 74.6 kDa calculated from a Sm heteroheptamer (named 7Sm-LLH, where LLH means "like-Lsm2-8 heptamer") (Supplementary Table S2). Reconstitution of 7Sm-LLH under different conditions (either incubated in regular buffer or regular buffer containing 5 M urea before GFC analysis) exhibited comparable GFC profiles (Supplementary Fig. S4). SDS–PAGE analysis of peak fractions revealed stoichiometric ratios of all Sm proteins (Fig. 2D and Supplementary Fig. S4). DLS characterization of reconstituted 7Sm-LLH at 4 and 20°C yielded diameters of 7.8–8.1 nm (Supplementary Fig. S5), matching the ∼80 Å dimensions of Sm/Lsm rings. DLS-estimated MWs were 80.2 ± 20.1 kDa (4°C) and 89 ± 18.8 kDa (20°C) (Supplementary Fig. S5), consistent with the calculated MW of 7Sm-LLH. Liquid chromatography (LC)-MS/MS analysis confirmed the presence of all seven Sm proteins, with identification scores, unique sequence coverage, and intensities significantly exceeding other candidates (Supplementary Table S4).

To detect if the reconstituted 7Sm-LLH is a ring-shaped complex, we employed negative staining EM. Indeed, we observed that most particles exhibited a doughnut-like shape with a diameter of ∼80 Å and central holes with a diameter of ∼20 Å (Fig. 2E and Supplementary Fig. S6). These dimensions are consistent with the diameters calculated from the Sm rings of U1 and U4 crystal structures [6, 7] and Lsm2–8 [8]. Additionally, we adopted a biochemical approach to probe the shape of 7Sm-LLH. As Gemin2 and snRNA exhibit negative cooperativity in binding to the horseshoe-shaped SmD1/D2/F/E/G (5Sm), snRNA binding to the interior of 5Sm causes it to splay open and reduce its curvature. This allows SmD3/B to join and subsequently reduces the affinity of Gemin2 for binding to 5Sm, leading to Gemin2's release [27]. Therefore, the mature Sm core cannot bind Gemin2, and neither can 7Sm-LLH, which is expected to adopt the same splayed conformation in its 5Sm component. To test if 7Sm-LLH binds to Gemin2/SMN(26–62), we conducted a pull-down assay. Similarly to pulling down wild-type SmD1/D2/F/E/G, His-tagged Gemin2/SMN(26–62) successfully pulled down $SmD1_H$/D2 and SmF/E/G together. However, it did not pull down either 7Sm-LLH or the mature Sm core (Fig. 2F, Sm core assembly on U4-snRNA is shown in Supplementary Fig. S7). These results further confirm the ring conformation of the assembled 7Sm-LLH.

As mentioned earlier, there is a significant difference in RNA interactions between the Lsm-type and Sm-type rings. The Lsm2–8 complex binds specifically to the 3′-Lsm site (AUAUUUU) of U6 snRNA but not to the snRNP code, a sequence motif comprising the Sm site and an adjacent 3′ stem-loop shared by other spliceosomal snRNAs. We next examined whether 7Sm-LLH could function in a similar way. First, we used GFC to test the binding of 7Sm-LLH to either U6 snRNA or a control U6ΔLsm snRNA (in which AUAUUUU was replaced with AUCCCCC). Although both experiments yielded similar OD traces with two peaks (∼11.5 and ∼14.5 ml, corresponding to RNA and 7Sm-LLH alone,

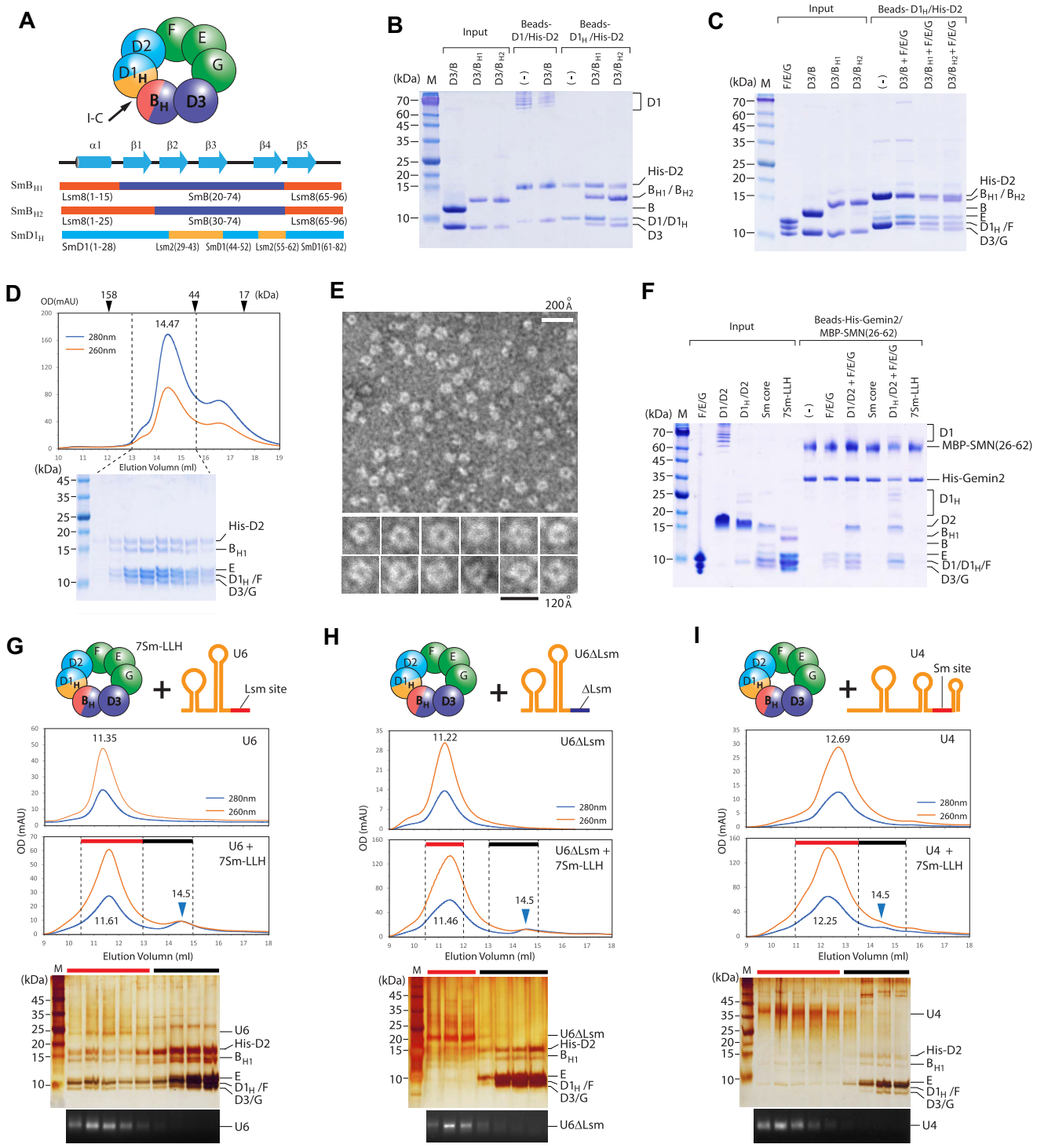


Figure 2. Increasing affinity between SC1 and SC3 in the Sm ring converts it into an Lsm-type ring. (**A**) The strategy to increase affinity between SC1 and SC3. The interface of SmD1 is replaced by that of Lsm2 ($SmD1_H$) and the interface of SmB by that of Lsm8 (SmB_{H1} and SmB_{H2}). (**B**) Increased affinity between SC1 ($SmD1_H$/D2) and SC3 ($SmD3/B_{H1}$ or $SmD3/B_{H2}$) was tested by pull-down assay. Wild-type SmD1/D2 and SmD3/B served as controls. (**C**) SC1 ($SmD1_H$/D2) pulled down SC3 ($SmD3/B_{H1}$ or $SmD3/B_{H2}$) and SC2 (SmF/E/G) together but could not pull down wild-type SmD3/B and SmF/E/G together. (**D**) Mixture of $SmD1_H$/D2, $SmD3/B_{H1}$, and SmF/E/G formed a heptameric complex in GFC (Superdex 200 10/300GL). (**E**) The heptameric complex (named as 7Sm-LLH) showed a doughnut-shaped conformation in negative staining EM. The diameters of the outside and inner rings are ~80 and ~20 Å, respectively. A representative region is shown. Several particles are shown in magnification. See more details in Supplementary Fig. S6 for additional EM images and analysis. (**F**) The reconstituted 7Sm-LLH could not bind to Gemin2/SMN(26–62). His-tagged Gemin2/MBP-SMN(26–62) bound to Ni-beads was tested to pull down 7Sm-LLH, SmF/E/G, $SmD1/D2$ + SmF/E/G, Sm core (see its reconstitution details in Supplementary Fig. S7), and $SmD1_H$/D2 + SmF/E/G served as controls. (**G–I**) Testing RNA-binding ability of 7Sm-LLH using GFC. (G) 7Sm-LLH was able to bind U6-snRNA. (H) 7Sm-LLH was unable to bind U6ΔSm-snRNA. (I) 7Sm-LLH was unable to bind U4-snRNA. For panels (G–I), the reconstituted 7Sm-LLH was incubated with RNAs (top, in cartoon) and the mixture were subject to GFC (middle, GFC trace of RNA alone is also shown for comparison). The elution fractions (marked by red and black bars) were analyzed by SDS–PAGE plus silver staining and GelRed staining (bottom). For each, one representative result of at least two independent experiments is shown.

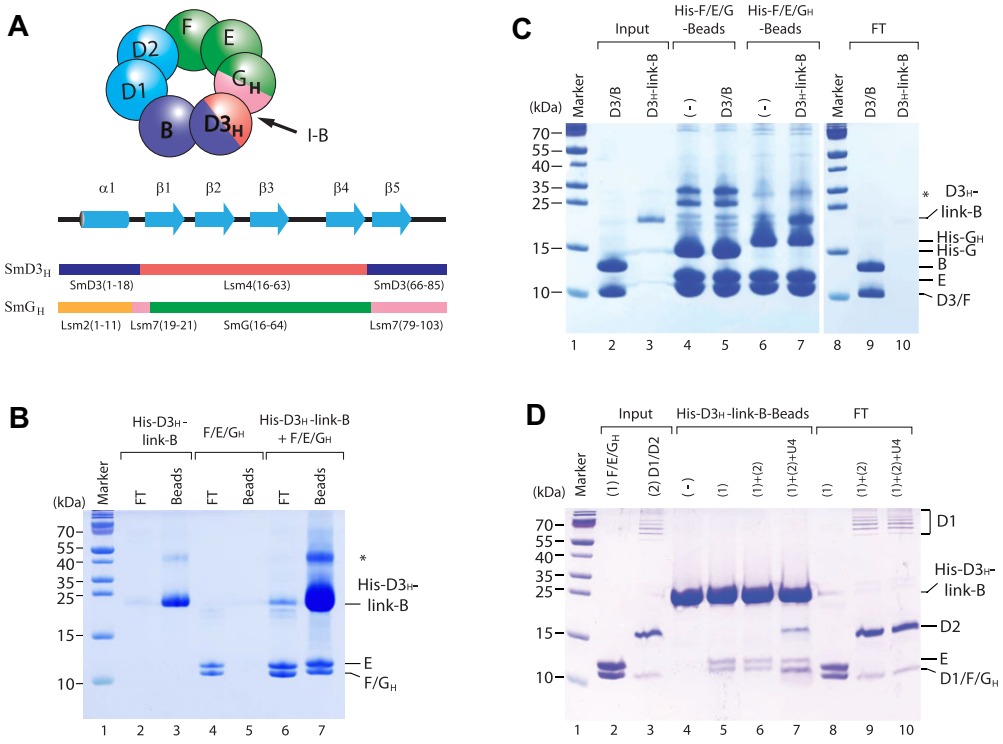


Figure 3. Increasing affinity between SC2 and SC3 in the Sm ring cannot convert it into an Lsm-type ring. (**A**) The strategy to increase affinity between SC2 and SC3. The interface of SmD3 is replaced by that of Lsm4 ($SmD3_H$), and the interface of SmG is replaced by that of Lsm7 with α1 of Lsm2 at its N-terminus (SmG_H). (**B** and **C**) Increased affinity between SC2 ($SmF/E/G_H$) and SC3 ($SmD3_H/B$, in a fused form of $D3_H$-link-B) was tested by pull-down assay. The His-tag was on SC2 in (B) and on SC3 in (C). Wild-type SmF/E/G and SmD3/B served as a control. (**D**) SC3 ($D3_H$-link-B) could not pull down SC2 and SC1 (SmD1/D2) together. However, in presence of U4-snRNA, SC3 ($D3_H$-link-B) could pull down SC2 and SC1 (SmD1/D2) together. For each in (B–D), one representative result of at least two independent experiments is shown. FT: Flow-through. Asterisks indicate impurities.

respectively), SDS–PAGE followed by silver staining showed a key difference: the first peak in the U6 snRNA experiment contained both RNA and Sm proteins, whereas the corresponding peak in the U6ΔLsm control contained only the mutant RNA (Fig. 2G and H). This demonstrates that 7Sm-LLH specifically binds U6 snRNA, mirroring the behavior of Lsm2–8. Second, we used GFC to test the interaction between the 7Sm-LLH ring and U4 snRNA, a representative spliceosomal snRNA that typically assembles with Sm proteins into an Sm ring. The peak containing RNA eluted at 12.25 ml, similar to the elution volume of snRNA alone (Fig. 2I). However, SDS–PAGE/Silver staining detected minimal Sm proteins in these fractions, while 7Sm-LLH eluted separately at 14.5 ml (Fig. 2I). In contrast, a reconstituted Sm core comprising wild-type 7Sm and U4 snRNA showed co-elution of RNA and Sm proteins at 11.9 ml (Supplementary Fig. S8). These results confirm that 7Sm-LLH cannot bind U4 snRNA. Furthermore, electrophoresis mobility shift assays (EMSAs) revealed that 7Sm-LLH failed to bind U1, U2 or U5 snRNAs, whereas wild-type 7Sm exhibited strong binding (Supplementary Fig. S9). This indicates that 7Sm-LLH has lost the ability to recognize RNAs containing the snRNP code. In summary, the above results demonstrate that increasing the binding affinity of I-C can convert Sm proteins from an Sm-type ring into an Lsm-type ring, both structurally and functionally.

7Sm cannot convert into an Lsm-type ring by enhancing the SC2–SC3 interaction

Next, we evaluated whether increasing affinity at interface B of the Sm core could yield an Lsm-type ring. Initially, we attempted the same approach as described above to replace I-B (between SmG and SmD3) with I-C between Lsm8 and Lsm2. However, we were unable to obtain a good expression of the mutant SC2 (data not shown). Then, we tried a different way. We replaced I-B in the Sm core with I-B in Lsm2–8 (between Lsm7 and Lsm4) and fused α1 of Lsm2 at the N-terminus of SmG mutant, creating SmG and SmD3 mutants, designed as SmG_H and $SmD3_H$, respectively (Fig. 3A and Supplementary Fig. S10). To enhance expression and facilitate size distinction, we also used a linked $SmD3_H/B$ ($SmD3_H$-link-B). Both $SmF/E/G_H$ and $SmD3_H$-link-B were expressed and behaved well (Fig. 3B and C). As anticipated, His-tagged $SmF/E/G_H$ was able to pull down $SmD3_H$-link-B, whereas His-tagged SmF/E/G could not pull down SmD3/B (Fig. 3C). Moreover, His-tagged $SmD3_H$-link-B pulled down $SmF/E/G_H$ (Fig. 3B). These data indicate that the affinity between the mutant SC2 and SC3 increased, as expected. Surprisingly, however, His-tagged $SmD3_H$-link-B could not pull down $SmF/E/G_H$ and SmD1/D2 together (Fig. 3D, lane 6). In contrast, when U4-snRNA was incubated with the three SCs, all the proteins and RNA were retained on the beads (Fig. 3D, lane 7), suggesting that both $SmF/E/G_H$ and $SmD3_H$-link-B have normal conformations similar to their wild-types. These results indicate that high affinity at both interfaces A and B does not enable the seven Sm proteins to form an Lsm-type ring. This finding is surprising, as it is against our initial hypothesis, suggesting that assembly dynamics and intermediate structures may inhibit the coalescence into a complete ring (Fig. 3A–D). This implies that the structures of the intermediates may prevent the integration of the final SC, which warrants further discussion.

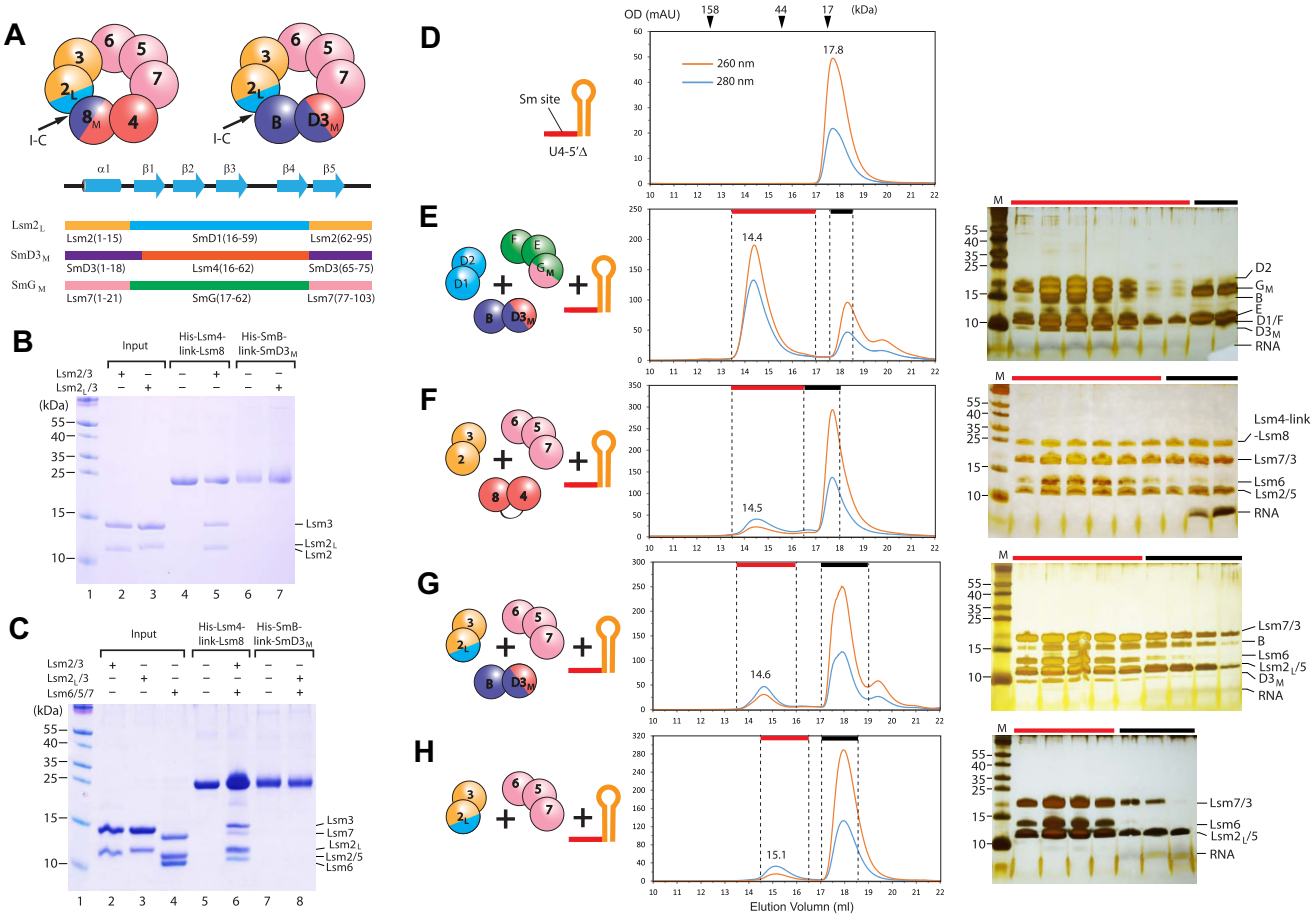


Figure 4. Reducing affinity between SC1 and SC3 in Lsm2–8 cannot form a Sm-type ring. (**A**) The strategy to reduce affinity between SC1 and SC3. The initial strategy was to replace the interface of Lsm2 with that of SmD1 ($Lsm2_L$) and the interface of Lsm8 with that of SmB ($Lsm8_M$) (upper left). However, this strategy was unsuccessful (data not shown). An alternative strategy was attempted, in which SmD3/B with the SmD3 interface replaced by that of Lsm4 ($SmD3_M$) was used to maintain the original interface for interaction with Lsm7 (upper right) (see more in Supplementary Fig. S12). (**B**) Reduced affinity between SC1 ($Lsm2_L/3$) and SC3 ($SmD3_M/B$) was tested by pull-down assay with Ni-beads. Wild-type Lsm2/3 and Lsm4/8 (in the form of Lsm4-link-Lsm8) served as controls. For better comparison, a linked version of $SmD3_M/B$ (SmB-link-$SmD3_L$) was used. (**C**) $SmD3_M/B$ (in the form of SmB-link-$SmD3_M$) could not pull-down $Lsm2_L/3$ and Lsm6/5/7, whereas Lsm4/8 (in the form of Lsm4-link-Lsm8) pulled down Lsm2/3 and Lsm6/5/7 together. (**D**–**H**) Binding test of various (L)Sm SCs to U4-5′Δ RNA through GFC. (D) U4-5′Δ RNA alone. (E) $SmD3_M/B$ was able to bind to SmD1/D2 and SmF/E/G_M and U4-5′Δ to form Sm core. (F) Lsm2/3, Lsm6/5/7, and Lsm4/8 (in the form of Lsm4-link-Lsm8) could not bind U4-5′Δ. (G) $Lsm2_L/3$, Lsm6/5/7 and $SmD3_M/B$ could not bind U4-5′Δ. (H) $Lsm2_L/3$ and Lsm6/5/7 could not bind U4-5′Δ to form a subcore. For each, one representative result of at least two independent experiments is shown.

Weakening the SC1–SC3 interaction of Lsm2–8 does not yield a Sm-type ring

Upon successful conversion of the Sm core into an Lsm-type ring by enhancing I-C affinity, we investigated whether reducing I-C affinity could convert Lsm2–8 into a Sm-type ring. We engineered a low-affinity interface to replace the high-affinity contact between Lsm2 and Lsm8 with the less stable interaction observed between SmD1 and SmB. The resulting Lsm2 mutant, denoted $Lsm2_L$, was created by fusing it with SmD1 segments (Fig. 4A and Supplementary Fig. S11). However, attempts to produce similar Lsm8 mutants resulted in either misfolded inclusion bodies or mutants that retained unwanted affinity with $Lsm2_L$ (Fig. 4A, upper left). Consequently, for SC3, we adopted SmB in its entirety combined with a modified SmD3 ($SmD3_M$), in which the β1–4 strands were replaced with those from Lsm4 to maintain the interface with Lsm7 (Fig. 4A, upper right, and Supplementary Figs S11 and S12). The resulting SC1 ($Lsm2_L/3$) and SC3 ($SmD3_M/B$) showed the anticipated reduced affinity, as evidenced by the absence of stable binding in pull-down analysis (Fig. 4B). Moreover, as anticipated, mixing $SmD3_M/B$ with Lsm6/5/7 and $Lsm2_L/3$ resulted in the loss of heptameric complex formation, in contrast to the native Lsm2–8 (Fig. 4C, compare lanes 8 and 6). To ensure accurate folding of $SmD3_M$, we tested its ability to integrate into the Sm core using the truncated U4-snRNA variant U4-5′Δ [27], which solely contains the Sm site and adjacent stem-loop (Fig. 4D). GFC confirmed the proper assembly of $SmD3_M/B$ into the Sm core (Fig. 4E). As a negative control, the native Lsm2–8 could not bind U4-5′Δ (Fig. 4F). At this stage, the engineered Lsm SCs, like their Sm counterparts, had lost spontaneous ring assembly capacity (Fig. 4C). However, unlike their Sm counterparts, these three Lsm SCs failed to form a stable RNA-dependent Sm-type ring when mixed with U4-5′Δ and analyzed by GFC (Fig. 4G). We further found that, unlike the formation of the Sm subcore from SmD1/D2, SmF/E/G and U-snRNA [27, 34], the intermediate $Lsm2_L/3/6/5/7$ complex displayed no binding to snRNA (Fig. 4H).

Converting an Lsm ring into Sm-type configuration

To adapt the $Lsm2_L/3/6/5/7$ complex for snRNA binding, we investigated structural differences between Lsm proteins and their Sm counterparts. We identified that, in addition to the signature loops 3 and 5, loop 2 in Sm proteins also interacts with RNA within the central channel (Supplementary Fig. S13). Based on this finding, we mutated the loop 2 sequences in Lsm3, Lsm6, and Lsm5 to match those of their Sm counterparts (Fig. 5A). Initial attempts to assemble the modified Lsm complex ($Lsm2_L/3_M$ and $Lsm6_M/5_{M1}/7$) with U4-5′Δ were unsuccessful, as the complex failed to form a stable RNA–protein complex efficiently (Supplementary Fig. S14). Subsequently, we introduced mutations F46Y/N75K in loops 3 and 5 of Lsm5 (generating $Lsm5_{M2}$) and L43F in loop 3 of Lsm7 (generating $Lsm7_M$) to enhance RNA binding (Fig. 5A). These modified Lsm proteins ($Lsm2_L/3_M$ and $Lsm6_M/5_{M2}/7_M$) successfully co-eluted with U4-5′Δ in GFC analysis, indicating RNA binding (Fig. 5B). Notably, the incorporation of $SmD3_M/B$ with these components yielded a co-eluted RNA–protein complex, suggesting the assembly of a RNA–Lsm complex resembling a fixed Sm-type ring (Fig. 5C). The specificity of RNA recognition was confirmed through control experiments with U4-5′Δ-ΔSm, a variant in which the 5Us in the Sm site were changed into 5Cs, where no RNA–protein complex was detected (Fig. 5D and E). Negative staining EM corroborated the doughnut-shaped structure of the RNA–Lsm complex with a diameter of ∼80 Å (Fig. 5F and Supplementary Fig. S6), matching previously characterized fixed rings [6, 7, 39]. Finally, pull-down assays demonstrated that $SmD3_M/B$ could bind $Lsm2_L/3_M$ and $Lsm6_M/5_{M2}/7_M$ in the presence of U4 snRNA, but not in the presence of U4ΔSm or U6 snRNA (Supplementary Fig. S15). This confirmed that the newly formed complex represents a Sm-type ring rather than an Lsm-type one.

Discussion

In most eukaryotic organisms, two distinct classes of (L)Sm heteroheptameric rings are discerned: the fixed Sm-type forms, as seen in spliceosomal snRNPs and the U7 snRNP, and the flexible Lsm-type forms, exemplified by Lsm2–8 and Lsm1–7. These rings are essential for various RNA metabolic processes. Of note, conversion of the Lsm heteroheptameric ring into the Sm ring is considered a key event in the evolution of spliceosome and eukaryogenesis [32]. However, the mechanistic rationale behind the formation of two disparate ring structures from a similar set of Sm/Lsm proteins remains an enigma. While phylogenetic studies have hinted at their evolutionary connection, the precise molecular mechanisms driving their divergence and the potential for interconversion remain unexplored.

In this study, we focused on the spliceosomal Sm ring and Lsm2–8, both ubiquitous to eukaryotic spliceosomes, to examine the possibility and biophysical principles of interconversion between Sm-type and Lsm-type ring architectures. By employing computational and biochemical analyses, we discerned contrasting affinity patterns across the three interfaces of the corresponding SCs in the Sm ring and Lsm2–8: the Sm core possesses a singular high-affinity interface, I-A, whereas the Lsm2–8 features two such sites, with a stronger likelihood of I-C > I-A (Fig. 1). This discrepancy implies divergent assembly pathways: the Sm core likely forms through an SC1–SC2 intermediate, while the Lsm ring's assembly could involve mostly SC1–SC3 (and possibly SC1–SC2 intermediates)—a stark divergence from previous models suggesting Lsm ring assembly mirrored Sm core processes [37, 38]. Our findings reveal that in Sm proteins, the presence of two high-affinity interfaces between SC1–SC2 and SC1–SC3 enables the formation of an Lsm-type ring resembling Lsm2–8 in both structure and function (Fig. 2). Contrary to expectations, enhancing the affinity of SC1–SC2 and SC2–SC3 interfaces in Sm proteins does not result in an Lsm-type ring (Fig. 3). Notably, an intermediate consisting of a pair of 2-subunit SCs (SC1 and SC3) leads to an Lsm-type ring; in contrast, an intermediate with a single 2-subunit SC and a 3-subunit SC (SC2) fails to do so, exemplified by the principle $(2 + 2) + 3 \neq (2 + 3) + 2$. This principle also appears consistent with the assembly patterns of Lsm1–7 and the U7 snRNP core (Fig. 1A). Interestingly, the formation of a Sm-type ring is not guaranteed by this (2 + 3) intermediate configuration alone (Fig. 4). To achieve a Sm-type ring, the (2 + 3) intermediate must incorporate additional RNA-binding residues within loops 2, 3, and 5, as illustrated by our successful conversion of the Lsm2/3/6/5/7 intermediate to be capable of snRNA binding (Fig. 5).

The question arises: why does the presence of a tetramer (2 + 2) intermediate facilitate the formation of an Lsm-type ring, while a pentamer (2 + 3) intermediate does not? The most plausible explanation posits that the horseshoe-shaped (2 + 3) intermediate formed by either SC1–SC2 or SC2–SC3 is burdened by a restricted entryway between edge subunits, thereby precluding the incorporation of the last SC. However, when SC1 and SC3 form a tetramer (2 + 2) intermediate first, the space between the side subunits is quite open, allowing the access of the third SC and adjustment of its conformation for binding. Although further investigation is required, a narrower opening in the pentamer is supported by several pieces of evidence: (i) The conformation of the horseshoe-shaped SmD1/D2/F/E/G (5Sm) in the crystal structure of the heteroheptameric complex 5Sm/Gemin2/SMN(26–62) (PDB: 5XJL) has a larger curvature than the counterpart in the mature Sm core [26, 27]. (ii) The calculated binding energy for assembling D3/B into SmD1/D2/F/E/G (the sum of binding energies of I-B and I-C, –9.4 kcal/mol) is in favor of Lsm-type ring formation if there is no spatial restriction (Fig. 1A). (iii) In our pull-down results, high-affinity interfaces A and B in Sm SCs failed to form an Lsm-type ring (Fig. 3). The major contributor to the narrow opening of the (2 + 3) intermediate is most likely SC2 (SmF/E/G or Lsm6/5/7), which may have a larger curvature than the other SCs. Several lines of evidence support it: (i) in their solution and crystal structure, both SmF/E/G and Lsm6/5/7 form hexameric rings, in which their curvatures are larger than those in the Sm core and Lsm2-8 ring [38, 40, 41]. (ii) In contrast, the crystal structure of Lsm2/3 shows that it forms a heptamer with an additional Lsm3 filling the gap (PDB: 4N0A) [42], rather than a hexameric ring.

Since eukaryotes possess at least fourteen (L)Sm proteins, including the Sm core and Lsm2–8, both of which are pivotal components of spliceosome, while prokaryotes harbor only 1–2 Lsm proteins, the evolution of these two sets of (L)Sm proteins must have occurred within the eukaryogenesis process, estimated to have taken place between ∼1.8 and 2.7 billion years ago [43]. A key event in eukaryogenesis was the symbiosis of an archaeal ancestor and a bacterial (proto-mitochondrial) partner [32, 43]. It is believed

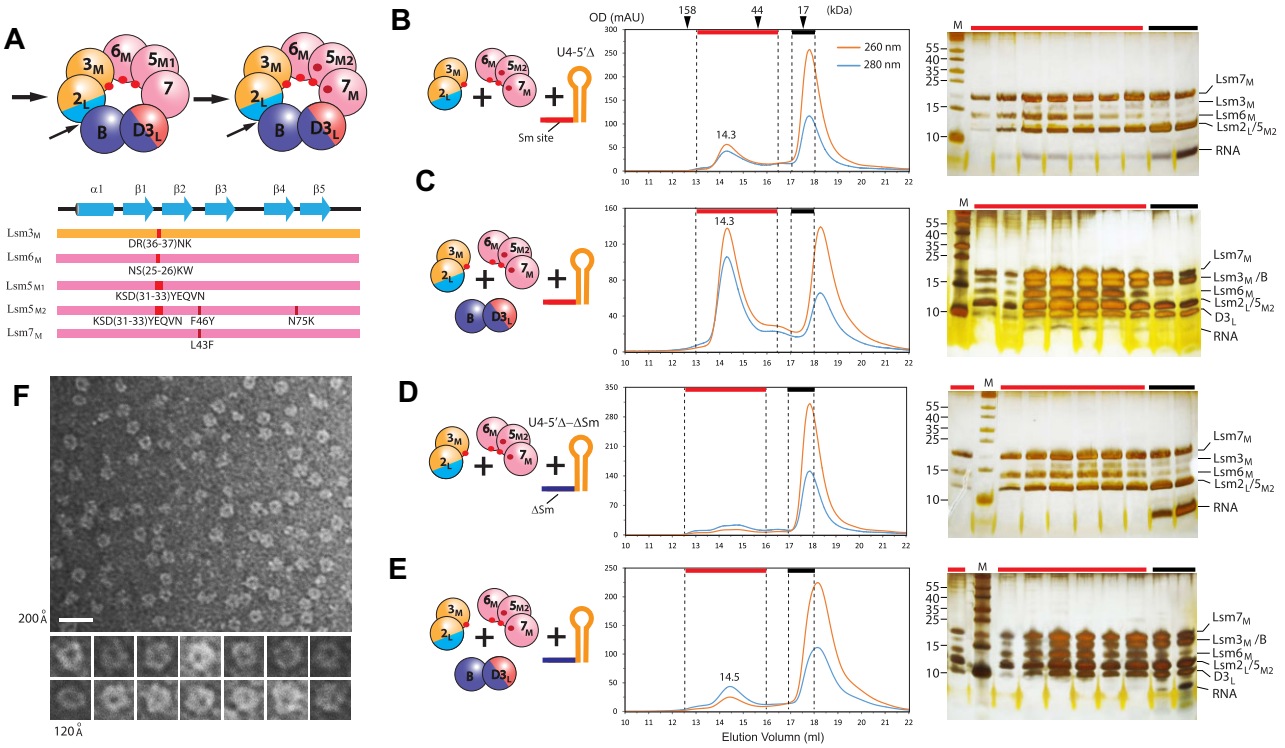


Figure 5. Converting Lsm2–8 ring into a Sm-type ring by additional mutations in SC1 and SC2. (**A**) The strategy to make mutations in SC1 and SC2. The initial step is replacing the RNA-interacting loop2s of Lsm3, Lsm6, and Lsm5 by those of SmD2, SmF, and SmE, respectively (see details in Supplementary Fig. S13). However, the combination (upper left) was still unable to efficiently bind U4-5′Δ (Supplementary Fig. S14). Then, in the second round of mutation, point mutations were further made in loops 3 and 5 of $Lsm5_{M1}$ and Loop 5 of Lsm7 to be the same as in SmE/G (upper right). (**B**–**E**) The bindings of $Lsm2_L/3_M$ and $Lsm6_M/5_{M2}/7_M$ to U4-5′Δ (B) and to U4-5′Δ-ΔSm (D), and the bindings of $Lsm2_L/3_M$, $Lsm6_M/5_{M2}/7_M$ and $SmD3_L/B$ to U4-5′Δ (C) and to U4-5′Δ-ΔSm (E) were tested with GFC. The input components are showed in cartoon (left), the mixtures were subjected to GFC (middle) followed by SDS-PAGE plus silver staining (right). For each in panels (B–E), one representative result of at least two independent experiments is shown. (**F**) The Lsm-U4-5′Δ complex showed a doughnut-shaped conformation in negative staining EM, with the diameters of the outside and inner rings being ~80 and ~20 Å, respectively. A representative region is shown. Several particles are shown in magnification. See more details in Supplementary Fig. S6.

that spliceosomal introns and many core spliceosome components originated from self-splicing group II introns present in the proto-mitochondria [32, 44, 45]. Specifically, snRNAs, including at least U2, U5, and U6, likely descended from fragmented group II introns, while the largest and key protein, Prp8, likely evolved from an intron-encoded protein (IEP), which possesses both maturase and reverse transcriptase functions [44–46]. Supporting this model, studies have identified intermediate forms of self-splicing group II introns in plant cytoplasmic organelles, consisting of 2–3 separate RNA pieces that form a *trans*-splicing complex structurally similar to ancestral group II introns [46, 47]. These findings provide evolutionary evidence for fragmentation events that likely preceded or coincided with the emergence of the heteroheptameric Lsm ring. The Lsm2–8 and the Sm ring likely emerged in the early stages of eukaryotic evolution, given their essential role in snRNP core formation, snRNA stabilization, and protein recruitment. Phylogenetic analyses indicate that Lsm2–8 and the Sm ring sequentially derived from prokaryotic Lsm proteins through two rapid waves of duplications followed by divergences [2–4]. The first wave involved the transformation of a homoheptameric ring into a heteroheptameric ring. Such conversions from homomeric to heteromeric ring-shaped complexes have occurred frequently in eukaryotic evolution, exemplified by the 20S CP [1] and exosome [48]. Typically, these derived heteromeric rings retain highly similar structures, functions and even assembly mechanisms compared to their homomeric counterparts, while offering enhanced spatial specificity through their rotational sweep. In the case of Lsm2–8, it assembles independently of RNA and functions as an RNA chaperone, akin to homomeric Lsm rings. The second wave involved the conversion of one heteroheptameric ring into another, exhibiting distinct functional and assembly characteristics. This type of conversion is relatively rare.

Based on the results of our protein engineering and reconstitution experiments, we propose a mechanistic model for the divergence between flexible Lsm-type rings and fixed Sm-type rings (Fig. 6). The model involves four key evolutionary steps (steps I–IV), driven by gene duplication and affinity modulation at subunit interfaces. During early eukaryogenesis, the ancestral Lsm2–8 ring (Ring 1) likely assembled via RNA-independent interactions, primarily through the high-affinity SC1–SC3 interface (pathway a). It prefers reversible binding to U6 snRNA (step a3) but disfavors binding to U1/U2/U4/U5 snRNAs (step a3'). Although an SC1–SC2 intermediate (pathway b) could form transiently, its inability to stably bind U-snRNA (step b2) or incorporate SC3 (step b2') would favor dissociation and redirect assembly through the SC1–SC3 pathway. Notably, despite theoretical potential, no natural SC2–SC3 interaction (pathway c) has been observed. Gene duplication of Lsm2–8 enabled mutational divergence in one

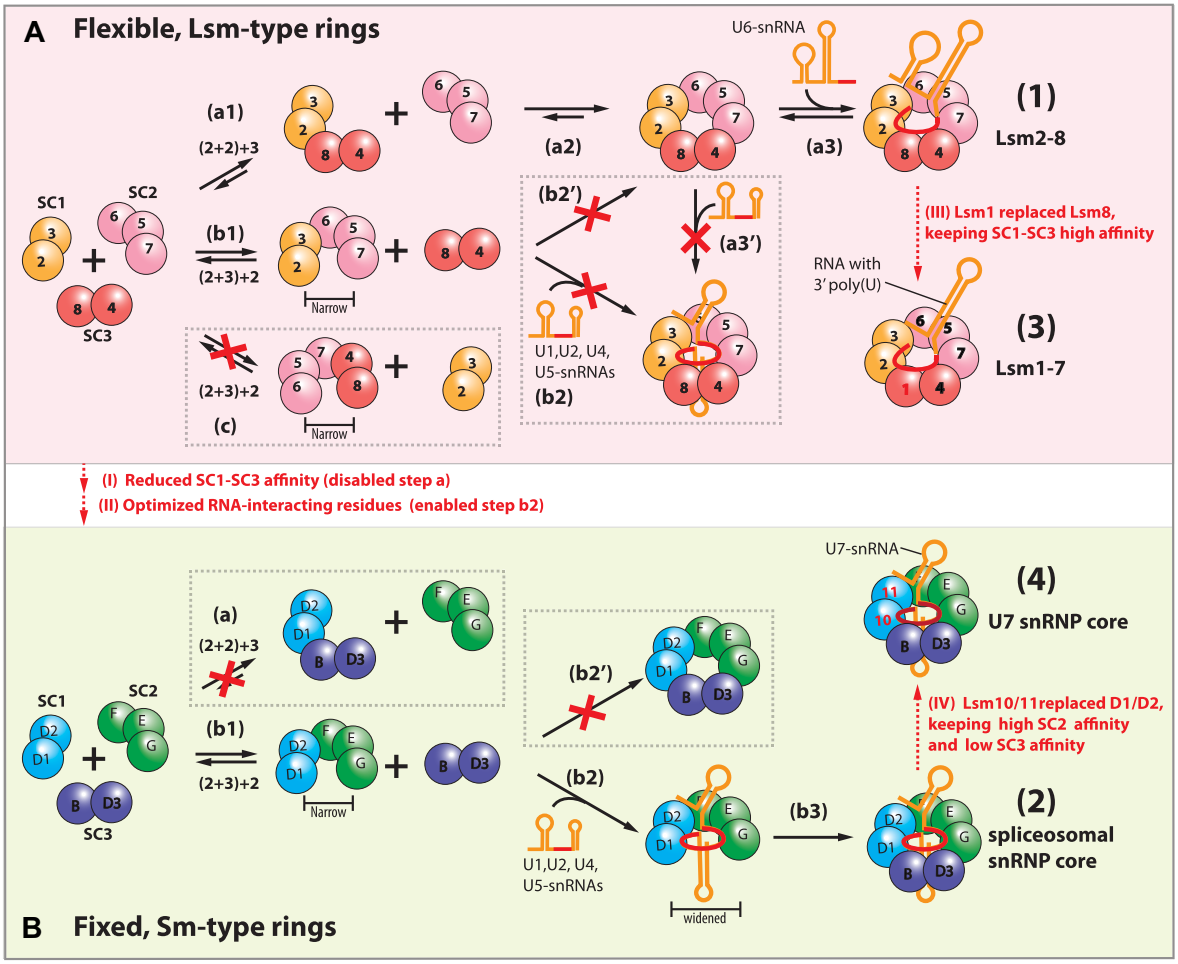


Figure 6. Schematic model of the assembly pathways of the Lsm-type and Sm-type heteroheptameric rings, along with the key steps in the evolution of four (L)Sm rings (I–IV, in red). See the main text for details. A square with a dash border indicates a theoretically possible but nonexistent pathway.

copy, with two critical steps driving Sm-type ring emergence: (i) Reduced SC1–SC3 affinity (Step I) abolished Lsm-type ring formation, leaving only the SC1–SC2 pentamer (pathway b) with poor RNA-binding capacity; (ii) Enhanced RNA-binding residues in loops 2/3/5 of the SC1–SC2 complex (Step II) enabled stable U1/U2/U4/U5 snRNA interactions (step b2) and subsequent SC3 incorporation (step b3) to form the Sm-type ring (Ring 2). These primordial (L)Sm rings likely supported a rudimentary protospliceosome, facilitating the evolution of the spliceosome and the emergence of the Last Eukaryotic Common Ancestor (LECA) [32]. Subsequent gene duplications further diversified the system. Lsm8's duplication produced Lsm1 (Step III), which retained high affinity for SC1 but formed the cytoplasmic Lsm1–7 (Ring 3) involved RNA decay. Like its ancestor, Lsm1–7 preferentially binds poly(U) tracts at the 3′ end of RNAs [15]. The emergence of Lsm10/11 in metazoans replaced SmD1/D2, preserving high affinity for SC2 but low affinity for SC3 (Step IV), to specialize in U7 snRNA binding (Ring 4) for histone mRNA processing [11, 21, 49]. Notably, U7 snRNA contains a slightly different Sm site (AAUUUGUCUAG, with key differences underlined) and a 3′ stem-loop—both of which are essential for U7 snRNP assembly [21, 49]. These four ring types (Lsm2–8, canonical Sm core, Lsm1–7, and U7 snRNP core) now dominate in metazoans, though minimal sets persist in some lineages (e.g. 7 Lsm + 7 Sm proteins) [50].

By integrating previous structural studies of individual Sm and Lsm rings [6–8, 11, 15–18], our studies further elucidate the mechanistic link between the assembly mechanisms and RNA-binding preferences of Lsm-type and Sm-type rings, revealing a tight coupling between the RNA-binding modes and assembly pathways. Lsm-type rings (homo- or heteroheptameric) assemble spontaneously without RNA and preferentially bind RNAs with a 3′-terminal poly(U) tract, as the poly(U) stretch encircles the central channel, positioning the 3′ end deeper within it [15]. Additional nucleotides downstream of poly(U) are disfavored because they either cannot thread through the central channel or, if retained on the 5′-side, introduce steric strain that counteracts the binding energy from poly(U). In this study, despite retaining all canonical RNA-binding residues, our engineered 7Sm–LLH complex spontaneously assembles into a heteroheptameric ring, shifting its binding preference from U1/U2/U4/U5 snRNAs to U6 snRNA (Fig. 2 and Supplementary Fig. S9). In contrast, Sm-type rings require RNA for stepwise assembly. They initially form a 5Sm intermediate, which presumably has a narrow opening to block complete heptameric ring formation. This 5Sm complex selectively binds RNAs containing poly(U) followed by a stem-loop structure: the horseshoe-shaped 5Sm permits unhindered RNA access, while loop 4 of specific Sm proteins interacts with downstream nucleotides, providing additional binding energy [18]. RNA binding in-

duces conformational opening, enabling incorporation of the final Sm heterodimer to complete the Sm core assembly. In this study, we show that blocking spontaneous assembly of Lsm proteins (allowing only the Lsm2/3/6/5/7 intermediate) and optimizing a few RNA-binding residues converts them into an RNA-dependent heteroheptameric ring that preferentially binds to RNAs containing the snRNP code (Fig. 5). This establishes an assembly-driven selectivity rule: constitutively assembled Lsm-type rings reversibly bind minimal 3′-poly(U), while RNA-dependent Sm-type ring assembly enables stable recruitment of substrates containing poly(U) followed by a stem-loop through cooperative RNA–protein interactions.

In this study, we investigated how the interaction affinity between SCs influences their assembly pathways and RNA-binding preferences. To modulate binding affinities, we primarily employed chimeric (L)Sm proteins. However, even single-residue mutations at the interface could alter SC binding affinities, particularly to weaken them. The interface between neighboring (L)Sm proteins is primarily stabilized by two types of interactions (using SmB–SmD1 as an example, Supplementary Fig. S3): (i) anti-parallel β-sheet interactions between β5 (SmB) and β4 (SmD1); (ii) hydrophobic and electrostatic contacts between α1 (SmB) and the β2-β3-β4 surface (SmD1). Focusing on the β-sheet interaction (i), since the main-chain hydrogen bonds between β5 (one subunit) and β4 (its neighbor) are conserved, the side-chain properties become crucial. Hydrophobic residues (e.g. Val and Phe) enhance binding affinity, whereas small or polar residues (e.g. Ser) tend to weaken it. Notably, the conserved Ser79 (β5, SmB) and Ser59 (β4, SmD1) likely contribute to the relatively low affinity of SmB–SmD1, in contrast to the stable Lsm8-Lsm2 interaction mediated by Val69 and Phe61 at equivalent positions. Such residue-based affinity differences await experimental confirmation.

A small number of single-residue mutations may suffice to interconvert Lsm- and Sm-type rings. However, the accumulation of additional mutations likely facilitated further functional specialization in (L)Sm rings. Three key examples illustrate this divergence: (i) the sequence divergence between Lsm1 and Lsm8 contributes to cytoplasmic retention of the Lsm1–7 complex for specialized RNA decay functions [19]. (ii) Distinct surface residues in SmD1/D2/F/E/G (versus Lsm2/3/6/5/7) prevent Pat1 binding while enabling selective Gemin2 interaction [10, 26]. (iii) Progressive interface mutations converted ancestral chaperone-independent Sm core assembly (observed in budding yeast via strong SmD1/D2-SmF/E/G interactions) into the chaperone-dependent pathway essential for metazoans [36].

Our studies of the interconversions between Lsm-type and Sm-type rings also enrich understanding of the divergence of oligomeric complexes, especially ring-shaped ones. Unlike the archetypal 20S core particle, which maintains a consistent hollow cylinder structure and function as a protein degradation machinery despite subunit differentiations from archaea to eukaryotes [1], the (L)Sm rings demonstrate variability among SC interfaces, underlining the complexity and plasticity of their assembly mechanisms. These insights may inform the design of novel self-assembling heteromeric structures for a range of applications.

In summary, our study reveals how distinct assembly pathways—driven by differential subunit affinities and RNA interactions—govern the functional divergence between Lsm-type and Sm-type rings. These findings provide mechanistic insights into (L)Sm ring evolution and a framework for engineering self-assembling oligomeric complexes.

Acknowledgements

Author contributions: L.M. contributed to conversion of the Sm ring to an Lsm-type ring and EM study. Y. Hou contributed to characterization of subcomplex interactions of wild-type Lsm2–8 and the Sm ring and conversion of Lsm2–8 to a Sm-type ring. Y. Hu, Y.W., and C.S. contributed to recombined protein preparations. Y.L. contributed to EM data analysis. D.S. contributed to funding acquisition and writing-review & editing. R.Z. contributed to funding acquisition, conceptualization, supervision, and writing.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by National Key R&D programs of China (No. 2017YFA0504300 and 2017YFA0505900) and Sichuan Science & Technology Plan Project (No. 2020YJ0209). Funding to pay the Open Access publication charges for this article was provided by the National Key R&D programs of China (No. 2017YFA0504300 and 2017YFA0505900) and Sichuan Science & Technology Plan Project (No. 2020YJ0209).

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

1. Fuchs ACD, Hartmann MD. On the origins of symmetry and modularity in the proteasome family: symmetry transitions are pivotal in the evolution and functional diversification of self-compartmentalizing proteases. *Bioessays* 2019;**41**:e1800237. https://doi.org/10.1002/bies.201800237
2. Scofield DG, Lynch M. Evolutionary diversification of the Sm family of RNA-associated proteins. *Mol Biol Evol* 2008;**25**:2255–67. https://doi.org/10.1093/molbev/msn175
3. Veretnik S, Wills C, Youkharibache P *et al.* Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. *PLoS Comput Biol* 2009;**5**:e1000315. https://doi.org/10.1371/journal.pcbi.1000315
4. Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S *et al.* Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J* 1999;**18**:3451–62. https://doi.org/10.1093/emboj/18.12.3451
5. Kambach C, Walke S, Young R *et al.* Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* 1999;**96**:375–87. https://doi.org/10.1016/S0092-8674(00)80550-4
6. Pomeranz Krummel DA, Oubridge C, Leung AK *et al.* Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* 2009;**458**:475–80. https://doi.org/10.1038/nature07851

7. Leung AK, Nagai K, Li J. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 2011;**473**:536–9. https://doi.org/10.1038/nature09956
8. Zhou L, Hang J, Zhou Y *et al.* Crystal structures of the Lsm complex bound to the 3′ end sequence of U6 small nuclear RNA. *Nature* 2014;**506**:116–20. https://doi.org/10.1038/nature12803
9. Zhou L, Zhou Y, Hang J *et al.* Crystal structure and biochemical analysis of the heptameric Lsm1–7 complex. *Cell Res* 2014;**24**:497–500. https://doi.org/10.1038/cr.2014.18
10. Sharif H, Conti E. Architecture of the Lsm1–7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell Rep* 2013;**5**:283–91. https://doi.org/10.1016/j.celrep.2013.10.004
11. Sun Y, Zhang Y, Aik WS *et al.* Structure of an active human histone pre-mRNA 3′-end processing machinery. *Science* 2020;**367**:700–3. https://doi.org/10.1126/science.aaz7758
12. Schumacher MA, Pearson RF, Moller T *et al.* Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* 2002;**21**:3546–56. https://doi.org/10.1093/emboj/cdf322
13. Toro I, Basquin J, Teo-Dreher H *et al.* Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus. *J Mol Biol* 2002;**320**:129–42. https://doi.org/10.1016/S0022-2836(02)00406-0
14. Toro I, Thore S, Mayer C *et al.* RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J* 2001;**20**:2293–303. https://doi.org/10.1093/emboj/20.9.2293
15. Montemayor EJ, Virta JM, Hayes SM *et al.* Molecular basis for the distinct cellular functions of the Lsm1–7 and Lsm2–8 complexes. *RNA* 2020;**26**:1400–13. https://doi.org/10.1261/rna.075879.120
16. Weber G, Trowitzsch S, Kastner B *et al.* Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J* 2010;**29**:4172–84. https://doi.org/10.1038/emboj.2010.295
17. Kondo Y, Oubridge C, van Roon AM *et al.* Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* 2015;**4**:e04986. https://doi.org/10.7554/eLife.04986
18. Li J, Leung AK, Kondo Y *et al.* Re-refinement of the spliceosomal U4 snRNP core-domain structure. *Acta Crystallogr D Struct Biol* 2016;**72**:131–46. https://doi.org/10.1107/S2059798315022111
19. Tharun S, He W, Mayes AE *et al.* Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* 2000;**404**:515–8. https://doi.org/10.1038/35006676
20. Bouveret E, Rigaut G, Shevchenko A *et al.* A Sm-like protein complex that participates in mRNA degradation. *EMBO J* 2000;**19**:1661–71. https://doi.org/10.1093/emboj/19.7.1661
21. Pillai RS, Grimmler M, Meister G *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**:2321–33. https://doi.org/10.1101/gad.274403
22. Pillai RS, Will CL, Luhrmann R *et al.* Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. *EMBO J* 2001;**20**:5470–9. https://doi.org/10.1093/emboj/20.19.5470
23. Seto AG, Zaug AJ, Sobel SG *et al.* Saccharomyces cerevisiae telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature* 1999;**401**:177–80. https://doi.org/10.1038/43694
24. Tang W, Kannan R, Blanchette M *et al.* Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* 2012;**484**:260–4. https://doi.org/10.1038/nature10924
25. Chari A, Golas MM, Klingenhager M *et al.* An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell* 2008;**135**:497–509. https://doi.org/10.1016/j.cell.2008.09.020
26. Zhang R, So BR, Li P *et al.* Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell* 2011;**146**:384–95. https://doi.org/10.1016/j.cell.2011.06.043
27. Yi H, Mu L, Shen C *et al.* Negative cooperativity between Gemin2 and RNA provides insights into RNA selection and the SMN complex's release in snRNP assembly. *Nucleic Acids Res* 2020;**48**:895–911. https://doi.org/10.1093/nar/gkz1135
28. Hu Y, Hou Y, Zhou S *et al.* Mechanism of assembly of snRNP cores assisted by ICln and the SMN complex in fission yeast. *iScience* 2023;**26**:107604. https://doi.org/10.1016/j.isci.2023.107604
29. Tisdale S, Lotti F, Saieva L *et al.* SMN is essential for the biogenesis of U7 small nuclear ribonucleoprotein and 3′-end formation of histone mRNAs. *Cell Rep* 2013;**5**:1187–95. https://doi.org/10.1016/j.celrep.2013.11.012
30. Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* 2014;**15**:108–21. https://doi.org/10.1038/nrm3742
31. Cauchi RJ. SMN and Gemins: 'we are family' … or are we?: insights into the partnership between Gemins and the spinal muscular atrophy disease protein SMN. *Bioessays* 2010;**32**:1077–89. https://doi.org/10.1002/bies.201000088
32. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 2006;**440**:41–5. https://doi.org/10.1038/nature04531
33. Roithova A, Feketova Z, Vanacova S *et al.* DIS3L2 and LSm proteins are involved in the surveillance of Sm ring-deficient snRNAs. *Nucleic Acids Res* 2020;**48**:6184–97. https://doi.org/10.1093/nar/gkaa301
34. Raker VA, Plessel G, Luhrmann R. The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *EMBO J* 1996;**15**:2256–69. https://doi.org/10.1002/j.1460-2075.1996.tb00579.x
35. Zaric B, Chami M, Remigy H *et al.* Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* 2005;**280**:16066–75. https://doi.org/10.1074/jbc.M414481200
36. Wang Y, Chen X, Kong X *et al.* A unique mechanism of snRNP core assembly. *Nat Commun* 2025;**16**:3166. https://doi.org/10.1038/s41467-025-58461-7
37. Sobti M, Cubeddu L, Haynes PA *et al.* Engineered rings of mixed yeast Lsm proteins show differential interactions with translation factors and U-rich RNA. *Biochemistry* 2010;**49**:2335–45. https://doi.org/10.1021/bi901767w
38. Mund M, Neu A, Ullmann J *et al.* Structure of the LSm657 complex: an assembly intermediate of the LSm1–7 and LSm2–8 rings. *J Mol Biol* 2011;**414**:165–76. https://doi.org/10.1016/j.jmb.2011.09.051
39. Kastner B, Bach M, Luhrmann R. Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc Natl Acad Sci USA* 1990;**87**:1710–4. https://doi.org/10.1073/pnas.87.5.1710
40. Wu D, Jiang S, Bowler MW *et al.* Crystal structures of Lsm3, Lsm4 and Lsm5/6/7 from *Schizosaccharomyces pombe*. *PLoS One* 2012;**7**:e36768. https://doi.org/10.1371/journal.pone.0036768
41. Plessel G, Luhrmann R, Kastner B. 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* 1997;**265**:87–94. https://doi.org/10.1006/jmbi.1996.0713
42. Wu D, Muhlrad D, Bowler MW *et al.* Lsm2 and Lsm3 bridge the interaction of the Lsm1–7 complex with Pat1 for decapping activation. *Cell Res* 2014;**24**:233–46. https://doi.org/10.1038/cr.2013.152
43. Vosseberg J, van Hooff JJE, Kostlbacher S *et al.* The emerging view on the origin and early evolution of eukaryotic cells. *Nature* 2024;**633**:295–305. https://doi.org/10.1038/s41586-024-07677-6
44. Vosseberg J, Snel B. Domestication of self-splicing introns during eukaryogenesis: the rise of the complex spliceosomal machinery. *Biol Direct* 2017;**12**:30. https://doi.org/10.1186/s13062-017-0201-6

45. Galej WP, Toor N, Newman AJ *et al.* Molecular mechanism and evolution of nuclear pre-mRNA and group II intron splicing: insights from cryo-electron microscopy structures. *Chem Rev* 2018;**118**:4156–76. https://doi.org/10.1021/acs.chemrev.7b00499
46. Sharp PA. Five easy pieces. *Science* 1991;**254**:663. https://doi.org/10.1126/science.1948046
47. Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J *et al.* A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. *Cell* 1991;**65**:135–43. https://doi.org/10.1016/0092-8674(91)90415-U
48. Januszyk K, Lima CD. The eukaryotic RNA exosome. *Curr Opin Struct Biol* 2014;**24**:132–40. https://doi.org/10.1016/j.sbi.2014.01.011
49. Kolev NG, Steitz JA. *In vivo* assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site. *Nat Struct Mol Biol* 2006;**13**:347–53. https://doi.org/10.1038/nsmb1075
50. Reimer KA, Stark MR, Aguilar LC *et al.* The sole LSm complex in *Cyanidioschyzon merolae* associates with pre-mRNA splicing and mRNA degradation factors. *RNA* 2017;**23**:952–67. https://doi.org/10.1261/rna.058487.116

Received: August 13, 2024. **Revised:** April 4, 2025. **Editorial Decision:** April 24, 2025. **Accepted:** May 17, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.