

Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex

Cell Research (2014) 24:497-500. doi:10.1038/cr.2014.18; published online 11 February 2014

Dear Editor,

The Lsm (Sm-like) family of proteins, characterized by the Sm fold [1] and conserved among eukaryotes, plays an important role in RNA biogenesis. The Sm and Lsm complexes play an essential role in pre-mRNA splicing [2]. Of the five small ribonucleoproteins (snRNPs), four (U1, U2, U4, U5) contain the Sm heptamer ring, whereas the U6 snRNP contains a specific Lsm2-8 heptamer that comprises Lsm2, Lsm3, Lsm4, Lsm5, Lsm6, Lsm7, and Lsm8 [3]. Another Lsm heptameric complex, Lsm1-7, which differs from Lsm2-8 by one Lsm protein, is thought to function in mRNA decapping, a crucial step in the mRNA degradation pathway [4].

In yeast, mRNA degradation begins with deadenylation catalyzed mainly by the deadenylase Ccr4-Not complex to shorten the polyA tail [5]. Subsequently, whether the 3'-end of mRNA is bound by the Lsm1-7 complexes is an essential determinant for the directionality of mRNA degradation. In the absence of Lsm1-7, the exposed 3'-end of mRNA is captured by exosome and degradation begins from the 3'-end [6]. In the presence of Lsm1-7, the 3'-end of mRNA is recognized by the Pat1-Lsm1-7 complex [6]. Then the decapping enzymes Dcp1 and Dcp2 are recruited to decap mRNA, followed by Xrn1-mediated degradation from the 5'-end [7].

The Sm and Lsm proteins exhibit contrasting functions and distinct biochemical properties. First, the Sm proteins are located in the nucleus and only function in the spliceosome pathway of mRNA splicing. The Lsm proteins appear both in the nucleus and cytoplasm and are involved in various pathways of RNA metabolism. Second, the Sm proteins only assemble into a heptamer ring in the presence of snRNA, whereas the Lsm proteins can form two stable heptamer rings by themselves, the Lsm1-7 and Lsm2-8 complexes. The Sm ring binds the central part of snRNA whereas the Lsm2-8 ring recognizes the 3'-end of U6 snRNA [8].

Unlike the Sm proteins, whose heptameric ring can be readily assembled from stable heteromeric subcomplexes, the individually purified Lsm proteins, either alone

or in heteromeric forms, failed to form a correctly assembled heptamer in the absence of denaturation/refolding. Lsm3 is known to form a stable homo-octamer by itself [9]; co-expression of Lsm5, Lsm6, and Lsm7 led to formation of a stable, but non-functional, hetero-hexamer [10]. In fact, every combination of hetero-dimeric or hetero-trimeric Lsm subcomplex would result in the formation of a stable oligomer, which can be reconstituted into a heptameric ring only by denaturation and refolding [11]. After numerous trials, we succeeded in co-expression of all seven Lsm proteins from *S. cerevisiae* using the pQLink plasmid [12]. The Lsm1-7 complex was biochemically purified to homogeneity (Supplementary information, Figure S1). Analysis by mass spectrometry confirmed the presence of all seven Lsm proteins.

Despite rigorous trials, we were unable to crystallize the Lsm1-7 complex. To solve this problem, we launched a systematic protein engineering effort that involved removal of flexible sequences and mutation of cysteine to serine. Eventually we crystallized the Lsm1-7 complex in the P21 space group (Supplementary information, Figure S2). To facilitate structure determination, we also generated and crystallized the selenomethionine-labeled Lsm1-7 complex in the P1 space group. The structure of Lsm1-7 was determined by a combination of selenium-based single anomalous dispersion (SAD) and molecular replacement using the atomic coordinates of an archael Sm-like homo-heptamer (PDB code 1I81 [13]). The final atomic model was refined at 3.0 Å resolution (Supplementary information, Table S1 and Figure S3).

The overall appearance of the heptameric Lsm1-7 complex from *S. cerevisiae* resembles a thick donut, with an outer diameter of approximately 70 Å, an inner diameter of 5 Å, and a thickness of 45 Å (Figure 1A). The seven Lsm components sequentially interact with each other to form a closed ring, with the order of Lsm1-Lsm2-Lsm3-Lsm6-Lsm5-Lsm7-Lsm4. Except for Lsm1, each component within the ring only interacts with two neighboring Lsm proteins. For example, Lsm7 is in contact only with Lsm5 and Lsm4. The C-terminal sequence of Lsm1 forms an extended α -helix that crosses over the

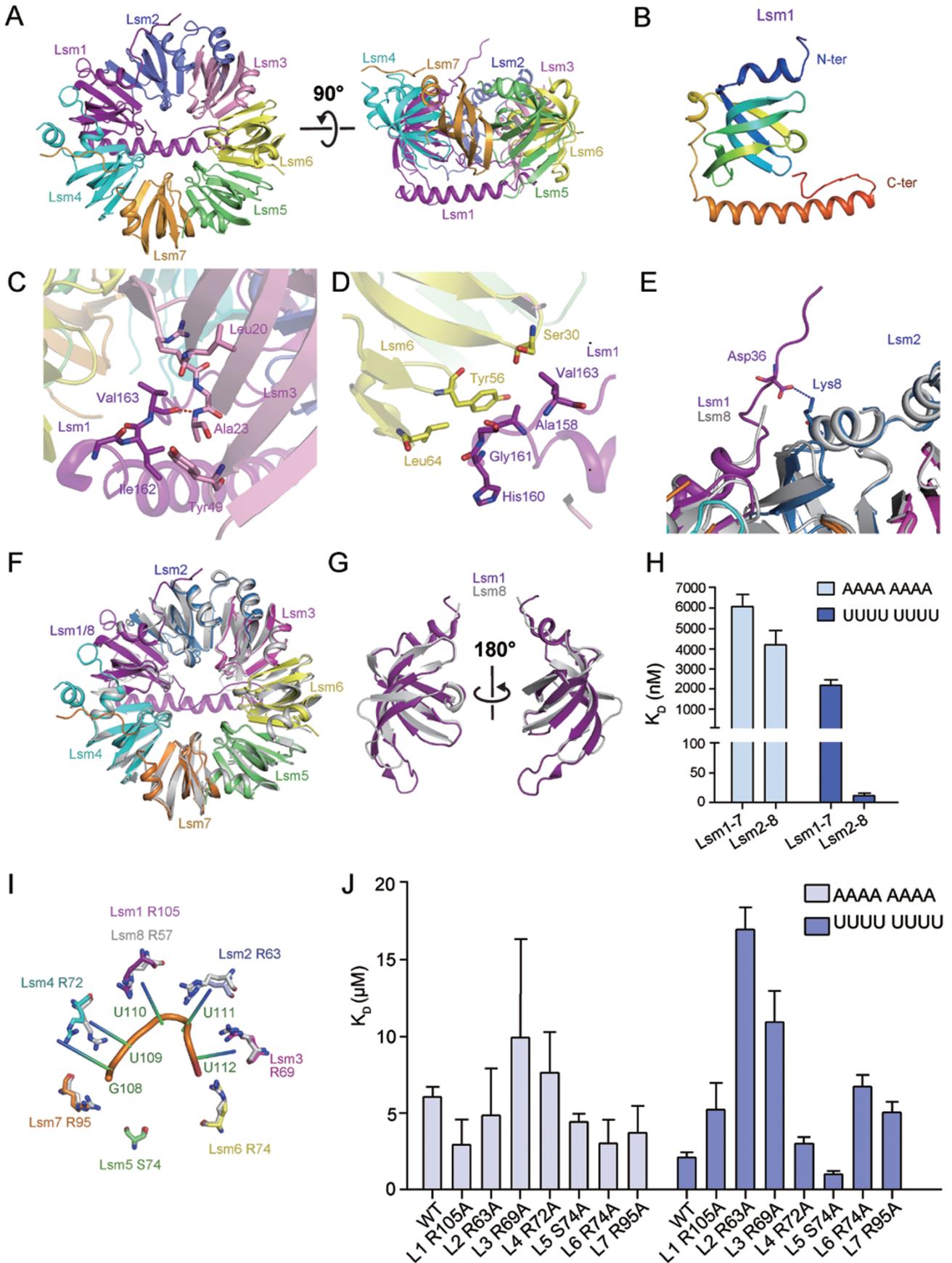


Figure 1 Structure of the Lsm1-7 heptameric complex from *S. cerevisiae*. **(A)** Overall structure of the Lsm1-7 complex in two perpendicular views. The seven Lsm subunits are color-coded. The central hole of the Lsm1-7 ring is partially blocked by a characteristic α -helix from the C-terminus of Lsm1. The crystallized Lsm1-7 complex contains five full-length Lsm proteins (Lsm2/3/5/6/7), residues 1-93 of Lsm4, and residues 30-173 of Lsm1. To facilitate crystallization, Cys45 of Lsm2 and Cys37/Cys63 of Lsm3 were mutated to Ser. **(B)** The Lsm1 subunit adopts an extended appearance in the Lsm1-7 ring. Unlike the other six Lsm subunits, the N- and C-terminal sequences of Lsm1 have well-defined structures that extend from the conserved Sm core fold. **(C)** Residues from the C-terminal sequence of Lsm1 interact with Lsm3 through hydrogen bonds (H-bonds) and van der Waals contacts. **(D)** Residues from the C-terminal sequence of Lsm1 interact with Lsm6 through van der Waals contacts. **(E)** Asp36 of Lsm1 forms a salt bridge with Lys8 of Lsm2. **(F)** Structural overlay of the Lsm1-7 and Lsm2-8 complexes. These two complexes exhibit an RMSD of 0.89 Å over 461 aligned C α atoms. Components of the Lsm1-7 ring are color-coded, whereas the Lsm2-8 ring (PDB code 4M78) is shown in grey. **(G)** Structural comparison between Lsm1 and Lsm8. **(H)** The Lsm1-7 ring binds to RNA sequences with micromolar affinities. Lsm1-7 only exhibits 3-fold lower binding affinity for the octa-nucleotide 5'-AAAAAAA-3' compared to 5'-UUUUUUU-3'. By contrast, the Lsm2-8 complex preferentially binds the oligo-U element but exhibits a similar binding affinity for the oligo-A sequence as Lsm1-7. **(I)** Structural alignment of the conserved Arg residues between Lsm1-7 and Lsm2-8 (PDB code 4M7A). The residues in Lsm1-7 are colored, whereas the residues in Lsm2-8 are shown in grey. **(J)** Biochemical analysis of RNA binding by variants of the Lsm1-7 complex. The K_D values in **H** and **J** were measured by Biolayer Interferometry (BLI) on a Fortebio Octet system. Each experiment was independently repeated three times, with the average value and standard deviation shown here.

ring, with the ensuing loop associating with Lsm3 and Lsm6. The central hole of the Lsm1-7 ring is partially blocked by this α -helix from Lsm1, leaving a very small opening at the center of the ring. This structural feature contrasts that in the Lsm2-8 complex, where the central hole has a diameter of 15 Å [8].

Unlike the other six Lsm components, the N- and C-terminal sequences of Lsm1 have well-defined structures that extend out from the conserved Sm core fold (Figure 1B) and interact with neighboring Lsm subunits in the heptameric Lsm1-7 complex. Notably, a few amino acids from the C-terminal sequence of Lsm1 interact with Lsm3 and Lsm6 through hydrogen bonds (H-bonds) and van der Waals contacts (Figure 1C and 1D). In addition, Asp36 at the N-terminal loop of Lsm1 interacts with Lys8 of Lsm2 through a salt bridge (Figure 1E). These specific interactions likely constitute the basis why Lsm1, but not any other of the Lsm proteins, can replace Lsm8 to form a heptameric complex with Lsm2/3/4/5/6/7.

The Lsm1-7 complex shares six common components with the heptameric Lsm2-8 complex, with Lsm1 in the Lsm1-7 complex substituted by Lsm8 in the Lsm2-8 complex. Consequently, the overall structures of these two complexes are very similar, with a root-mean-squared deviation (RMSD) of 0.89 Å over 461 aligned C α atoms (Figure 1F). The main difference is between Lsm1 and Lsm8, which can be aligned to each other with an RMSD of 0.97 Å over 54 aligned C α atoms in the core Sm fold (Figure 1G). Notably, these 54 aligned C α atoms come from residues 32-115 of Lsm1 and residues 1-67 of Lsm8.

Although the N- and C-terminal sequences of Lsm1 mediate interactions that are unique to the Lsm1-7 complex, the core Sm fold of Lsm1 appears to make a conserved set of interactions compared to Lsm8. Specifically, interactions between Lsm1 and Lsm2, including both main chain H-bonds and side chain van der Waals

contacts, are similar to those between Lsm8 and Lsm2. The majority of the Lsm1-Lsm4 interactions are also preserved in the Lsm8-Lsm4 interface (Supplementary information, Figure S4).

The Lsm1-7 complex plays an important role in RNA metabolism by facilitating degradation of mRNA. In yeast, the Lsm1-7 complex binds the 3'-end of mRNA and localizes to the P-bodies, ultimately resulting in mRNA degradation from the 5'-end [4]. We examined the RNA binding affinities of the Lsm1-7 complex using biolayer interferometry (BLI) on a Fortebio Octet system. The Lsm1-7 complex exhibits binding affinities of approximately 6 μ M and 2 μ M for the octa-nucleotides 5'-AAAAAAA-3' and 5'-UUUUUUU-3', respectively (Figure 1H). The binding affinity difference between the oligo-A and oligo-U sequences is only about 3-fold, suggesting that the Lsm1-7 complex may not recognize RNA sequences with the same level of specificity as the Lsm2-8 complex [8]. By contrast, the Lsm2-8 complex exhibits a binding affinity of about 20 nM for the oligo-U sequence, approximately 200-fold tighter than that for the oligo-A sequence (Figure 1H). These results suggest that the RNA binding mode of the Lsm1-7 complex may be similar to that of the Lsm2-8 complex towards the oligo-A, but not the oligo-U, sequence.

Although Lsm1 and Lsm8 share limited sequence similarity, the RNA-binding residues in Lsm8 are largely preserved in Lsm1. For example, the D₃₁X₃₂N₃₅ and R₅₇GX motifs in Lsm8 correspond to D₇₂X₇₃N₇₆ and R₁₀₅GX in Lsm1, respectively. In addition, modeling studies suggest that the Lsm1-7 complex can bind to the 3'-end of oligo-U sequence similarly as the Lsm2-8 complex. However, the oligo-U binding affinity of Lsm1-7 is approximately 100-fold lower than that of Lsm2-8 (Figure 1H). Obviously, the sequence variation between Lsm1 and Lsm8 helps determine their RNA binding affinity as

well as specificity. The molecular basis for this observation is likely to be revealed by the crystal structure of Lsm1-7 bound to RNA.

An invariant Arg residue in Lsm6/3/2/8/4 plays an important role in RNA recognition by the Lsm2-8 complex [8] (Figure 1I). We generated seven variants of the Lsm1-7 complex, each containing replacement of the conserved Arg (Ser in Lsm5) by Ala in one Lsm component, and examined their interactions with the oligo-A and oligo-U RNA elements (Figure 1J). As recently reported [8], mutation of the conserved Arg to Ala in any of the Lsm6/3/2/8/4 subunits results in drastic reduction of binding affinity between the 3'-end U-rich sequence of U6 snRNA and the Lsm2-8 complex. In the Lsm1-7 complex, such point mutation in Lsm2 or Lsm3 caused the most pronounced reduction of RNA binding affinity to the oligo-U sequence (Figure 1J). By sharp contrast, such point mutation only caused modest changes (less than 2-fold) in binding affinities for the oligo-A sequence. This result is consistent with the observation that the Lsm1-7 complex exhibited a weak specificity for the oligo-U sequences (Figure 1H).

In summary, we report the crystal structure of the heptameric Lsm1-7 complex and preliminary characterization of its RNA-binding properties. Our structural and biochemical characterization serves as a framework for mechanistic understanding of the function of Lsm1-7 complex in RNA metabolism. In the final phase of manuscript preparation, we noted publication of two related manuscripts [14, 15], of which Sharif *et al.* [14] reported the crystal structure of Lsm1-7 without RNA binding studies.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31130002 and 31021002). The atomic coordinates have been deposited in the Protein Data Bank with the accession code 4M75.

Lijun Zhou^{1,3,*}, Yulin Zhou^{1,*}, Jing Hang^{2,3},
Ruixue Wan², Guifeng Lu², Chuangye Yan^{2,3},
Yigong Shi^{1,3}

¹Ministry of Education Key Laboratory of Protein Science, ²State Key Laboratory of Bio-membrane and Membrane Biotechnology, ³Tsinghua-Peking Joint Center for Life Sciences, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China

*These two authors contributed equally to this work.

Correspondence: Yigong Shi

E-mail: shi-lab@tsinghua.edu.cn

References

- 1 Kambach C, Walke S, Young R, *et al.* *Cell* 1999; **96**:375-387.
- 2 Moore MJ, Query CC, Sharp PA. Splicing of precursors to messenger RNAs by the spliceosome. In: Atkins RGJ. eds. *The RNA World*. New York: Cold Spring Harbor Laboratory Press, 1993:303-357.
- 3 Achsel T, Brahm H, Kastner B, *et al.* *EMBO J* 1999; **18**:5789-5802.
- 4 Tharun S, He W, Mayes AE, *et al.* *Nature* 2000; **404**:515-518.
- 5 Tucker M, Valencia-Sanchez MA, Staples RR, *et al.* *Cell* 2001; **104**:377-386.
- 6 Tharun S. *RNA Biol* 2009; **6**:228-232.
- 7 Fillman C, Lykke-Andersen J. *Curr Opin Cell Biol* 2005; **17**:326-331.
- 8 Zhou L, Hang J, Zhou Y, *et al.* *Nature* 2013; **506**:116-120.
- 9 Naidoo N, Harrop SJ, Sobti M, *et al.* *J Mol Biol* 2008; **377**:1357-1371.
- 10 Mund M, Neu A, Ullmann J, *et al.* *J Mol Biol* 2011; **414**:165-176.
- 11 Zaric B, Chami M, Remigy H, *et al.* *J Biol Chem* 2005; **280**:16066-16075.
- 12 Scheich C, Kummel D, Soumailakakis D, *et al.* *Nucleic Acids Res* 2007; **35**:e43.
- 13 Collins BM, Harrop SJ, Kornfeld GD, *et al.* *J Mol Biol* 2001; **309**:915-923.
- 14 Sharif H, Conti E. *Cell Rep* 2013; **5**:283-291.
- 15 Wu D, Muhrad D, Bowler MW, *et al.* *Cell Res* 2014; **24**:233-246.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)



This work is licensed under the Creative Commons Attribution-NonCommercial-Share Alike Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>