

# The sole LSm complex in *Cyanidioschyzon merolae* associates with pre-mRNA splicing and mRNA degradation factors

KIRSTEN A. REIMER,<sup>1,9</sup> MARTHA R. STARK,<sup>1</sup> LISBETH-CAROLINA AGUILAR,<sup>2</sup> SIERRA R. STARK,<sup>1</sup> ROBERT D. BURKE,<sup>3</sup> JACK MOORE,<sup>4</sup> RICHARD P. FAHLMAN,<sup>4,5</sup> CALVIN K. YIP,<sup>6</sup> HARUKO KUROIWA,<sup>7</sup> MARLENE OEFFINGER,<sup>2,8</sup> and STEPHEN D. RADER<sup>1</sup>

<sup>1</sup>Department of Chemistry, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada

<sup>2</sup>Laboratory of RNP Biochemistry, Institut de Recherches Cliniques de Montréal (IRCM), Faculty of Medicine, McGill University, Montreal, QC H3A 0G4, Canada

<sup>3</sup>Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, V8W 3P6, Canada

<sup>4</sup>Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

<sup>5</sup>Department of Oncology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

<sup>6</sup>Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada

<sup>7</sup>Kuroiwa Initiative Research Unit, College of Science, Rikkyo University, Toshima, Tokyo 171-8501, Japan

<sup>8</sup>Département de Biochimie, Université de Montréal, Montréal, QC H2W 1R7, Canada

## ABSTRACT

Proteins of the Sm and Sm-like (LSm) families, referred to collectively as (L)Sm proteins, are found in all three domains of life and are known to promote a variety of RNA processes such as base-pair formation, unwinding, RNA degradation, and RNA stabilization. In eukaryotes, (L)Sm proteins have been studied, inter alia, for their role in pre-mRNA splicing. In many organisms, the LSm proteins form two distinct complexes, one consisting of LSm1–7 that is involved in mRNA degradation in the cytoplasm, and the other consisting of LSm2–8 that binds spliceosomal U6 snRNA in the nucleus. We recently characterized the splicing proteins from the red alga *Cyanidioschyzon merolae* and found that it has only seven LSm proteins. The identities of CmLSm2–CmLSm7 were unambiguous, but the seventh protein was similar to LSm1 and LSm8. Here, we use in vitro binding measurements, microscopy, and affinity purification-mass spectrometry to demonstrate a canonical splicing function for the *C. merolae* LSm complex and experimentally validate our bioinformatic predictions of a reduced spliceosome in this organism. Copurification of Pat1 and its associated mRNA degradation proteins with the LSm proteins, along with evidence of a cytoplasmic fraction of CmLSm complexes, argues that this complex is involved in both splicing and cytoplasmic mRNA degradation. Intriguingly, the Pat1 complex also copurifies with all four snRNAs, suggesting the possibility of a spliceosome-associated pre-mRNA degradation complex in the nucleus.

**Keywords:** *Cyanidioschyzon merolae*; LSm complex; U6 snRNA; pre-mRNA splicing; mRNA degradation; Pat1

## INTRODUCTION

Nuclear pre-mRNA splicing is the eukaryotic process of removing introns from pre-messenger RNA (Berget et al. 1977; Chow et al. 1977). In the stepwise splicing reaction, U1, U2, U4, U5, and U6 snRNAs assemble with proteins to form discrete small, nuclear ribonucleoproteins (snRNPs) that assemble on the pre-mRNA and catalyze the splicing reaction (Wahl et al. 2009). In addition to snRNA-specific proteins, four snRNPs (U1, U2, U4, and U5) contain a common

heteroheptameric Sm protein complex that binds to the 3' end of the snRNA (Lerner and Steitz 1979). In contrast, U6 associates with a heteroheptameric complex of Sm-like (LSm) proteins (Séraphin 1995).

(L)Sm proteins form a variety of RNA-binding complexes in Eukaryotes and Archaea (Wilusz and Wilusz 2013). Nine different LSm proteins have been identified in yeast, eight of which form two major complexes: the LSm 2–8 proteins form a complex involved in pre-mRNA splicing (Mayes et al. 1999; Salgado-Garrido et al. 1999), and the LSm 1–7

<sup>9</sup>Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

Corresponding author: [rader@unbc.ca](mailto:rader@unbc.ca)

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.058487.116>.

© 2017 Reimer et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

proteins form a similar complex involved in mRNA degradation (Tharun et al. 2000). These two distinct complexes share six of the seven subunits (LSm 2–7); however, they have dramatically different roles and cellular localizations (splicing in the nucleus and mRNA degradation in the cytoplasm). The splicing-associated LSm complex binds U6 snRNA (Vidal et al. 1999), whereas the mRNA degradation complex is found to be associated with enzymes recruited for mRNA degradation, namely Pat1, Xrn1, Dhh1, Edc3, Edc4, Scd6, and Dcp1/2 (Bouveret et al. 2000; Franks and Lykke-Andersen 2008; Fromm et al. 2012; Cary et al. 2015). In addition, LSm proteins have been found to interact with U8 snRNA in *Xenopus* (Tomasevic and Peculis 2002) and other small RNAs (Fischer et al. 2010), and have been implicated in pre-tRNA and pre-rRNA processing (Beggs 2005) and telomerase RNA processing (Tang et al. 2012). The presence of LSm and Sm-like proteins in eukarya, archaea, and bacteria (which contain the Sm-motif-containing Hfq complex), as well as their wide variety of functions, indicates that (L)Sm proteins are important in modulating several aspects of RNA and RNP biogenesis.

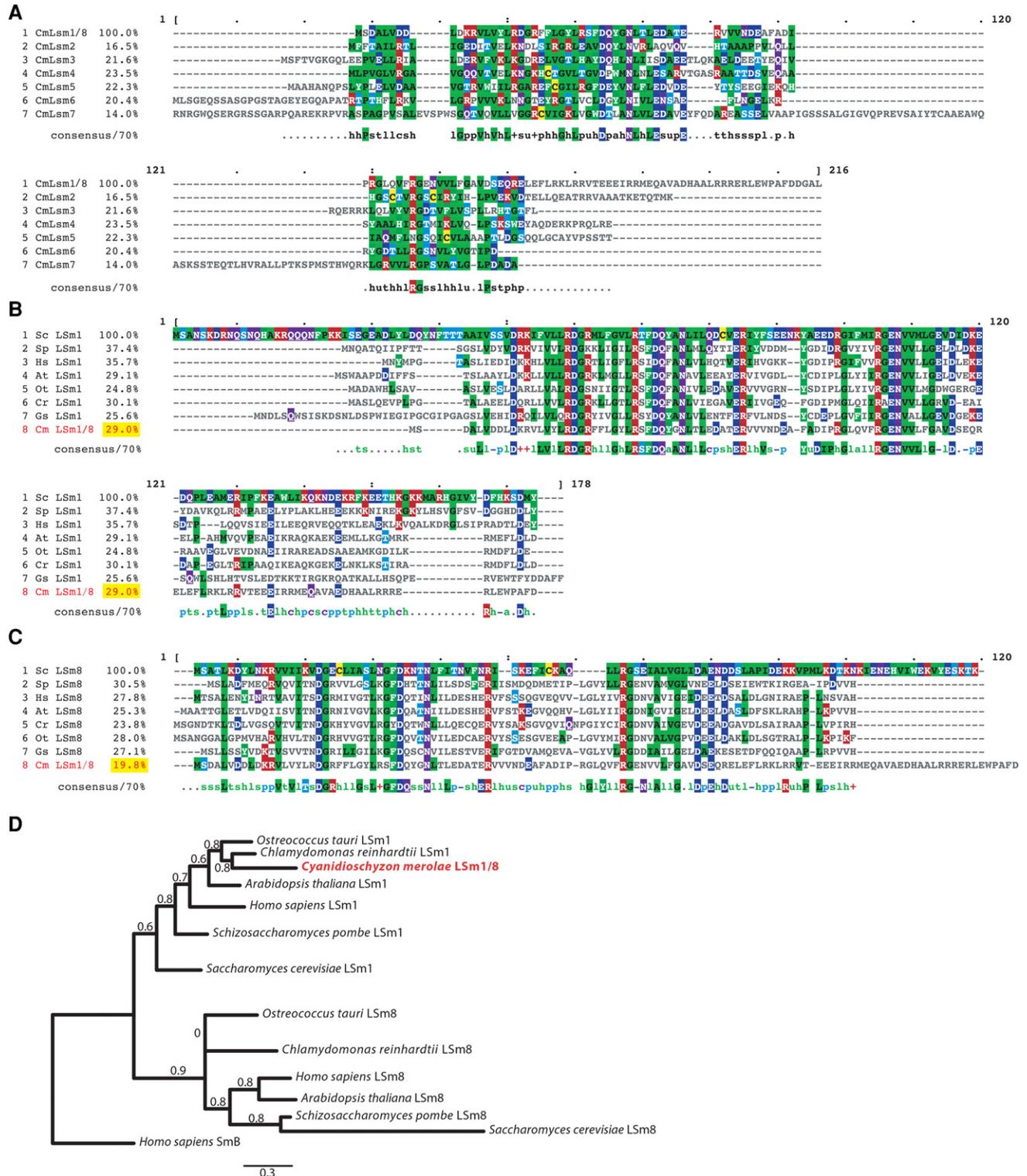
Recently, we reported a dramatically reduced set of splicing components in the red alga *Cyanidioschyzon merolae* (Stark et al. 2015), whose genome had been found to contain only 27 introns (Matsuzaki et al. 2004). We proposed that this organism offers a more tractable system for studying the complex process of splicing, as it harbors only 31 proteins predicted to assemble into snRNPs. Furthermore, we found few snRNP biogenesis factors, and a startling absence of the U1 snRNA and U1-associated proteins. Interestingly, we found only seven LSm proteins, in contrast to the eight or more LSm proteins found in other eukaryotes. This suggests that only one LSm complex forms in *C. merolae*. We were able to unambiguously identify the CmLSm 2–7 subunits by sequence comparison; however, the remaining subunit showed similarity to LSm1 and LSm8. Thus, it was unclear whether the CmLSm complex is involved in splicing or in mRNA degradation.

In order to determine the function of this singular LSm complex, we investigated its association with the U6 snRNA, which would indicate a role in splicing. Here, we show that recombinantly purified *C. merolae* LSm complex binds *C. merolae* U6 snRNA in vitro. We report that immunoprecipitating the LSm complex copurifies U6 snRNA along with many other splicing proteins from *C. merolae* extract, and that in the reciprocal experiment, U6 snRNA pull-down copurifies the LSm proteins. These data, in combination with the observation of a nuclear fraction of LSm proteins, support a splicing function for the CmLSm complex. Nevertheless, we also observed the Pat1-associated mRNA degradation complex, not only in CmLSm immunoprecipitation, but also in all of the snRNA pull-downs. Together with a clear cytoplasmic fraction of CmLSm proteins, this supports an mRNA degradation function for the CmLSm complex.

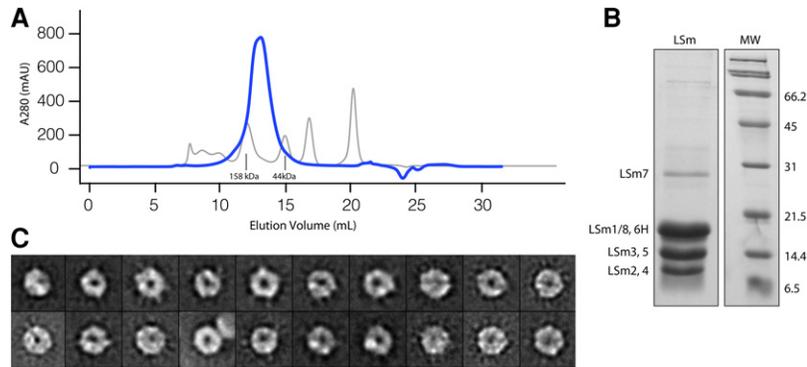
## RESULTS

While looking for splicing proteins in *C. merolae*, we identified CmLSm proteins 2–7 as the top hits from human homologs (Fig. 1A). The alignment highlights the conservation of known residues in the Sm motif (Cooper et al. 1995; Séraphin 1995); however, BLAST searches were unable to clearly distinguish whether the remaining protein was LSm1 or LSm8. In order to determine which protein the CmLSm1/8 candidate was most similar to, and therefore which LSm function the *C. merolae* proteins would be implicated in, we aligned the sequence of the CmLSm1/8 candidate with LSm1 and LSm8 protein sequences from other organisms (Fig. 1B,C). The CmLSm1/8 protein showed greatest similarity to the LSm1 proteins in terms of sequence conservation. For example, the CmLSm1/8 protein is 29% identical to *Saccharomyces cerevisiae* (Sc) LSm1 (Fig. 1B), but only 20% identical to Sc LSm8 (Fig. 1C). To further test the evolutionary relationship of the *C. merolae* protein to LSm1 and LSm8 proteins, we calculated phylogenetic trees with a variety of homologs, using distantly related proteins as outgroups. In all trees calculated, CmLSm1/8 unambiguously segregated with the LSm1 proteins (Fig. 1D). This suggests that the CmLSm complex is more similar to the cytosolic LSm1–7 complex involved in mRNA degradation, leaving open the question of whether these proteins have any role in pre-mRNA splicing. If the CmLSm complex is not associated with U6 during splicing, however, U6 would be predicted to have no associated proteins (since *C. merolae* lacks the canonical U6 snRNP protein Prp24 [Stark et al. 2015]). We therefore hypothesized that, even though the composition of the CmLSm complex appeared more similar to the mRNA degradation complex, the CmLSms are nevertheless associated with U6 snRNA.

In order to address the function of the CmLSm complex in vitro, we expressed and reconstituted the recombinant CmLSm complex from *Escherichia coli*. We generated an expression vector for the seven CmLSm genes using the pQLink-based expression system we previously developed for the yeast LSm complex (Dunn 2014). A peak eluted from a gel filtration column in a volume intermediate to the 158 kDa and 44 kDa size standards, consistent with the predicted complex mass of 92 kDa (Fig. 2A). SDS-PAGE showed bands corresponding to four unique sizes (Fig. 2B). Since several of the CmLSm subunits are close in size, we expect bands to comigrate (see annotations at left, Fig. 2B). Human and yeast LSm complexes form a torus (Zaric et al. 2005; Karaduman et al. 2008), so we analyzed the CmLSm complex by negative stain electron microscopy (EM). Two-dimensional analysis revealed that, similar to the human and yeast complexes, the purified CmLSm complex adopts an overall toroidal architecture (Fig. 2C), consistent with its predicted properties. To confirm the composition of the complex, we analyzed the purified sample by mass spectrometry, which showed the presence of all seven expressed



**FIGURE 1.** The putative CmLsm1/8 protein sequence is most similar to Lsm1 proteins. (A) Multiple sequence alignment of *C. merolae* LSm proteins. The Lsm7 sequence (XP\_005537866.1) begins at amino acid 35. Percent identities are normalized by aligned length. Residues are colored by identity and property. The consensus sequence at a 70% threshold is shown below, with symbols as defined in MVIEW (Brown et al. 1998). *C. merolae* Lsm1/8 aligned with (B) Lsm1 proteins and (C) Lsm8 proteins of *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Homo sapiens* (Hs), *Arabidopsis thaliana* (At), *Ostreococcus tauri* (Ot), *Chlamydomonas reinhardtii* (Cr), and *Galdieria sulphuraria* (Gs), formatted as in A. (D) Phylogenetic tree of Lsm1 and Lsm8 sequences, showing that CmLsm1/8 clusters within the Lsm1 sequences. Branch support values are calculated with PhyML (Guindon et al. 2010), and the scale bar indicates the number of amino acid substitutions per site.



**FIGURE 2.** The *C. merolae* LSm proteins associate into a toroidal complex. (A) Gel filtration chromatography of the CmLSm proteins expressed in *E. coli*. The blue line corresponds to the LSm proteins, while the light gray line shows the elution peaks for gel filtration standards. Molecular masses of the closest standards are *below* the corresponding peaks. (B) SDS-PAGE analysis of the CmLSm complex. Identity of the LSm proteins is given at *left* based on mass spectrometric analysis (Supplemental Table S1) (6H = Lsm6 with a His tag). (*Right*) Molecular weight marker sizes in kDa. (C) Representative class averages of CmLSm complexes by negative stain electron microscopy. Each average image represents ~800 particles.

proteins (see Supplemental Material). These observations suggested that the CmLSm complex was organized similarly to other LSm complexes.

To test directly the hypothesized interaction between the CmLSm complex and U6 snRNA, we performed electrophoretic mobility shift assays (EMSA) with the recombinantly purified CmLSm complex and full length, *in vitro* transcribed U6. High concentrations of CmLSm complex (>100 nM) resulted in a quantitative shift of U6 snRNA from the free-to-bound form as detected by native gel electrophoresis (Fig. 3A). To calculate the dissociation constant ( $K_d$ ), we plotted the fraction of bound U6 against the concentration of LSm protein and fit the binding data to the Hill equation as described in Materials and Methods (Fig. 3B). The  $K_d$  for full-length U6 binding the LSm complex was calculated to be  $120 \pm 15$  nM, and the line fit gave a Hill coefficient of  $n = 1.2 \pm 0.2$ . This value is consistent with the LSm complex binding as a single particle, rather than each protein assembling individually onto the RNA.

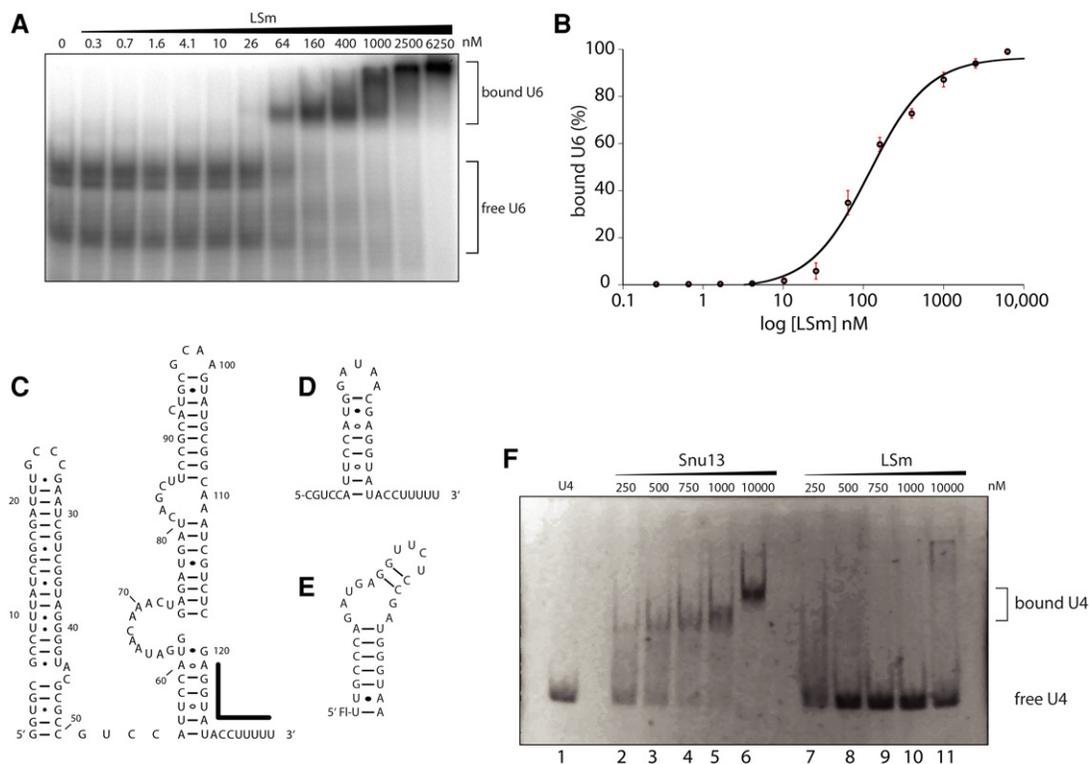
Previous reports have shown that the 3' uridine-rich end of U6 is necessary for LSm binding (Achsel et al. 1999). Similarly, cross-links have been observed in *S. cerevisiae* between the LSm complex and the base of the 3' stem loop of U6 (Karaduman et al. 2006). Both of these potential binding elements are conserved in the predicted secondary structure of CmU6 (Fig. 3C; Stark et al. 2015). In order to investigate whether these sites are important for LSm binding in *C. merolae*, we designed two oligonucleotides corresponding to the above-mentioned regions of U6 (ro62, 3' end: Fig. 3C, highlighted region, and ro63, 3' end+stem: Fig. 3D) and repeated the EMSAs. Increasing concentrations of LSm complex were capable of shifting both oligos from free-to-bound forms (Supplemental Fig. S1). The ro62 oligo gave a  $K_d$  of 150 nM, and the ro63 gave a  $K_d$  of 180 nM (Table 1). These values

indicate that the 3' U-rich end is sufficient for LSm binding, as including residues to encompass more of U6 does not substantially increase the binding affinity of the LSm complex. In contrast to some reports (Licht et al. 2008), but consistent with others (Zhou et al. 2014), the similarity of  $K_d$  values between the oligonucleotides (with 3' OH) and full-length U6 (with a 3' cyclic phosphate) suggests that a 3' cyclic phosphate is not an important determinant for LSm binding in *C. merolae*. These data show that the CmLSm complex binds U6 quantitatively, implying a role for the LSm complex in splicing.

To establish the specificity of the CmLSm complex for U6, we measured the CmLSm complex's binding affinity for a small fragment of U4 snRNA. We did not expect the CmLSm complex to bind this fragment, since it does not con-

tain the canonical uridine-rich LSm binding site. Using an oligo corresponding to the 5' kink-turn of *C. merolae* U4 (ro52; Fig. 3E), we observed no binding interaction at LSm concentrations up to 10  $\mu$ M (Fig. 3F, lanes 7–11). In contrast, we observed binding between the ro52 oligo and CmSnu13 (Fig. 3F, lanes 2–6), as demonstrated previously (Black et al. 2016). Together, these results indicate that the *C. merolae* LSm complex binds specifically, and with high affinity, to U6 snRNA.

In light of these *in vitro* results, we sought evidence for U6: LSm binding in *C. merolae* whole-cell extract. We immunoprecipitated the LSm proteins from extract using anti-LSm polyclonal antiserum raised against the recombinant LSm complex (Fig. 4A, lane 2), but not with non-immune serum (lane 3). The immunoprecipitated proteins comigrate with recombinant CmLSm proteins (lane 1). We extracted RNA from the immunoprecipitated pellet and analyzed the resulting RNA by Northern blotting for all four snRNAs (Fig. 4B). We observed a band of the expected size for U6 in the coimmunoprecipitated RNA (Fig. 4B, lane 5), but not in the non-immune control (lane 3). By comparing RNA from the supernatant (S) and the immunoprecipitated pellet (P), we found that, on average, 42% of U6 in extract was precipitated by the anti-LSm antiserum (Table 2; Fig. 4C). Interestingly, when we probed for the other three *C. merolae* snRNAs, we observed 38% of total U4, 39% of U5, and 12% of U2 in the precipitates. In contrast, the non-immune serum pulled down 2% of U6, 3% of U4, 1% of U5, and 4% of U2 (Table 2). While copurification of U4, U5, and U6 can be explained by their association in the tri-snRNP, the unexpected copurification of U2 snRNA may be due to cross-reactivity of the LSm antiserum with Sm proteins (MR Stark, unpubl.). These results supported our *in vitro* experiments by demonstrating an interaction between U6 and the LSm complex in *C. merolae* extract. The copurification



**FIGURE 3.** The CmLSm complex binds U6 snRNA in vitro. (A) Electrophoretic mobility shift assay with recombinantly purified LSm complex and in vitro transcribed,  $^{32}$ P-labeled U6. Protein concentrations are indicated (*top*) and free U6 is shown in lane 1. (B) U6:LSm binding data (open circles) and line fit (solid line). Error bars are the standard error from three replicates. (C) Predicted secondary structure of *C. merolae* full-length U6 (Stark et al. 2015). The sequence used for the 3'-end oligo (ro62) is highlighted by the dark line. (D) Predicted structure of the 3' end + stem oligo (ro63) corresponding to the base of the U6 stem. (E) Predicted structure of the U4 oligo (ro52) corresponding to the *C. merolae* U4 kink-turn. "Fl" denotes the 5' fluorescein moiety. (F) Fluorescent EMSA with the U4 snRNA oligo. Free U4 oligo is shown in the first lane, with increasing amounts of a known binding partner, CmSnu13, as indicated, and the CmLSm complex at concentrations of 250–10,000 nM.

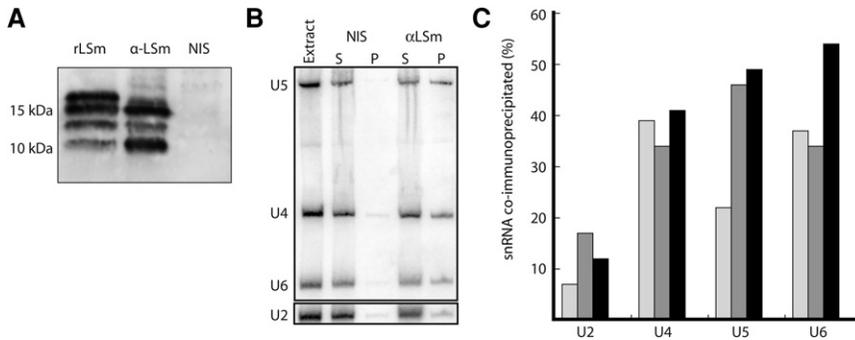
of U2, U4, and U5 further supports our contention that the CmLSm proteins are involved in splicing.

To investigate the composition of particle(s) immunoprecipitated with the LSm antiserum, we analyzed copurified proteins by mass spectrometry (IP-MS). We identified 58 proteins that yielded at least two unique peptides and that were more than twofold enriched relative to the IgG control (Table 3, coIP columns). Aside from the LSm complex, we identified 17 splicing proteins previously found in *C. merolae* (Stark et al. 2015), including six of the seven Sm proteins canonically associated with U2, U4, and U5; U4/U6 snRNP proteins Snu13 and Prp3; U5 proteins Prp8, Brr2, and Snu14; U2 protein Rse1; Sub2 from the A complex; EJC proteins THOC2 and Yra1; the splicing regulator Quaking;

and Pab1 (Table 3). In addition, 11 ribosomal and chloroplast proteins, and 12 proteins with other nonsplicing annotations, were identified that we assumed to be contaminants. Notably, 11 of the detected proteins were not annotated in the genome and not previously identified as part of the spliceosome, and therefore could be splicing proteins that were too divergent to detect in our original analysis (Stark et al. 2015). Of these, seven had no significant BLAST hits, but two, CMR356C and CMS485C, appeared to be homologs of Prp4 and Prp31. BLAST searches with CMR356C yielded known Prp4 homologs in five organisms (Table 4), albeit with E-values above the cutoff we used in our original searches (Stark et al. 2015). BLAST searches with CMS485C yielded Prp31, but not always as the top hit, and with relatively poor E-values (Table 4). While sequence alignments of CMR356C (Supplemental Fig. S2) and CMS485C (Supplemental Fig. S3) only confirm the presence of WD and Nop motifs, respectively (Horowitz et al. 1997; Bizarro et al. 2015), and have remarkably low sequence identities to their *S. cerevisiae* homologs of 14% and 12%, the identification of these proteins as CmPrp4 and CmPrp31 is consistent with their substantial abundance in the list of proteins coimmunoprecipitated with the LSm proteins.

**TABLE 1.** U6 snRNA binding parameters

U6 construct	$K_d$ (nM)	Relative $K_d$	Hill coefficient
Full length	120	1	$1.2 \pm 0.2$
3' end (ro62)	150	1	$2.2 \pm 0.1$
3' end + stem (ro63)	180	2	$1.8 \pm 0.2$
Control (ro52)	>10,000	>100	n/d



**FIGURE 4.** U6 snRNA associates with the LSm complex in *C. merolae* extract. (A) Western blot of CmLSm proteins immunoprecipitated from *C. merolae* whole-cell extract using anti-CmLSm antibodies. Lane 1, 50 ng recombinantly purified CmLSm protein; lane 2, immunoprecipitate from anti-CmLSm serum; and lane 3, immunoprecipitate from non-immune serum (NIS). Numbers at left indicate position of molecular weight standards. (B) Northern blot of coimmunoprecipitated RNA probed for U5, U4, and U6 (top panel), and U2 (bottom panel). Lane 1, total RNA from *C. merolae* whole-cell extract; lanes 2–3, supernatant (S) and pellet (P) from control immunoprecipitation with non-immune serum; lanes 4–5, supernatant (S) and pellet (P) from immunoprecipitation with anti-CmLSm serum. (C) Percentage of U2, U4, U5, and U6 snRNAs coimmunoprecipitated by the anti-CmLSm antiserum ( $n = 3$ ).

One of the annotated, nonsplicing proteins in the IP-MS experiment was Dcp1 (Table 3), which raised the possibility that CmLSm proteins are associated with mRNA degradation machinery as well as splicing factors. As we had not previously sought mRNA degradation proteins in *C. merolae*, we performed searches for the known members of this pathway. These searches revealed clear homologs for Dcp2, Xrn1, and Dhh1, the latter having an astounding 65%–71% identity with homologs (Table 5). We also found good candidates for Edc4 and Scd6, and confirmed the identity of Dcp1, but we were unable to identify homologs of Pat1 or Edc3. In addition to Dcp1, we were therefore able to identify Xrn1, Edc4, and Scd6 in the coimmunoprecipitated proteins (Table 3). We noticed, however, that CMB102C, annotated only as “hypothetical protein,” was the second most highly enriched protein in the IP-MS experiment. PSI-BLAST searches yielded Pat1 homologs, but with high E-values (Table 5). We therefore aligned CMB102C with a number of previously identified Pat1 homologs (Supplemental Fig. S4). The alignments revealed a protein with a comparable length to its homologs, comparable divergence from the *S. cerevisiae* protein, and conserved topo II binding motifs (Wang et al. 1996).

To further test the possibility that CMB102C could be a Pat1 ortholog, we used the LOMETS homology modeling metaserver (Wu and Zhang 2007) to find proteins with whose three-dimensional structures CMB102C’s sequence was consistent. Five of the top 10 results, as ranked by the LOMETS confidence score, corresponded to the C-terminal portion of Pat1 homologs, and the top two had Z-scores of 80, demonstrating high-quality modeling results (Table 6; Wu and Zhang 2007). Structure-based alignment of CMB102C with Pat1 (Supplemental Fig. S5) was slightly different from sequence-based alignment (Supplemental Fig. S4), although they were generally within 10 amino acids of one another.

Figure 5 shows the fit between the CMB102C model and the Pat1 structure (panel A), demonstrating the overall consistency between the CMB102C sequence and Pat1’s three-dimensional fold (Sharif and Conti 2013; Wu et al. 2014). Amino acids predicted to be at the interface between CMB102C and the LSm proteins have similar properties and orientations to those in Pat1’s interface (Fig. 5B). We have therefore provisionally identified CMB102C as the Pat1 ortholog in *C. merolae*.

Based on these observations, we hypothesized that there are two separate CmLSm complexes, one nuclear splicing complex and one cytoplasmic degradation complex, and that in making whole-cell extract they became mixed, resulting in immunoprecipitation of both.

To test this, we used 2’OMe antisense oligonucleotide pull-downs to investigate whether only splicing proteins would copurify with U6 snRNA. Northern analysis of the pull-down showed that we isolated ~70% of U6 from extract, compared to <1% when using a control oligo (Fig. 6, lanes 8 and 2). In addition, ~50% of U4 was isolated, consistent with known base-pairing between *C. merolae* U4 and U6 (Stark et al. 2015). Mass spectrometric identification of U6-associated proteins (2’OMe-MS) revealed the LSms, as expected, as well as U4-associated Sm proteins and Prp3, Prp4, and Prp31, supporting the identification of the latter two proteins in the IP-MS experiment (Table 3). We also observed U5-associated proteins Prp8, Brr2, and Snu114, consistent with the probable existence of tri-snRNP in *C. merolae*, and with the low levels of U5 visible in the Northern blot (Fig. 6, lane 8). We detected a variety of other splicing proteins, including some from the U2 snRNP and various step-specific factors (Table 3). Unexpectedly, we also found all of the mRNA degradation proteins except Dcp1 to be substantially enriched relative to the control.

While the U6–LSm interaction might conceivably reassort during cell lysis and complex purification, resulting in LSm-associated degradation complexes becoming associated with U6, it seemed less likely that this could happen with other snRNAs, particularly U2. We therefore performed pull-

**TABLE 2.** snRNA coimmunoprecipitation with LSm antibodies

snRNA	$\alpha$ LSm IP (%)	SEM	NIS IP (%)	SEM
U2	12	3	4	0
U4	38	2	3	0
U5	39	9	1	0
U6	42	6	2	1

TABLE 3. MS–MS results

Particle or step	Protein	MW (kDa)	Accession number	colP		2'O-methyl pull-down				
				IgG	Lsm	Control	U6	U4	U5	U2
U6	Lsm1	15	gi 544218471	0	144	9	75	25	33	36
	Lsm2	11	gi 544209633	0	71	0	25	12	12	11
	Lsm3	12	gi 544218259	0	48	7	22	14	12	14
	Lsm4	11	gi 544210944 gi 544218709	0	48	4	16	9	10	0
	Lsm5	11	gi 544215363	0	18	0	11	2	6	4
	Lsm6	11	gi 544215335	0	39	0	13	10	7	7
	Lsm7	22	gi 544215441	0	67	19	54	32	38	33
Sm	SmB	9	gi 544212616	0	7	0	7	3	0	15
	SmD1	15	gi 544210697	0	9	0	23	17	11	45
	SmD2	36	gi 544214527	0	18	0	28	18	13	94
	SmD3	19	gi 544213674	0	19	0	41	21	15	108
	SmE	12	gi 544211511 gi 544213736	0	10	0	15	11	11	60
	SmF	10	gi 544215924	0	0	0	0	3	0	19
	SmG	11	gi 544215110	0	5	0	7	3	0	36
U4/U6	Prp3	59	gi 544218113	0	85	0	97	30	0	8
	Snu13	16	gi 544215625	2	9	7	12	16	15	9
	Prp4	51	gi 544216891	0	33	0	64	25	0	0
	Prp31	41	gi 544217836	0	18	3	40	3	3	2
U5	Prp8	274	gi 544211441	0	5	0	10	16	61	3
	Brr2	205	gi 544213359	0	3	2	17	15	51	12
	Snu114	122	gi 544212916	0	4	0	6	3	51	4
U2	Prp9	60	gi 544216280	0	0	0	4	0	12	269
	Prp11	19	gi 544211339 gi 544214213	0	0	0	0	0	3	73
	Prp21	50	gi 544212559	0	0	0	4	2	6	203
	Hsh155	104	gi 544209427	0	0	0	9	10	11	397
	Rse1	179	gi 544213209	0	3	9	48	43	17	198
	Hsh49	14	gi 544210354	0	0	0	0	0	0	100
	Cus1	29	gi 544218405	0	0	0	0	0	0	60
	Rds3	14	gi 544217118	0	0	0	0	0	0	12
	Prp5	113	gi 544217005	0	0	0	8	12	28	66

Continued

TABLE 3. Continued

Particle or step	Protein	MW (kDa)	Accession number	colP		2'O-methyl pull-down				
				IgG	LSm	Control	U6	U4	U5	U2
U2 related	Prp43	83	gi 544213652			0	0	4	3	3
	Mud2	143	gi 544217774			0	10	5	40	21
Complex A	Sub2	51	gi 544210372	5	12					
NTC	Cef1	47	gi 544216515			10	14	14	15	39
	Prp46	46	gi 544216817			0	8	8	8	22
	Bud31	27	gi 544210870			0	0	0	4	0
Complex B	Prp38	21	gi 544212271			0	7	6	3	0
Complex Bact	Yju2	25	gi 544214477			2	2	3	3	25
	Prp2	77	gi 544210522			0	0	0	11	0
Second step	Prp22	140	gi 544210916			0	8	3	4	19
EJC	THOC2	186	gi 544210920	0	4	43	106	109	160	100
	Fal1	47	gi 544212622			0	3	0	0	0
	Yra1	30	gi 544211387	4	8					
SR	RSp31	34	gi 544214602	5	6	25	16	27	49	69
Misc.	Pab1	104	gi 544212541	10	30	14	76	101	49	98
	Quaking	68	gi 544209332	0	2	11	15	67	9	37
	Rpg1	143	gi 544211285			8	15	19	13	19
	Mtr4	119	gi 544209328	0	4	49	56	115	82	85
	Tub2	52	gi 544214469	0	5					
mRNA degradation	RPSA	32	gi 544218493			0	11	8	10	6
	Pat1	76	gi 544209591	0	113	18	138	75	86	85
	Dhh1	52	gi 544213271			0	7	7	9	2
	Dcp1	46	gi 544213684	0	2 <sup>a</sup>					
	Dcp2	42	gi 544212453			0	3	2	0	2
	Edc4	100	gi 544210815	4	18	16	50	57	35	50
	Scd6	60	gi 544218435	0	9	281	649	515	624	494
	Xrn1	168	gi 544217023	3	23	409	1612	505	1617	798

Blank cells indicate no peptides above threshold in any experiment; colored cells are at least twofold more enriched than the control; boxed cells are at least twofold more enriched than in other 2'O-methyl experiments.

<sup>a</sup>Only in one colP.

**TABLE 4.** BLAST search results for previously unidentified splicing proteins

Cm ID of mass spec protein	Protein query (species)	Cm hit rank	E-value	Reciprocal hit	RBH E-value	Identity
CMR356C	Prp4 (Sc)	24	$8 \times 10^{-7}$	1.Prp4	$9 \times 10^{-7}$	27%
	Prp (Sp)	24	$6 \times 10^{-8}$	1.Prp4	$5 \times 10^{-8}$	27%
	PRP4 (Hs)	24	$7 \times 10^{-7}$	1.PR4	$3 \times 10^{-5}$	21%
	U4-U6 60K (Dm)	25	$4 \times 10^{-8}$	1.U4-U6 60K	$2 \times 10^{-7}$	24%
	Prp4 (Gs)	26	$4 \times 10^{-7}$	1.Prp4	$2 \times 10^{-7}$	26%
CMS485C	Prp31 (Sc)	3	$3 \times 10^{-4}$	1.Nop58	$3 \times 10^{-6}$	
				2.Prp31	$7 \times 10^{-6}$	23%
	Prp31 (Sp)	3	$3 \times 10^{-3}$	1.Nop56	$8 \times 10^{-8}$	
				3.Prp31	$2 \times 10^{-3}$	36%
	PRP31 (Hs)	3	$1 \times 10^{-6}$	1.PR31	$2 \times 10^{-5}$	27%
	Prp31 (Dm)	3	$9 \times 10^{-5}$	1.Nop56	$2 \times 10^{-7}$	
				4.Prp31	$3 \times 10^{-5}$	27%
			1.Prp31	$1 \times 10^{-8}$	25%	
			1.Prp31	$1 \times 10^{-8}$	24%	
CME166C	Prp2 (Sc)	3	$3 \times 10^{-90}$	1.Prp43	$5 \times 10^{-123}$	
				5.Prp2	$3 \times 10^{-99}$	33%
	Cdc28 (Sp)	3	$5 \times 10^{-95}$	1.Prh1	$1 \times 10^{-133}$	
				5.Prp2	$3 \times 10^{-104}$	34%
	DHX16 (Hs)	3	$2 \times 10^{-102}$	1.DHX8	$6 \times 10^{-127}$	
				25.DHX16	$5 \times 10^{-112}$	36%
	lethal(2)37Cb (Dm)	3	$2 \times 10^{-103}$	1.Prp22	$9 \times 10^{-126}$	
			4.lethal(2)37Cb	$7 \times 10^{-111}$	35%	
			1.PR22	$6 \times 10^{-121}$		
			5.ESP3	$2 \times 10^{-118}$	35%	

The rank and identity of the reciprocal hit is noted. RBH E-values are reported for the top hit as well as for the predicted protein when it is not the top hit.

Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Gs, *Galdieria sulphuraria*; Cr, *Chlamydomonas reinhardtii*; At, *Arabidopsis thaliana*.

downs with oligos directed against U4, U5, and U2 snRNAs. Northern analysis again confirmed successful purification of the targeted snRNAs relative to controls (Fig. 6, lanes 4,6,10), with little copurification except for U6 in the U4 pull-down. Mass spectrometric results showed strong enrichment for the expected snRNP-specific proteins, although the Sm proteins were more highly enriched in the U2 pull-down than in any of the others, perhaps due to the higher efficiency snRNA isolation (Table 3). Unexpectedly, we detected substantial enrichment of NTC proteins, particularly with U2, as well as B complex and B<sup>act</sup> complex proteins, and several other miscellaneous splicing proteins. We also observed enrichment of several U2 proteins in the U4, U5, and U6 snRNA pull-downs, despite low levels of U2 snRNA detected on the Northern. Consistent with the existence of a U4/U6.U5 tri-snRNP, U5 proteins copurify with the U4 and U6 snRNAs, but their presence in the U2 pull-down was unexpected. One of the copurified proteins was clearly a DEAH-box helicase, which BLAST searches suggested might be Prp2 (Table 4), although it does not have the N-terminal extension canonically associated with Prp2 (Supplemental Fig. S6; King and Beggs 1990). Notably, mRNA degradation proteins copurified at similar levels with all of the snRNAs, raising the intriguing possibility of a spliceosome-associated RNA degradation complex.

The existence of just one CmLSm complex, along with the unexpected association of Pat1 mRNA degradation proteins with the splicing machinery, suggested that *C. merolae* may not harbor a cytoplasmic fraction of the Pat1 complex. In order to determine its cellular localization, we used anti-LSm antiserum to perform indirect immunofluorescence microscopy on *C. merolae* cells (Fig. 7). The chloroplast of *C. merolae* cells is autofluorescent over a broad range of wavelengths, and is therefore visible in the FITC (green) and TXRED (red) channels (Fig. 7A). We could readily detect the nucleus and chloroplast from DAPI staining in cells with no antiserum treatment (Fig. 7B), and we detected no green autofluorescence signal outside of the chloroplast when the outline of the DAPI signal was superimposed on the FITC (green) image (Fig. 7C,D arrow). As a further test, we digitally subtracted the DAPI (blue) image from the FITC (green) image, which again showed that there is no green signal outside of the chloroplast (Fig. 7E). In contrast, we easily detected a green signal outside the chloroplast in cells stained with anti-LSm antiserum (Fig. 7F). This demonstrated that the antibodies specifically recognized an antigen in *C. merolae* cells and were readily visible above the background autofluorescence.

To determine whether the CmLSm proteins are exclusively nuclear, we compared the FITC (green) signal to the DAPI (blue), as shown in two representative cells (Fig. 7G–N). In

**TABLE 5.** BLAST search results for mRNA degradation proteins

Protein	Organism	Top hit in <i>C. merolae</i> strain 10D	Cm ID	E-value	Identity
<b>Pat1</b>					
PAT1	Hs	ATP-binding cassette sub-family C	CMD133C	1	29%
Pat1	Sc	Transcription factor APF1	CMM052C	$1 \times 10^{-1}$	24%
Pat1	Sp	Probable serine-rich pumilio family RNA-binding domain protein	CMR037C	$1 \times 10^{-2}$	23%
PAT1	At	Hypothetical protein	CMB102C	$1 \times 10^{-1}$	24%
Patr-1	Dm	Probable sodium/hydrogen antiporter	CMS154C	3	25%
<b>Dhh1</b>					
DDX6	Hs	RNA helicase	CML140C	0	67%
Dhh1	Sc	RNA helicase	CML140C	0	67%
Ste13	Sp	RNA helicase	CML140C	0	71%
RH8	At	RNA helicase	CML140C	0	70%
Me31B	Dm	RNA helicase	CML140C	0	65%
<b>Dcp1</b>					
DCP1	Hs	Probable mRNA-decapping enzyme complex component DCP1	CMM070C	$4 \times 10^{-11}$	30%
Dcp1	Sc	Probable mRNA-decapping enzyme complex component DCP1	CMM070C	4	29%
Dcp1	Sp	Probable mRNA-decapping enzyme complex component DCP1	CMM070C	$5 \times 10^{-11}$	35%
DCP1	At	Probable mRNA-decapping enzyme complex component DCP1	CMM070C	$4 \times 10^{-18}$	33%
Dcp1	Gs	Probable mRNA-decapping enzyme complex component DCP1	CMM070C	$4 \times 10^{-4}$	31%
<b>Dcp2</b>					
DCP2	Hs	mRNA-decapping enzyme complex component DCP2	CMJ226C	$9 \times 10^{-50}$	38%
Dcp2	Sc	mRNA-decapping enzyme complex component DCP2	CMJ226C	$1 \times 10^{-42}$	31%
Dcp2	Sp	mRNA-decapping enzyme complex component DCP2	CMJ226C	$5 \times 10^{-52}$	37%
DCP2	At	mRNA-decapping enzyme complex component DCP2	CMJ226C	$3 \times 10^{-53}$	37%
Dcp2	Gs	mRNA-decapping enzyme complex component DCP2	CMJ226C	$1 \times 10^{-65}$	42%
<b>Edc4</b>					
EDC4	Hs	Similar to autoantigen	CMF168C	$5 \times 10^{-14}$	31%
VCS	At	Similar to autoantigen	CMF168C	$2 \times 10^{-10}$	34%
Edc4	Gs	Similar to autoantigen	CMF168C	$3 \times 10^{-14}$	29%
Ge-1	Dm	Similar to autoantigen	CMF168C	$3 \times 10^{-7}$	34%
Edc4	Dr	Similar to autoantigen	CMF168C	$3 \times 10^{-12}$	26%
<b>Xrn1</b>					
XRN1	Hs	Exonuclease	CMQ316C	$4 \times 10^{-152}$	40%
Xrn1	Sc	Deoxyribonuclease	CMR447C	$3 \times 10^{-134}$	35%
Exo2	Sp	Deoxyribonuclease	CMR447C	$1 \times 10^{-157}$	30%
XRN1	At	Deoxyribonuclease	CMR447C	$2 \times 10^{-137}$	39%
Xrn1	Gs	Deoxyribonuclease	CMR447C	0	34%
<b>Edc3</b>					
EDC3	Hs	Tryptophan synthase $\alpha$ chain	CymeCp007	2	32%
Edc3	Sc	Hypothetical protein, conserved	CMM009C	$7 \times 10^{-3}$	20%
Edc3	Sp	Hypothetical protein, conserved	CMM009C	$5 \times 10^{-1}$	24%
Edc3	Dm	Arogenate/prephenate dehydrogenase	CMS326C	$8 \times 10^{-1}$	26%
Edc3	Dr	Similar to GATA transcription factor areB $\gamma$	CMB029C	2	38%
<b>Scd6</b>					
LSm14B	Hs	Hypothetical protein, conserved	CMT375C	$1 \times 10^{-18}$	51%
Scd6	Sc	Hypothetical protein, conserved	CMT375C	$2 \times 10^{-11}$	41%
Sum2	Sp	Hypothetical protein, conserved	CMT375C	$7 \times 10^{-20}$	53%
DCP5	At	Hypothetical protein, conserved	CMT375C	$1 \times 10^{-19}$	55%
Tral	Dm	Hypothetical protein, conserved	CMT375C	$2 \times 10^{-16}$	47%

Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; At, *Arabidopsis thaliana*; Gs, *Galdiaaria sulphuraria*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*.

merged images, the antiserum staining appears to have a diffuse, nuclear portion, as well as a stronger, punctate pattern outside the nucleus (Fig. 7G,K). To test this, we made outlines of the merged chloroplast autofluorescence and DAPI signals (Fig. 7J,N) and superimposed them on the anti-LSm FITC (green) images (Fig. 7I,M). The arrows indicate punctae distributed at least partially outside the nuclear and chloroplast boundaries. As a further test, we digitally

subtracted the DAPI (blue) signal from the anti-LSm (green) signal, again demonstrating substantial LSm staining outside the nuclear boundary (Fig. 7O). Taken together, these data support the presence of LSm proteins in the cytoplasm.

To investigate the localization of CmLSm proteins in more detail, we used immunoelectron microscopy (IEM) with gold-labeled antibodies to assess the distribution of CmLSm proteins in *C. merolae* cells. In transverse sections

**TABLE 6.** LOMETS results for homology models of CMB102C

Rank	Template	Protein (species)	Alignment length	Coverage	Z-score	Identity	Confidence score	Program
1	4n0a_H	Pat1 (Sc)	257	0.366	79.353	0.15	High	HHSEARCH2
2	4ogp_A	Pat1 (Sc)	249	0.355	80.002	0.14	High	HHSEARCH2
3	2xesA0	Pat1 (Hs)	229	0.326	9.721	0.18	Medium	pGenTHREADER
4	4ui9O	APC5 (Hs)	654	0.932	15.113	0.09	Medium	Neff-PPAS
5	1vw1A	TcdA1 (Pl)	678	0.967	4.544	0.14	Medium	PROSPECT2
6	4ogp_A	Pat1 (Sc)	248	0.353	13.539	0.15	Medium	HHSEARCH
7	4fyqa	Aminopeptidase N (Hs)	650	0.927	9.032	0.15	Medium	SP3
8	3jav_A	IP3R1 channel (Rn)	640	0.912	36.400	0.12	Low	FFAS-3D
9	4ogp_A	Pat1 (Sc)	249	0.355	16.386	0.15	Low	HHSEARCH 1
10	5a9q1	NUP160 (Hs)	654	0.932	7.440	0.09	Low	SPARKS-X

Sc, *Saccharomyces cerevisiae*; Hs, *Homo sapiens*; Pl, *Photobacterium luminescens*; Rn, *Rattus norvegicus*.

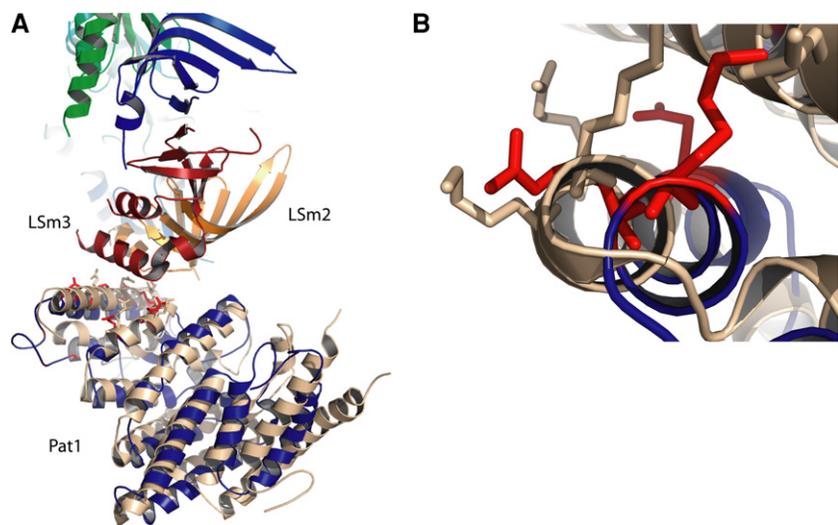
through the nucleus and surrounding cytoplasm (Fig. 8, left panel) and in longitudinal sections through the nucleus, mitochondrion, and chloroplast (Fig. 8, right panel) gold particles were observed in the nucleus (tailed arrows) and cytoplasm (solid arrows), but not in other organelles or cellular compartments. This clearly demonstrates that a substantial proportion of CmLSm proteins are indeed cytoplasmic. Our results support a nuclear function for the CmLSm complex in splicing, while the cytoplasmic fraction is consistent with an LSm-associated mRNA degradation complex.

## DISCUSSION

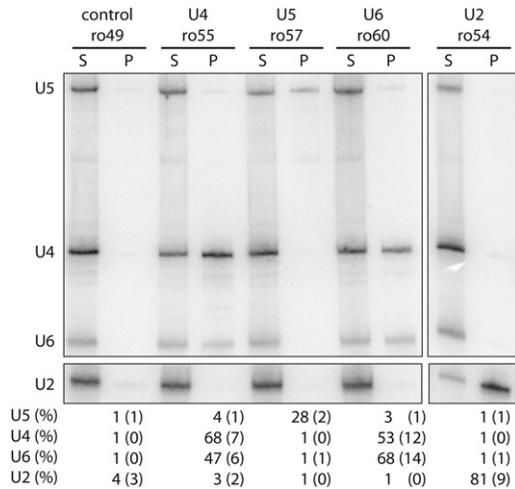
Bioinformatic searches identified seven distinct LSm proteins in *C. merolae* (Stark et al. 2015). LSm4 is encoded by paralogous genes (CMG061c and CMT545c) that differ at only three positions but produce identical proteins. Since most organisms in which LSm proteins have been studied have at least eight LSm proteins belonging to two complexes, we sought to determine whether the CmLSm proteins have a splicing function in the nucleus or an mRNA degradation function in the cytoplasm. Our IP-MS and 2'OMe-MS data, in vitro binding measurements, and microscopy are consistent with the CmLSm complex playing a role in both processes. In addition, the data raise the possibility of a spliceosome-associated RNA degradation complex in *C. merolae*.

Previous work has characterized LSm complexes in other organisms with reduced spliceosomes. In *Leishmania tarentolae*, proteomic analysis in the absence of a sequenced genome revealed the presence of LSm2, LSm3, LSm4, LSm5, and LSm8 (Tkacz et al. 2010). This complex was associated with other splicing components, but not with mRNA degradation

factors. Earlier work in *Trypanosoma brucei* found a single complex, lacking LSm1, that associated with U6, but was not detectable in cytoplasmic P-bodies or stress granules, again arguing against a role in mRNA degradation (Tkacz et al. 2008). In contrast, however, depletion of *T. brucei* LSm8 resulted in increased mRNA stability (Liu et al. 2004), raising the possibility that the single *T. brucei* LSm complex may function in both splicing and mRNA decay. Our microscopic evidence demonstrates the presence of CmLSm proteins in the cytoplasm that could be involved in mRNA degradation. Consistent with this, we found a possible *C. merolae* Pat1 homolog as well as the other known Pat1-associated mRNA degradation proteins (except Edc3) that copurified with the CmLSm complex, and, unexpectedly, with all four snRNAs. Further work will be required to confirm the function of the Pat1 complex in mRNA degradation, as well as to test the possibility of a spliceosome-associated RNA



**FIGURE 5.** Homology modeling of *C. merolae* Pat1 candidate on LSm/Pat1 structure 4N0A. (A) Overlap of CmPat1 model (blue) with Pat1 (wheat) from *S. cerevisiae* (Wu et al. 2014) showing interactions with LSm2 and LSm3. CmPat1 side chains predicted to contact the LSm proteins are highlighted in red. (B) Close-up of Pat1 helix 2a, showing similar position and orientation of CmPat1 (red) and ScPat1 (wheat) side chains that contact the LSm proteins.



**FIGURE 6.** 2′O-methyl antisense oligonucleotide pull-downs of *C. merolae* snRNAs. Northern analysis of snRNAs in pull-down supernatants (S) and pellets (P). The identity of snRNAs on the blot is indicated at left, while the oligo used in the pull-down, and its target snRNA, is indicated above each lane. The average fraction of each snRNA isolated in each experiment is given below ( $n \geq 3$ ) with standard deviations in parentheses.

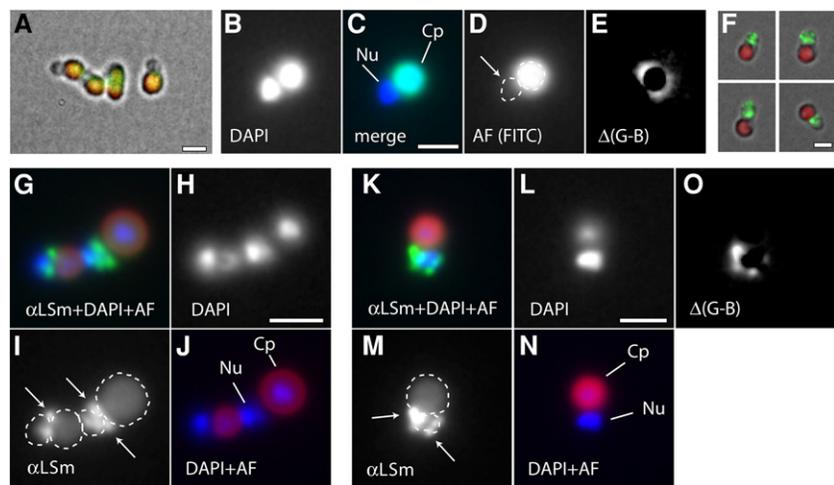
degradation particle. A splicing-associated degradosome would be consistent with evidence for a spliceosomal discard pathway that rejects aberrant pre-mRNA transcripts for degradation (Burgess and Guthrie 1993; Koodathingal et al. 2010; Mayas et al. 2010; Koodathingal and Staley 2013). It would also be consistent with recent evidence for a nuclear Pat1 fraction (Marnef et al. 2012), although experiments in *S. cerevisiae* suggest that nuclear Pat1 is not involved in pre-mRNA degradation (Muppavarapu et al. 2016).

These observations raise the question of how one LSm complex would be targeted to two cellular compartments. In *S. cerevisiae*, no single LSm is sufficient for nuclear exclusion or accumulation, although the N termini of LSm1 and LSm8 play a role in targeting (Reijns et al. 2009). One possibility in *C. merolae* is that cytoplasmic localization of the LSm/Pat1 complex is the default, while targeting to the nucleus occurs via Pat1–snRNP interactions. Interestingly, it has been shown in trypanosomes that Sm proteins can functionally substitute for LSm proteins (Palfi et al. 2000), so it is conceivable that the nuclear CmLSm complex contains an Sm protein responsible for nuclear localization in

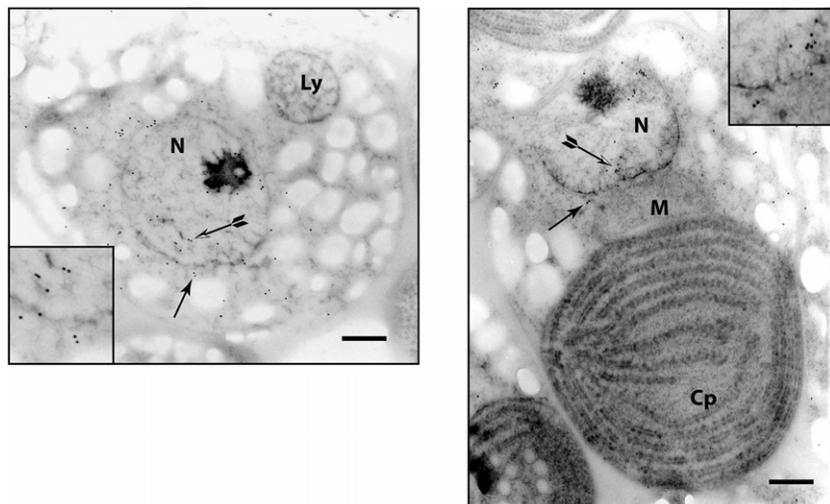
place of LSm8. Our in vitro data, however, make it clear that the CmLSm complex is competent to bind U6 snRNA without the participation of Sm proteins.

The protein identifications presented here empirically substantiate our bioinformatic predictions of splicing proteins in *C. merolae* (Stark et al. 2015). Of the 42 predicted proteins, we found all but Dib1, Msl5/BBP, Prp16, and the cap binding complex proteins Sto1 and Cbc2 enriched at least twofold above background in our copurification experiments. We also observed peripheral, putative splicing proteins, specifically Fal1, Rsp31, Rpg1, Mtr4, and RPSA. In addition, we found candidates for Prp4 and Prp31 that are clearly enriched in U4 and U6 snRNA pull-downs, although their homology with known proteins is tenuous. We observed a protein with some similarity to Prp2 in association with the U5 snRNA, bringing the count of core splicing proteins in *C. merolae* to 45. Notably, despite copurifying the mRNA degradation machinery and several previously unidentified splicing proteins, we failed to observe any U1-associated proteins, supporting our previous conclusion that the U1 snRNP is absent in this organism.

The  $K_d$  value of  $120 \pm 15$  nM that we observed for the interaction between full-length U6 and the CmLSm complex is similar to the reported  $K_d$  of  $52 \pm 7$  nM for the interaction between yeast U6 and LSm2-8 (Zhou et al. 2014). This suggests that the CmLSm complex might interact with U6 in a similar



**FIGURE 7.** CmLSm proteins localize in the nucleus and in bright, cytoplasmic foci. (A) Merged bright field and autofluorescence (green, red) images of control *C. merolae* cells (no primary antisense). (B) DAPI signal from control cells. (C) Merged DAPI and autofluorescence (FITC) images. Nucleus (Nu) and chloroplast (Cp) are indicated. (D) Green autofluorescence with dashed outline of the DAPI signal superimposed. Arrow indicates the absence of signal in the nuclear region. (E) Digitally subtracted image (green-blue) of the same cell as in B–D. (F) Merged bright field, autofluorescent (red), and anti-LSm (green) images from anti-LSm-probed cells. (G) Merged DAPI (blue), autofluorescent (red), and anti-LSm (green) images of two cells. (H) DAPI, (I) anti-LSm, and (J) merge of DAPI plus autofluorescence with organelles indicated as in C. The outline of the signal from the latter image (dashed lines) was superimposed on the anti-LSm image (I), and arrows indicate green (LSm) signal extending beyond the borders of the nucleus and chloroplast. (K–N) As in (G–J). (O) Digital subtraction of the DAPI signal (L) from the anti-LSm signal (M) demonstrating LSm signal outside of the nucleus. Scale bars, 2  $\mu$ m.



**FIGURE 8.** Immunoelectron microscopy confirms cytoplasmic fraction of CmLSm proteins. Transverse (*left*) and longitudinal (*right*) sections of *C. merolae* cells showing nuclear (N), lysosomal (Ly), mitochondrial (M), and chloroplast (Cp) compartments. Gold particles coupled to anti-CmLSm antibodies demonstrate nuclear (tailed arrows) and cytoplasmic (solid arrows) localization. *Insets* show particles on either side of the nuclear membrane in regions highlighted by arrows. Scale bar, 200 nm.

manner to how it does in yeast. Additionally, our CmLSm complex looks comparable to the yeast LSm complex at the level of electron microscopy on negatively stained samples (see Fig. 2 of Achsel et al. 1999). The toroidal appearance of the complex, its purification by a single tag on LSm6, and the Hill coefficient for binding of approximately one all argue for a preformed complex that binds RNA in a single step. Although a putative cross-link between the yeast LSm complex and the U6 stem was found (Karaduman et al. 2006), an oligo containing only the CmLSm binding site bound comparably to an oligo with the adjacent stem ( $K_d = 150$  nM versus 180 nM, respectively), as well as to the complete snRNA ( $K_d = 120$  nM). These results suggest that the CmLSm complex only binds the extreme 3' end of U6, but further work will be required to fully characterize the CmLSm binding determinants.

Our results suggest that the only proteins associated with the U6 snRNP are the LSm proteins. Since *C. merolae* is missing Prp24, which usually associates with the U6 snRNP and promotes U4/U6 di-snRNP formation, this could indicate that the LSm proteins carry out this function. LSm proteins have previously been shown to promote U4/U6 base-pairing in vitro (Achsel et al. 1999). We have shown that U4 and U6 snRNAs are indeed capable of base-pairing in *C. merolae* (Stark et al. 2015), but have yet to investigate whether the presence of LSm proteins enhances this interaction.

## MATERIALS AND METHODS

See Supplemental Material for more details.

## Cyanidioschyzon merolae bioinformatic analysis

We obtained protein sequences for the LSm, Prp4, Prp31, Prp2, and mRNA degradation pathway genes from the NCBI Homologene database and analyzed these sequences using NCBI Protein BLAST to find homologs in *C. merolae* using reciprocal best hit methodology (Ward and Moreno-Hagelsieb 2014; Stark et al. 2015). Protein sequences were aligned using MUSCLE (Edgar 2004) and formatted with MView (Brown et al. 1998). For the phylogenetic tree, sequences were aligned with Clustal Omega (Li et al. 2015), and trees were calculated with PhyML (Guindon et al. 2010), and visualized and edited with FigTree. We used PSI-BLAST (Altschul et al. 1997) to identify homologs of the uncharacterized proteins from *C. merolae* that were enriched in the IP-MS and/or 2'OMe-MS experiments. Additionally, to confirm the identity of the proposed CmPat1 homolog we used the structure modeling (threading) program, LOMETS (Wu and Zhang 2007), visualized with PyMol (Schrödinger).

## LSm cloning, protein preparation, and verification by mass spectrometry

We amplified all seven LSm genes from *C. merolae* genomic DNA by PCR, and combined these genes sequentially into a single coexpression plasmid, pQLink, using ligation-independent cloning (Scheich et al. 2007; Dunn 2014). We expressed the resulting plasmid, with His-tagged LSm6, in Rosetta(DE3)pLysS cells using auto-inducing media ZYM-5052 (Studier 2005), then lysed the cells and purified the protein complex in two steps using nickel affinity chromatography, followed by gel filtration. To confirm by mass spectrometry that all seven LSm subunits were present in the final concentrated sample, the sample was denatured and digested with trypsin. Reactions were quenched with formic acid and loaded directly onto a C18 reverse-phase column for LC-MS/MS.

## Production of CmLSm antibodies and affinity purification

We injected a rabbit with purified CmLSm protein complex in order to generate polyclonal antibodies against the LSm proteins. We affinity purified the anti-LSm antibodies from the crude serum using cross-linked CmLSm complex coupled to sepharose, followed by acid elution (Harlow and Lane 1988).

## Electrophoretic mobility shift assays

EMSA reactions contained  $^{32}$ P-labeled, in vitro transcribed (IVT) U6 snRNA at a final concentration of 10 nM, and LSm protein complex at the concentrations indicated in Figure 3. Following incubation for 15 min at room temperature, we electrophoresed samples on a native polyacrylamide gel at 4°C and imaged the gel on a phosphorimager screen. For U4 binding measurements, we followed the

same procedure as above, except we used a fluorescein-labeled U4 oligo at a final concentration of 100 nM, and visualized the gel with a fluorescent scanner. We used Kaleidagraph (Synergy Software) to fit the data, measured in triplicate, to a modified Hill equation and generate  $K_d$  values.

### Preparation of *C. merolae* whole-cell extract

Extract from the 10D strain of *C. merolae* was prepared following the cryo-grinding method for yeast splicing extract using a mortar and pestle (Ansari and Schwer 1995; Dunn and Rader 2014), or using a planetary ball mill (Trahan et al. 2016), with some modifications. Briefly, we grew cells in MA2 media, harvested cells in log phase, and injected them into liquid nitrogen using a syringe. We ground the cells to a fine powder in the presence of liquid nitrogen so that the grindate remained frozen throughout. Grindate was thawed quickly by the addition of cold lysis buffer. The cell lysate was sonicated briefly and then centrifuged to remove starch and cellular debris. Extract was either used immediately, or glycerol was added to a final concentration of 10% and the extract was snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

### LSm coimmunoprecipitation

To immunoprecipitate the LSm complex from *C. merolae* extract, we cross-linked anti-LSm serum or non-immune serum to Protein A Sepharose using dimethylpimelimidate. For coimmunoprecipitation followed by mass spectrometry (IP-MS) we conjugated affinity purified anti-LSm antibodies or rabbit IgG to magnetic M270 Epoxy Dynabeads (Oeffinger et al. 2007). *C. merolae* extract was incubated with cross-linked beads followed by extensive washing to remove nonspecific proteins. Proteins were eluted with Laemmli buffer for analysis by Western blot, or proteins were removed by treatment with Proteinase K followed by phenol:chloroform extraction, and the RNA was EtOH-precipitated for analysis by Northern blot. For IP-MS, we performed on-bead trypsin digestion of the bound proteins with 750 ng of trypsin in 50  $\mu\text{L}$  of 20 mM Tris-HCl pH 8.0 at  $37^{\circ}\text{C}$  overnight in a thermomixer set at 900 RPM (Gingras et al. 2007). These reactions were quenched with formic acid and then cleaned on a C18 ZipTip prior to loading onto a C18 reverse-phase column for LC-MS/MS analysis as previously described (Trahan et al. 2016).

### 2'O-methyl oligo pull-downs

To isolate the snRNAs and their associated proteins from *C. merolae* extract, we first incubated extract with a biotinylated RNA oligo complementary to a short region of each snRNA, or a control oligo, and then added magnetic beads coated with Neutravidin. Beads were washed extensively and then the same procedures as in the LSm coimmunoprecipitation were followed for Northern blot analysis and mass spectrometry.

### Mass spectrometry

For details on sample preparation, see [Supplemental Methods](#). Raw files were first converted to mzML format using ProteoWizard (v3.0.9322) and the AB SCIEX MS Data Converter (v1.3 beta),

and then searched using Mascot and Comet (v2014.02 rev. 2) search engines. The searches were performed against the RefSeq database release 57 including a decoy set. One missed cleavage was allowed in the search parameters for +2 to 4+ precursor ions with a 10 ppm error tolerance, and a 0.6 Da error tolerance on fragmented ions. The output from each search engine were analyzed through the Trans-Proteomic Pipeline (Deutsch et al. 2010) (v4.7 POLAR VORTEX rev 1) by means of the iProphet pipeline using a 5% FDR (Shteynberg et al. 2011).

### Immunofluorescence microscopy

We synchronized the division of *C. merolae* cells by subjecting them to a 12 h dark–12 h light cycle. We collected cells 10 h into the second light cycle to isolate cells in interphase (Suzuki et al. 1994). Cells were fixed in paraformaldehyde/methanol and permeabilized with Triton X-100. Cells were blocked with BSA and then incubated with anti-LSm serum, followed by a fluorescent secondary antibody. Control reactions were performed as above, with the exception of incubating in  $1\times$  PBS instead of anti-LSm antiserum. We visualized the cells with an Olympus BX61 fluorescence microscope.

### Electron microscopy

We prepared negative stain specimens by adsorbing purified proteins to glow discharged carbon-coated copper grids and staining with uranyl formate. We took images of the specimens on a Tecnai Spirit transmission electron microscope (FEI) and interactively selected particles from the micrographs.

### Immunolectron microscopy

Immunolectron microscopy was performed as described previously ((Yagisawa et al. 2007). Grid sections were incubated with affinity purified anti-LSm antibodies followed by goat antirabbit IgG conjugated with colloidal gold. Samples were stained with uranyl acetate and visualized with an electron microscope.

### SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

### ACKNOWLEDGMENTS

We thank Matt Halstead and Kelly Hrywkiw for initial work on recombinant LSm protein expression. Dr. Takayuki Fujiwara shared important tips for *C. merolae* microscopy. Naomi Fast and Andrew MacMillan provided helpful discussions on this project. Farid Jalali (Olympus) was instrumental in helping us optimize data collection on the fluorescence microscope. We thank the members of the Rader and Oeffinger laboratories for thoughtful discussion of this work. Denis Faubert and his team in the mass spec facility at IRCM provided invaluable assistance and advice. This work was supported by an undergraduate Research Project Award from the UNBC Office of Research (K.A.R.), the Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant 298521 to S.D.R., NSERC Discovery Grant RGPIN-2016-03737 to R.D.B., NSERC Discovery Grant RGPIN 418157-12 to

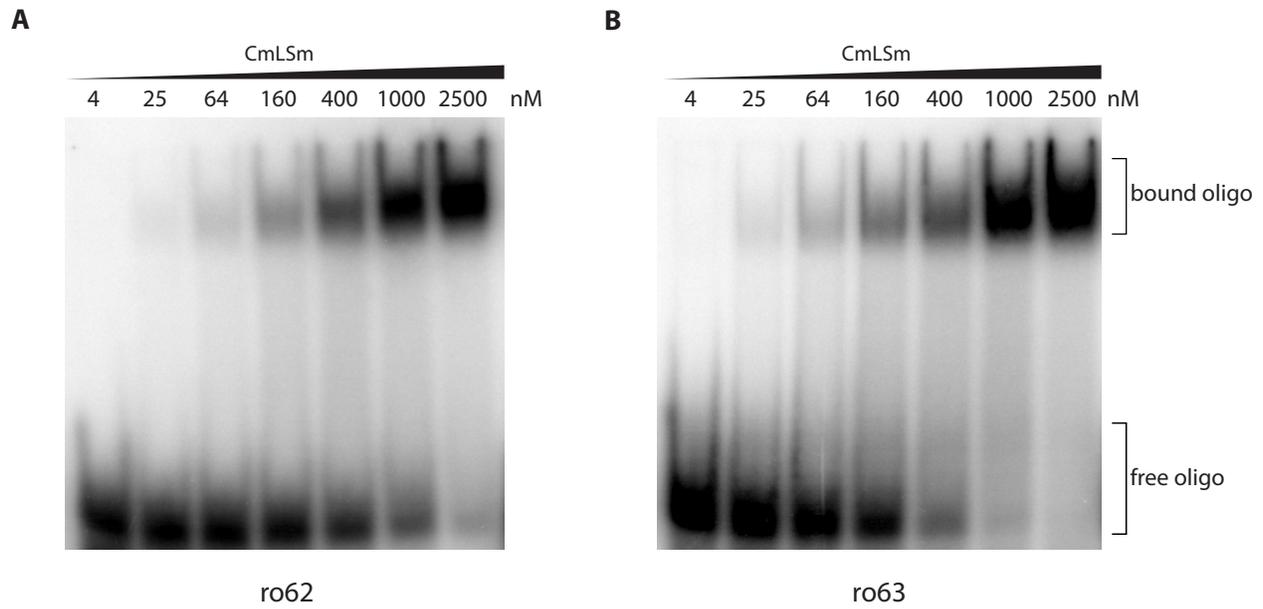
C.K.Y., NSERC Discovery Grant 341453-12 to R.P.F., and NSERC Discovery Grant RGPIN-2015-06586 to M.O.

Received August 15, 2016; accepted March 15, 2017.

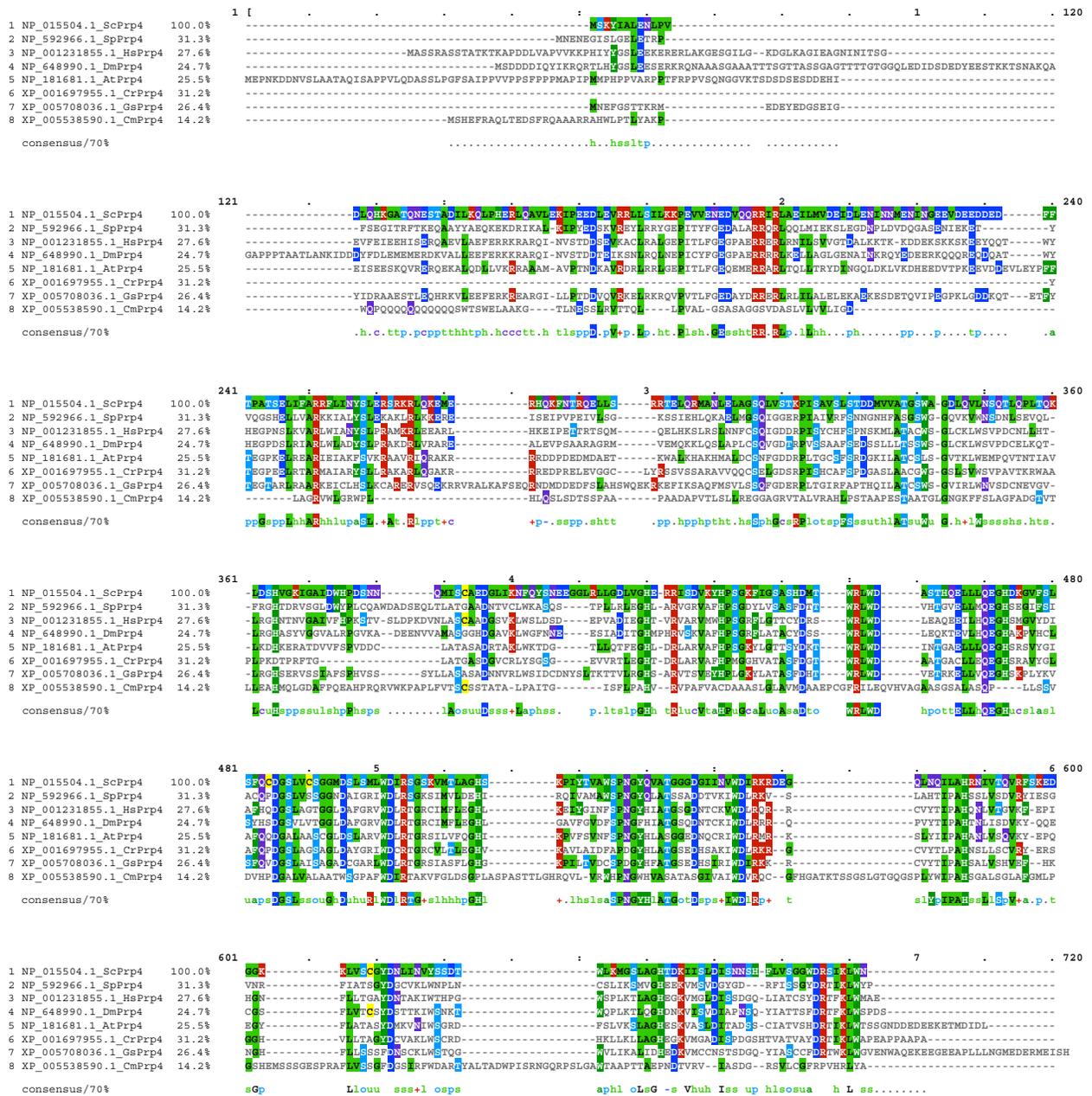
## REFERENCES

- Achsel T, Brahm H, Kastner B, Bachi A, Wilm M, Lührmann R. 1999. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J* **18**: 5789–5802.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Ansari A, Schwer B. 1995. SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J* **14**: 4001–4009.
- Beggs JD. 2005. Lsm proteins and RNA processing. *Biochem Soc Trans* **33**: 433–438.
- Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci* **74**: 3171–3175.
- Bizarro J, Dodré M, Huttin A, Charpentier B, Schlotter F, Brantlant C, Verheggen C, Massenet S, Bertrand E. 2015. NUFIP and the HSP90/R2TP chaperone bind the SMN complex and facilitate assembly of U4-specific proteins. *Nucleic Acids Res* **43**: 8973–8989.
- Black CS, Garside EL, MacMillan AM, Rader SD. 2016. Conserved structure of Snu13 from the highly reduced spliceosome of *Cyanidioschyzon merolae*. *Protein Sci* **25**: 911–916.
- Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B. 2000. An Sm-like protein complex that participates in mRNA degradation. *EMBO J* **19**: 1661–1671.
- Brown NP, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**: 380–381.
- Burgess SM, Guthrie C. 1993. A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell* **73**: 1377–1391.
- Cary GA, Vinh DBN, May P, Kuestner R, Dudley AM. 2015. Proteomic analysis of Dhh1 complexes reveals a role for Hsp40 chaperone Ydj1 in yeast P-body assembly. *G3 (Bethesda)* **5**: 2497–2511.
- Chow LT, Gelinis RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**: 1–8.
- Cooper M, Johnston LH, Beggs JD. 1995. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J* **14**: 2066–2075.
- Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, et al. 2010. A guided tour of the trans-proteomic pipeline. *Proteomics* **10**: 1150–1159.
- Dunn EA. 2014. "Investigation of free U6 small nuclear ribonucleoprotein structure and function." Ph.D. thesis, University of British Columbia, Vancouver.
- Dunn EA, Rader SD. 2014. Preparation of yeast whole cell splicing extract. *Methods Mol Biol* **1126**: 123–135.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Fischer S, Benz J, Späth B, Maier L-K, Straub J, Granzow M, Raabe M, Urlaub H, Hoffmann J, Brutschy B, et al. 2010. The archaeal Lsm protein binds to small RNAs. *J Biol Chem* **285**: 34429–34438.
- Franks TM, Lykke-Andersen J. 2008. The control of mRNA decapping and P-body formation. *Mol Cell* **32**: 605–615.
- Fromm SA, Truffault V, Kamenz J, Braun JE, Hoffmann NA, Izaurralde E, Sprangers R. 2012. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J* **31**: 279–290.
- Gingras A-C, Gstaiger M, Raught B, Aebersold R. 2007. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**: 645–654.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.
- Harlow E, Lane D. 1988. *Antibodies*. Cold Spring Harbor Laboratory, NY.
- Horowitz DS, Kobayashi R, Krainer AR. 1997. A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* **3**: 1374–1387.
- Karaduman R, Fabrizio P, Hartmuth K, Urlaub H, Lührmann R. 2006. RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J Mol Biol* **356**: 1248–1262.
- Karaduman R, Dube P, Stark H, Fabrizio P, Kastner B, Lührmann R. 2008. Structure of yeast U6 snRNPs: arrangement of Prp24p and the Lsm complex as revealed by electron microscopy. *RNA* **14**: 2528–2537.
- King DS, Beggs JD. 1990. Interactions of PRP2 protein with pre-mRNA splicing complexes in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **18**: 6559–6564.
- Koodathingal P, Staley JP. 2013. Splicing fidelity: DEAD/H-box ATPases as molecular clocks. *RNA Biol* **10**: 1073–1079.
- Koodathingal P, Novak T, Piccirilli JA, Staley JP. 2010. The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. *Mol Cell* **39**: 385–395.
- Lerner MR, Steitz JA. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci* **76**: 5495–5499.
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* **43**: W580–W584.
- Licht K, Medenbach J, Lührmann R, Kambach C, Bindereif A. 2008. 3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through Lsm proteins. *RNA* **14**: 1532–1538.
- Liu Q, Liang X-H, Uliel S, Belahcen M, Unger R, Michaeli S. 2004. Identification and functional characterization of Lsm proteins in *Trypanosoma brucei*. *J Biol Chem* **279**: 18210–18219.
- Marnef A, Weil D, Standart N. 2012. RNA-related nuclear functions of human Pat1b, the P-body mRNA decay factor. *Mol Biol Cell* **23**: 213–224.
- Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima S-Y, Mori T, Nishida K, Yagisawa F, Nishida K, et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**: 653–657.
- Mayas RM, Maita H, Semlow DR, Staley JP. 2010. Spliceosome discards intermediates via the DEAH box ATPase Prp43p. *Proc Natl Acad Sci* **107**: 10020–10025.
- Mayes AE, Verdone L, Legrain P, Beggs JD. 1999. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J* **18**: 4321–4331.
- Muppavarapu M, Huch S, Nissan T. 2016. The cytoplasmic mRNA degradation factor Pat1 is required for rRNA processing. *RNA Biol* **13**: 455–465.
- Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD, Rout MP. 2007. Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat Methods* **4**: 951–956.
- Palfi Z, Lücke S, Lahm HW, Lane WS, Kruff V, Bragado-Nilsson E, Séraphin B, Bindereif A. 2000. The spliceosomal snRNP core complex of *Trypanosoma brucei*: cloning and functional analysis reveals seven Sm protein constituents. *Proc Natl Acad Sci* **97**: 8967–8972.
- Reijns MAM, Auchynnikava T, Beggs JD. 2009. Analysis of Lsm1p and Lsm8p domains in the cellular localization of Lsm complexes in budding yeast. *FEBS J* **276**: 3602–3617.
- Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Séraphin B. 1999. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J* **18**: 3451–3462.

- Scheich C, Kümmel D, Soumailakakis D, Heinemann U, Büsow K. 2007. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res* **35**: e43.
- Séraphin B. 1995. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J* **14**: 2089–2098.
- Sharif H, Conti E. 2013. Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell Rep* **5**: 283–291.
- Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, Tasman N, Mendoza L, Moritz RL, Aebersold R, Nesvizhskii AI. 2011. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics* **10**: M111.007690.
- Stark MR, Dunn EA, Dunn WSC, Grisdale CJ, Daniele AR, Halstead MRG, Fast NM, Rader SD. 2015. Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc Natl Acad Sci* **112**: E1191–E200.
- Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.
- Suzuki K, Ehara T, Osafune T, Kuroiwa H, Kawano S, Kuroiwa T. 1994. Behavior of mitochondria, chloroplasts and their nuclei during the mitotic cycle in the ultramicroalga *Cyanidioschyzon merolae*. *Eur J Cell Biol* **63**: 280–288.
- Tang W, Kannan R, Blanchette M, Baumann P. 2012. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484**: 260–264.
- Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. 2000. Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* **404**: 515–518.
- Tkacz ID, Cohen S, Salmon-Divon M, Michaeli S. 2008. Identification of the heptameric Lsm complex that binds U6 snRNA in *Trypanosoma brucei*. *Mol Biochem Parasitol* **160**: 22–31.
- Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, Leberthal I, Michaeli S. 2010. Analysis of spliceosomal proteins in *Trypanosomatids* reveals novel functions in mRNA processing. *J Biol Chem* **285**: 27982–27999.
- Tomasevic N, Peculis BA. 2002. *Xenopus* LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. *Mol Cell Biol* **22**: 4101–4112.
- Trahan C, Aguilar L-C, Oeffinger M. 2016. Single-step affinity purification (ssAP) and mass spectrometry of macromolecular complexes in the yeast *S. cerevisiae*. *Methods Mol Biol* **1361**: 265–287.
- Vidal VP, Verdone L, Mayes AE, Beggs JD. 1999. Characterization of U6 snRNA-protein interactions. *RNA* **5**: 1470–1481.
- Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701–718.
- Wang X, Watt PM, Louis EJ, Borts RH, Hickson ID. 1996. Pat1: a topoisomerase II-associated protein required for faithful chromosome transmission in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **24**: 4791–4797.
- Ward N, Moreno-Hagelsieb G. 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS One* **9**: e101850.
- Wilusz CJ, Wilusz J. 2013. Lsm proteins and Hfq: life at the 3' end. *RNA Biol* **10**: 592–601.
- Wu S, Zhang Y. 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**: 3375–3382.
- Wu D, Muhrad D, Bowler MW, Jiang S, Liu Z, Parker R, Song H. 2014. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res* **24**: 233–246.
- Yagisawa F, Nishida K, Kuroiwa H, Nagata T, Kuroiwa T. 2007. Identification and mitotic partitioning strategies of vacuoles in the unicellular red alga *Cyanidioschyzon merolae*. *Planta* **226**: 1017–1029.
- Zaric B, Chami M, Rémy H, Engel A, Ballmer-Hofer K, Winkler FK, Kambach C. 2005. Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* **280**: 16066–16075.
- Zhou L, Hang J, Zhou Y, Wan R, Lu G, Yin P, Yan C, Shi Y. 2014. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* **506**: 116–120.



**Figure S1.** The CmLSm complex binds oligonucleotides. **(A)** Electrophoretic mobility shift assay with recombinantly purified LSm complex and  $^{32}\text{P}$ -labelled oligonucleotide ro62 corresponding to the 3' end of U6 as shown in Figure 3C (highlighted by dark line). Protein concentrations are indicated (top). **(B)** As in A, but with oligonucleotide ro63 corresponding to the 3' end of U6 with an adjacent, truncated stem loop as indicated in Figure 3D.

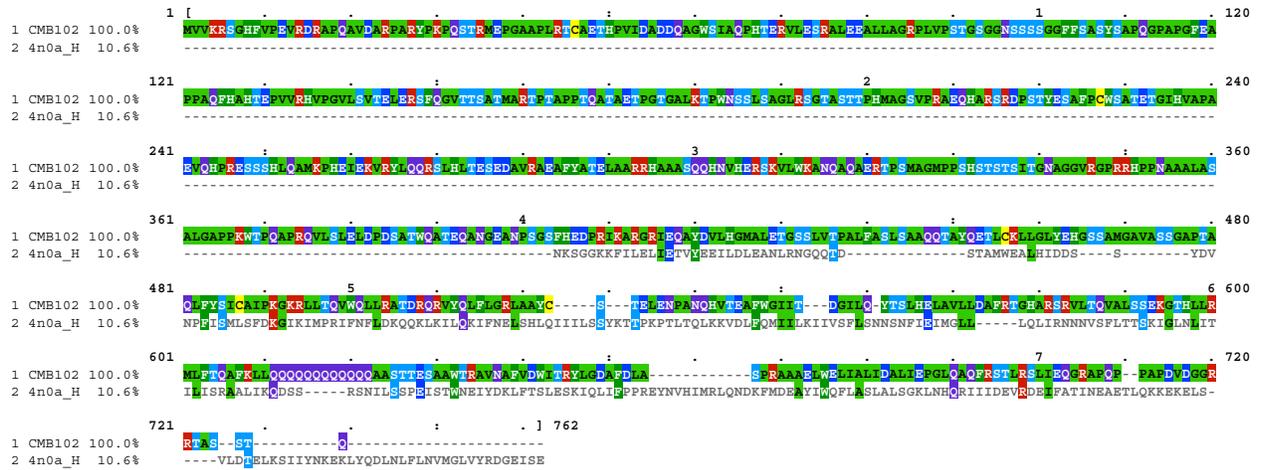


**Figure S2.** MUSCLE (Edgar 2004) alignment of CMR356C with Prp4 sequences from *S. cerevisiae* (Sc), *S. pombe* (Sp), *H. sapiens* (Hs), *D. melanogaster* (Dm), *A. thaliana* (At), *C. reinhardtii* (Cr), and *G. sulphuraria* (Gs). Sequence identity to ScPrp4 is shown at left, consensus sequences at different thresholds calculated by MVIEW (Brown et al. 1998) are shown below the alignment. Sequences are colored by identity and amino acid property





**Figure S4.** Alignment of CMB102C with Pat1 sequences from *S. cerevisiae* (Sc), *S. pombe* (Sp), *H. sapiens* (Hs), *D. melanogaster* (Dm), *A. thaliana* (At), *C. reinhardtii* (Cr), and *G. sulphuraria* (Gs). Sequence identity to ScPrp4 is shown at left, consensus sequences at different thresholds calculated by MVIEW (Brown et al. 1998) are shown below the alignment. Sequences are colored by identity and amino acid property as described at <https://desmid.github.io/mview/manual/manual.html>.



**Figure S5.** HHSEARCH2 (Söding 2005) structure-based alignment of CMB102C with the C-terminal domain of ScPat1 from PDB structure 4N0A (Wu et al. 2014).

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

1 NP\_014408.1\_ScPrp2 100.0%  
2 NP\_595686.2\_SpPrp2 35.4%  
3 NP\_003578.2\_HsPrp2 33.4%  
4 NP\_609946.1\_DmPrp2 39.2%  
5 NP\_174527.2\_AtPrp2 33.4%  
6 XP\_005535662.1\_CmPrp2 29.5%  
consensus/70%

**Figure S6.** Alignment of CME166C with Prp2 sequences from *S. cerevisiae* (Sc), *S. pombe* (Sp), *H. sapiens* (Hs), *D. melanogaster* (Dm), *A. thaliana* (At), *C. reinhardtii* (Cr), and *G. sulphuraria* (Gs). Sequence identity to ScPrp4 is shown at left, consensus sequences at different thresholds calculated by MVIEW (Brown et al. 1998) are shown below the alignment. Sequences are colored by identity and amino acid property as described at <https://desmid.github.io/mview/manual/manual.html>.

**Table S1. Identification of CmLSm proteins by SDS-PAGE mass spectrometry.**

<b>band</b>	<b>dominant protein(s)</b>	<b>Mr (kDa)</b>	<b>other proteins</b>
1	LSm2, LSm3	11.3, 11.8	LSm6, LSm4, LSm1
2	LSm7	21.7	LSm3, LSm2, LSm4, LSm6
3	LSm6	10.7	LSm1
4	LSm3	11.8	LSm2, LSm6
5	LSm2	11.3	LSm3, LSm4

**Table 2. Mass spectrometric data for the recombinantly expressed LSm complex**

protein	score*	coverage	unique peptides	peptide sequences
LSm1 (CMT394)	26	64%	8	GENVVLFGAVDSEQR VVVNDEAFADIPR SFDQYGNLTLEDATER VTEEEIR MEQAVAEDHAALR ERLEWPAFDDGAL RMEQAVAEDHAALR RVTEEEIR
LSm2 (CMB130)	374	77%	10	GRLEAVDQYLNVR VDTELLQEATR LEAVDQYLNVR MFFTAILR YIHLPEK LAQVQVHTAAAPPVLQLLHGSCVTR TLIGEDITVELK YIHLPEKVDTELLQEATR YIHLPEKVDTELLQEATRR VDTELLQEATRR
LSm3 (CMT262)	279	72%	9	AELDEETYEQIVR GQLEEPVELLR GDTVFLVSPLLR KLQLVYVR LQLVYVR ELVGTLHAYDQHLNLIISDAEETLQK KLQLVYVRGDTVFLVSPLLR LQLVYVRGDTVFLVSPLLR GDRELVGTLHAYDQHLNLIISDAEETLQK
LSm4 (CMT545)	36	71%	5	AATDSVEQAASYAALHIR GAVGQQVTVELK HCTGVLTVGDPWMNLNLESAR SWEYAQDER MLPVGLVR
LSm5 (CMP159)	0.5	13%	2	VWIILR EFCGILR

LSm6 (CMP138)	88	73%	13	RYGDTLLR VLGRPVVVK LNNGTEYR YGDTLLR TPTHFLR GSNVLYVGTIPD GTLVCLDGYNIVLENSAEFLNGELK GTLVCLDGYNIVLENSAEFLNGELKR VLGRPVVVKLNNGTEYR RYGDTLLRGSNVLYVGTIPD KVLGRPVVVK KVLGRPVVVKLNNGTEYR YGDTLLRGSNVLYVGTIPD
LSm7 (CMP206)	119	53%	8	EASSELVAAPIGSSSALGIGVQPR GPSVATLGLPDADA ASPAGPVSALVSPWSGQTVQVLLVGGGR EVSAIYTCAAEAWQASK EVSAIYTCAAEAWQASKSSTEQTLHVR ALLPTKSPMSTHWQR SSTEQTLHVR SPMSTHWQR

\* Mass spectrometric parameters were calculated using Proteome Discoverer 1.4.

**Table 3. Mass spectrometry of proteins in purified CmLSm complex.**

<b>protein</b>	<b># PSMs*</b>
LSm2	518
LSm3	473
LSm1	239
LSm7	187
LSm6	70
LSm4	34
LSm5	9
Hfq	5

\*average number of peptide spectral matches (n=2)

**Table S4. Oligonucleotides used in this work.**

<b>oligo</b>	<b>sequence</b>
oSDR919 LSm1/8 FWD	GCCCTTGGATCCATGAGCGACGCGCTGGTCCG
oSDR908 LSm1/8 REV	CCGGAAGTCGACGCGGCCGCTTAGAGGGCACCATCGTCGAAGGCAGG
oSDR907 LSm2 FWD	GGGCTTGGATCCATGTTCTTTACCGCGATTTTGCGGACGC
oSDR910 LSm2 REV	CCGGAAGTCGACGCGGCCGCTTACTTCATGGTCTGGGTTTCCTTCGTCCG
oSDR909 LSm3 FWD	GCCCTTGGATCCATGTCGTTCACTGTCGGAAAAGGGCAATTAGAGG
oSDR912 LSm3 REV	CCGGAAGTCGACGCGGCCGCTTACAGGAAGTTCCTGTATGACGCAAAGCGG
oSDR911 LSm4 FWD	GCCCTTGGATCCATGCTGCCGGTCCGACTCGTCC
oSDR914 LSm4 REV	CCGGAAGTCGACGCGGCCGCTTACTCGCGGAGCTGCCGTGG
oSDR915 LSm5 FWD	GGGCTTGGATCCATGGCGGCACACGCGAACCAGC
oSDR916 LSm5 REV	CCGGAAGTCGACGCGGCCGCTTACGTCGTCGATGAAGGTACATAGGCACATCCGAGCT GCTGAGATCCATCCAGTGTCGGGGCTGCTGC
oSDR913 LSm6 FWD	GCCCTTGGATCCATGCTCAGCGGTGAGCAGTCAAGTGC
oSDR918 LSm6 REV	CCGGAAGTCGACGCGGCCGCTAATCGGGAATCGTCCCTACGTAAAGTACATACTACC
oSDR917 LSm7 FWD	GCCCTTGGATCCATGGACCCGAAACATCCCAGAGATCGG
oSDR920 LSm7 REV	CCGGAAGTCGACGCGGCCGCTTACGCGTCAGCGTCAGGCAGG
oSDR1093 5'HH FWD	TACTTCCAATCCCACG AAGGCGCACC <u>CTGATGAGTCCGTGAGGACG</u>
oSDR1087 5'HH REV	GACGGTACCGGGTACCGTTT <u>CGTCCTCACGGACTCATCAG</u>
oSDR1096 3'HH FWD	ACCGGATGTGCTTCCGGT <u>CTGATGAGTCCGTGAGGACG</u>
oSDR1097 3'HH REV	TTATCCACTTCCCACG AATACCTTTTTT <u>CGTCCTCACGGACTCATCAG</u>
oSDR1094 HH U6 FWD	GTACCCGGTACCGTC <u>GGTGCGCCTT TATCGGCG</u>
oSDR1095 U6 HH REV	GGAAAGCACATCCGGT <u>AAA AAGGTATACCTCGAGACGATTGTCCG</u>
oSDR1127 HH U6 HH FWD	TACTTCCAATCCCACG
oSDR1128 HH U6 HH REV	TTATCCACTTCCCACG
oSDR1092 T7 HH U6 FWD	AAATTTAATACGACTCACTATAGGGA GAAGGCGCACCCCTGATGAG
oSDR1076 3'HH final REV	GGGAGA ATACCTTTTTT <u>CGTCCTCACG</u>

<b>oligo</b>	<b>sequence</b>
ro62 U6 3' end	AGGUAUACCUUUUU
ro63 U6 3'end+stem	CGUCCAUUCCAUGGAUAACGAGGUAAUACCUUUUU
ro52 U4 kink-turn	5'FAM-UUGCCCAGAUGAGGUUCUCCGAUGGGUAA
oSDR604 U2 Northern	GATGCAGGCTCCCTGGAATATAAAATATCCC
oSDR597 U4 Northern	AAATTGTTTGTGTTTCAGCATACCGTT
oSDR768 U5 Northern	GGACACCGCAAGTAAAAGGCATGG
oSDR598 U6 Northern	AAAAAGGTATACCTCGAGACGATTGTC
ro54 U2	mUmCmGmAmUmAmCmAmCmCmAmUmGmAmG-3'Biotin-TEG
ro55 U4	mUmUmGmUmUmUmGmUmGmUmUmCmAmGmC-3'Biotin-TEG
ro57 U5	mCmAmAmGmUmAmAmAmAmGmGmCmAmUmG-3'Biotin-TEG
ro60 U6	mUmUmGmUmUmAmUmCmCmAmUmGmGmA-3'Biotin-TEG
ro49 control	CACGATGACATCmGmGmGmAmAmCmUmGmCmUmGmAmU-3'Biotin

## SUPPLEMENTAL METHODS

**Cyanidioschyzon merolae bioinformatic analysis.** We obtained protein sequences for the LS<sub>m</sub>, Prp4, Prp31, Prp2, and mRNA degradation pathway genes from the NCBI homologue database (<https://www.ncbi.nlm.nih.gov/homologene>) and analyzed these sequences using NCBI Protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to find homologs in *C. merolae* using reciprocal best hit methodology (Ward and Moreno-Hagelsieb 2014; Stark et al. 2015). Additional protein sequences not found in the homologue database were identified by using one of the homologue sequences to search against a specific organism with BLAST. Protein sequences were aligned using MUSCLE (Edgar 2004) and formatted using MView (Brown et al. 1998). For the phylogenetic tree, sequences were aligned with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>; Li et al. 2015), and trees were calculated with PhyML (<http://www.atgc-montpellier.fr/phyml/>; Guindon et al. 2010), and visualized and edited with FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). We used PSI BLAST (Altschul et al. 1997) to identify homologs of the uncharacterized proteins from *C. merolae* that were enriched in the IP-MS and/or 2'OMe-MS experiments. Additionally, to confirm the identity of the proposed CmPat1 homolog we used the structure modeling (threading) program, LOMETS (Wu and Zhang 2007), visualized with PyMol (Schrödinger).

**LS<sub>m</sub> cloning and protein preparation.** We amplified all seven LS<sub>m</sub> genes (*LSm1/8* through *LSm7*; see oligos listed in Supplemental Table S4) from *C. merolae* genomic DNA by PCR, and combined these genes sequentially into a single co-expression plasmid, pQLink (Addgene plasmids 13670, 13667), using ligation-independent cloning. This approach has been described in detail previously (Scheich et al. 2007; Dunn 2014). Individual LS<sub>m</sub>

genes were combined in pairwise steps using the *Swa*I and *Pac*I sites in pQLinkN and T4 DNA polymerase. In this way, we constructed a plasmid containing the seven *C. merolae* genes, each containing a start and stop codon, and with LSm6 containing an N-terminal 6xHis tag (in pQLinkH).

We transformed the pQLink vector containing all seven LSm genes into Rosetta(DE3)pLysS cells (Novagen) and grew in auto-inducing media ZYM-5052 (Studier 2005) for 24 h at 37 C. Cells were harvested, snap frozen in liquid nitrogen, lysed in buffer A (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 20 mM imidazole, 5 mM  $\beta$ -mercaptoethanol) with cOmplete EDTA-free protease inhibitor cocktail (Roche), sonicated, and cleared of cell debris and insoluble proteins by centrifugation. Cell lysate was passed over a HisTrap HP Ni sepharose column (GE Healthcare), washed in buffer A, and eluted in a single step in buffer B (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 500 mM imidazole, 5 mM  $\beta$ -mercaptoethanol). The LSm complex was then loaded onto a gel filtration column (Superdex 200 10/300 GL, GE Healthcare) equilibrated in buffer A without imidazole. Peak fractions were collected, pooled, and concentrated using a YM-30 Centriprep centrifugal filter unit (Millipore). The Superdex 200 column was calibrated using gel filtration standards (BioRad), with the following sizes: thyroglobulin (670 kDa), gamma globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.4 kDa). Protein aggregates come out in the void volume of the column (~8 mL). The purified protein complex was used to generate anti-serum, and for EMSAs.

Hfq is known to contaminate recombinant LSm protein preparations expressed in bacteria (Milojevic et al. 2013). We confirmed the composition of the CmLSm complex by

separating the bands with SDS-PAGE and then subjecting individual bands to mass spectrometry. The recombinant proteins resolved into five bands, which contained the proteins shown in Table S1. We did not detect LSm5 and LSm7. Notably, despite substantial cross-contamination of LSm proteins between bands, no bacterial Hfq was observed. Peptides confirming the identity of each protein present, along with the protein coverage and score are shown in Table S2.

We subsequently assessed the purity of our CmLSm preparation by subjecting the entire sample to mass spectrometry without prior electrophoresis. This time, all seven LSm proteins were detected (Table S3). In this sample, we did detect 6 spectra corresponding to 3 unique peptides from Hfq among the 29 *E. coli* contaminants in the sample. We conclude that our CmLSm prep is relatively free of contaminating Hfq.

**Preparation of U6 with a 2'3'-cyclic phosphate at the 3' end.** We prepared U6 snRNA with homogeneous 5' and 3' ends by *in vitro* transcription (IVT) and subsequent cleavage by 5' and 3' hammerhead ribozymes (Price et al. 1995). Hammerhead ribozymes were made by overlap extension of oligonucleotide pairs (oSDR1093/oSDR1087, oSDR1096/oSDR1097; see Supplemental Table S4), and then overlap extension and gene SOEing (Vallejo et al. 2008) were used to connect the hammerhead ribozymes to the U6 gene which had been amplified by PCR from genomic DNA (oSDR1094/oSDR1095). The underlined sequences in the oligos indicate the regions of complementarity between hammerhead oligo pairs or between the U6 primers and the hammerheads. The final extended/SOEed product was amplified using outside primers oSDR1127 and oSDR1128. Nested primer oSDR1092, containing the T7 promoter sequence, was used with

oSDR1076 to generate the template for IVT. The HiScribe T7 High Yield RNA Synthesis Kit (NEB) was used according to the manufacturer's instructions with 0.5 – 1 µg of template per reaction. After incubation for 4 h at 37 C, we separated the cleaved U6 snRNA from the hammerhead fragments and uncleaved transcript on a denaturing polyacrylamide gel and visualized the RNA by UV shadowing. Gel slices were homogenized in a microcentrifuge tube using a small disposable pestle, mixed with 200 - 400 µL dH<sub>2</sub>O and incubated at 70 C for 10 min before being passed over a DTR column (D-Mark Biosciences) to remove acrylamide. RNA was EtOH precipitated and resuspended in dH<sub>2</sub>O.

**Electrophoretic Mobility Shift Assays.** We prepared U6 snRNA with homogeneous 2',3' cyclic phosphodiester ends by *in vitro* transcription (IVT) of a U6 gene flanked by self-cleaving hammerhead (HH) ribozyme sequences (Price et al. 1995). We purified the cleaved U6 snRNA using denaturing PAGE with UV shadowing, followed by gel extraction and EtOH precipitation. In order to maintain the 2'3'-cyclic phosphate on the 3' end of the full length U6 transcript, we end-labeled it, as well as the two U6 RNA oligos (ro62 and r063), using T4 polynucleotide kinase (PNK) minus 3' phosphatase activity (New England Biolabs) and <sup>32</sup>P-γATP. Unincorporated nucleotides were removed by purifying over a G-25 spin column (SantaCruz Biotechnology) according to the manufacturer's protocol. EMSA reactions were 20 µL containing 12 mM HEPES-NaOH, pH 7.5, 1.5 mM MgCl<sub>2</sub>, 50 mM NaCl, 10% glycerol, 0.1% Triton X-100, and 5 µg *E. coli* tRNA. We added RNA to a final concentration of 10 nM, and LSm protein complex at the concentrations indicated in Figure 3. We incubated the reactions for 15 min at room temperature, then loaded directly

onto a 6% non-denaturing polyacrylamide gel and electrophoresed for 60 min at 200 V, 4 C. For U4 binding measurements, we followed the same procedure as above, except we used a fluorescein-labeled U4 oligo (ro52, IDT) at a final concentration of 100 nM in place of <sup>32</sup>P-labeled U6 snRNA.

Radioactive EMSAs were imaged on a phosphor imager screen overnight and visualized with a Cyclone Phosphor Imager and OptiQuant software (Perkin Elmer). The resulting images were uniformly adjusted for contrast. We used Kaleidagraph (Synergy Software) to fit the data, measured in triplicate, to a modified Hill equation (Equation 1) and generate *Kd* values. Fluorescent EMSAs were scanned with a Kodak Image Station 4000 PRO using Carestream Molecular Imaging Software v.5.0 and quantified as above.

$$\text{Equation 1: } \theta = a + (b - a)/(1 + (Kd/[protein])^n)$$

where  $\theta$  is the fraction of RNA bound, *a* is the minimum asymptote, *b* is the maximum asymptote *Kd* is the equilibrium binding constant, and *n* the Hill coefficient.

**Preparation of *C. merolae* whole-cell extract.** *C. merolae* whole-cell extract was prepared loosely following the cryo-grinding method for yeast splicing extract using a mortar and pestle (Ansari and Schwer 1995; Dunn and Rader 2014), or using a planetary ball mill (Trahan et al. 2016). The 10D strain of *C. merolae* was obtained from the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan ([mcc.nies.go.jp/](http://mcc.nies.go.jp/)) and cultured as described (Stark et al. 2015). Four liters of cells were harvested at OD<sub>750</sub> 2.0 – 3.0 by centrifuging 15 min at 3000 x g. Cell pellets were scraped into a 20 mL syringe and squeezed into liquid nitrogen, forming noodles. Frozen noodles

were ground either with a pestle into a fine powder, adding liquid nitrogen to the mortar, as needed, to keep the cells frozen, or in a Retsch PM-100 planetary ball mill (8 cycles of 3 min each, 400 rpm, 35% power). Four liters of cells yields approximately 10-12 g of grindate.

For small-scale testing in pull down experiments (either colP or 2'O-methyl (2'OMe) oligo) we quickly thawed the powder from 4L of cells by the addition of 75 mL of cold HSWB (20 mM Hepes-KOH, pH 7.9, 300 mM KCl, 10 mM MgCl<sub>2</sub>, 0.1% IGEPAL, 1 mM DTT) plus Roche cOmplete EDTA-free protease inhibitor cocktail, sonicated the cell lysate briefly to break up genomic DNA, then centrifuged 10 min at 50,000 x g. Glycerol was added to a final concentration of 10%. Total protein concentration of the extract was determined by Bradford assay to be approximately 25 mg/mL. Cell extract was aliquoted and snap frozen in liquid nitrogen before storage at -80 C.

For large-scale IP-MS experiments we made the extract immediately before performing the colP, and therefore, did not add any glycerol to the extract. We weighed out 4 g of grindate/colP, resuspended in 36 mL of cold modified RNP buffer (20mM Hepes-KOH, pH 7.4, 110 mM KOAc, 75 mM NaCl, 1.5mM MgCl<sub>2</sub>, 0.25% TritonX-100, 0.1% Tween-20) plus 1:5000 antifoam B and Roche cOmplete EDTA-free protease inhibitor cocktail (Oeffinger et al. 2007), sonicated and centrifuged as above, and then used immediately.

For the large-scale 2'OMe-MS experiments the buffer we used was a modified lower salt HSWB buffer (20mM Hepes-KOH, pH 7.4, 200mM KCl, 1.5mM MgCl<sub>2</sub>, 0.1% TritonX-100, 0.1% Tween-20) plus 1:5000 antifoam B and protease inhibitors. These extracts had a total protein concentration of approximately 10 mg/mL.

**Production of CmLSm antibodies and affinity purification.** Polyclonal anti-serum against recombinantly purified LSm complex was prepared by immunizing a 5-7 week old female New Zealand White rabbit with subcutaneous injections of 100 µg LSm protein in 750 µL phosphate buffered saline (PBS) combined 1:1 with Freund's complete adjuvant for the initial immunization, or with Freund's incomplete adjuvant for the booster injections. We performed the booster injections 22 and 43 days after the initial immunization. We anaesthetized the rabbit 57 days after the initial immunization, and performed the terminal bleed by cardiac puncture. We collected blood without heparin and incubated at 37 C for 1 h, then at 4 C overnight to facilitate clotting before carefully collecting serum by centrifugation (400 x g, 15 min).

We affinity purified the anti-CmLSm antibodies from crude serum. Purified CmLSm protein was coupled to CNBr-activated Sepharose 4B (GE Health Sciences) following the manufacturers instructions. After coupling, the LSm proteins were cross-linked to each other using dimethylpimelimidate (DMP) (Harlow and Lane 1988). Crude anti-CmLSm serum was bound in batch to the CmLSm-sepharose at 4 C overnight, washed, transferred to a BioRad polyprep 10 mL column, and antibodies were eluted with 100 mM glycine, pH 2.5 (Harlow and Lane 1988). This was immediately neutralized with 1 M Tris pH 8 and then dialyzed into PBS.

**LSm co-immunoprecipitation.** To immunoprecipitate the LSm complex from *C. merolae* extract we cross-linked anti-LSm serum or non-immune serum to Protein A Sepharose 4 Fast Flow (GE Healthcare) using DMP. Serum was incubated with the sepharose for 2 h at room temperature, beads were washed 3 times in binding buffer (0.2 M sodium borate, pH

9), and IgGs cross-linked for 40 min with 46 mM DMP in binding buffer. The reaction was quenched for 1.5 h with 0.2 M ethanolamine in 50 mM ammonium bicarbonate, pH 8. The antibody-bound beads were then washed 3 times in acid wash buffer (0.58% v/v acetic acid, 150 mM NaCl) to remove uncoupled IgGs, and then 3 times in PBS to neutralize. For co-immunoprecipitation followed by mass spectrometry (IP-MS) we densely conjugated affinity purified anti-LSm antibodies or rabbit IgG to magnetic M270 Epoxy Dynabeads (160 µg/mg; Trahan et al. 2016).

For small-scale reactions (done in triplicate), 100 µL of *C. merolae* extract was incubated with 10 µL cross-linked beads (anti-LSm or pre-immune) and 300 µL IPP<sub>150</sub> (20 mM Hepes-KOH, pH 7.9, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.05% IGEPAL) for 2 h at 4 C on a head-over-tail mixer. Beads were washed 3 times with IPP<sub>150</sub>. Protein was eluted for analysis by Western blot with 20 µL Laemmli buffer, 70 C for 10 min. Samples were electrophoresed on a 12% high TEMED SDS polyacrylamide gel (50 µL 10% APS, 50 µL TEMED/15 mL gel). The gel was transferred to nitrocellulose in Towbin buffer using a semi-dry blotter (Owl Separation Systems) at 1.5 mA/cm<sup>2</sup> for 40 min. The membrane was blocked, incubated with anti-LSm antiserum diluted 1:1000 in PBST for 1 h, washed, and then incubated for 1 h with goat anti-rabbit-HRP antibody (Santa Cruz Biotechnology) diluted 1:5000, before being visualized using the SantaCruz ImmunoCruz Western Blotting Luminol Reagent (SC-2048) according to the manufacturer's protocol. RNA was eluted from the beads for analysis by Northern blot by digestion with proteinase K (50 µg proteinase K in 200 µL buffer (10 mM Tris-HCl, pH 7.6, 100 mM NaCl, 1 mM EDTA, 0.5% SDS), 65 C for 20 min) and phenol:chloroform extraction. RNA was EtOH-precipitated in the presence of 30 µg glycogen and 5 µg total yeast RNA, resuspended and

electrophoresed on a 6%, 7 M urea polyacrylamide gel. The RNA was transferred onto nylon membrane in 1x TBE using a semi-dry blotter (2.5 mA/cm<sup>2</sup> for 20 min). The membranes were probed for all 4 snRNAs using <sup>32</sup>P-labelled oligonucleotides complementary to each of the snRNAs (see Supplemental Table S4; Stark et al. 2015) in RapidHyb buffer (GE Healthcare), 1 h at 60 C. Blots were washed and visualized on a phosphorimager after overnight exposure.

For IP-MS we incubated 36 mL of extract with 45 µL of either rabbit IgG- or anti-LSm antibody-conjugated Dynabeads (6.75 mg beads covered with ~1 mg antibody) for 2 h at 4 C, with slow rotation. The beads were recovered with a Dynal magnet and washed 10 times with the extract buffer, followed by several washes in NH<sub>4</sub>OAc, and finally in Tris-HCl before on-bead trypsin digestion (Trahan et al. 2016). The reactions were quenched with formic acid, and then cleaned on a C18 ZipTip (Millipore) following the manufacturers instructions. IP-MS reactions were done in duplicate.

**2'O-methyl oligo pull downs.** To isolate the snRNAs and their associated proteins from *C. merolae* extract, we first incubated the extract with a biotinylated RNA oligo complimentary to a short region of each snRNA, or, as a negative control, to an oligo complimentary to a region of *Saccharomyces cerevisiae* U6 snRNA (see Supplemental Table S4 - ro54, ro55, ro57, ro60, ro49). The RNA oligo was added in ~200x molar excess over the snRNA in the extract. Magnetic beads coated with Neutraavidin (GE SeraMag Neutraavidin SpeedBeads) were added, and the biotinylated oligo-associated RNP complex allowed to bind. For small-scale reactions (done in triplicate) 25 µL of extract was incubated with 2 pmol of oligo for 30 min at 30 C, or for 1 h at 4 C. Beads were pre-

blocked for 30 min at 4 C (20 mM Hepes, pH 7.4, 100 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01% Triton X-100, 1 mg/mL BSA, 200 µg/mL glycogen, 200 µg/mL yeast tRNA). After washing the beads, 5 µL were added to the extract/oligo, along with HSWB to bring the volume up to 400 µL. Reactions were incubated for 20 min at 4 C on a head-over-tail mixer. Beads were washed 3x in HSWB, and then the RNA was isolated for Northern blot analysis, as described above. For LC-MS/MS analysis, extract volume was increased to 36 mL. Two nmol of oligo were added to each reaction and nutated for 1 h 45 min at 4 C. Two hundred µL of blocked Neutravidin beads were then added and the mixture was nutated for another 1 h 45 min at 4 C. Washes, on-bead trypsin digestion, and ZipTip clean up were the same as described for the IP-MS. Each large scale oligo pull down was performed once.

**Mass spectrometry.** *In-gel digestion and sample analysis.* We separated recombinant CmLSm proteins by SDS-PAGE and confirmed their identities by mass spectrometry. Excised gel bands were destained twice in 100 mM ammonium bicarbonate/acetonitrile (50:50). The samples were then reduced (10mM β-mercaptoethanol in 100 mM bicarbonate) and alkylated (55 mM iodoacetamide in 100 mM bicarbonate). After dehydration, enough trypsin (6 ng/µl) was added to just cover the gel pieces and the digestion was allowed to proceed overnight at room temperature. Tryptic peptides were first extracted from the gel using 97% water/2% acetonitrile/1% formic acid followed by a second extraction using 50% of the first extraction buffer and 50% acetonitrile. Fractions containing tryptic peptides dissolved in aqueous 5% v/v acetonitrile and 1% v/v formic acid were resolved and ionized by using nanoflow HPLC (Easy-nLC II, Thermo Scientific)

coupled to an LTQ XL-Orbitrap hybrid mass spectrometer (Thermo Scientific). Nanoflow chromatography and electrospray ionization were accomplished by using a PicoFrit fused silica capillary column (ProteoPepII, C18) with 100  $\mu\text{m}$  inner diameter (300  $\text{\AA}$ , 5  $\mu\text{m}$ , New Objective). Peptide mixtures were injected onto the column and resolved at 500 nL/min using 75 min linear gradients from 0 to 40% v/v aqueous acetonitrile in 0.2% v/v formic acid. The mass spectrometer was operated in data-dependent acquisition mode, recording high-accuracy and high-resolution survey Orbitrap spectra using external mass calibration, with a resolution of 30,000 and m/z range of 400–2000. The fourteen most intense multiply charged ions were sequentially fragmented by collision induced dissociation (CID), and spectra of their fragments were recorded in the linear ion trap; after two fragmentations, all precursors selected for dissociation were dynamically excluded for 60 sec. Data were processed using Proteome Discoverer 1.4 (Thermo Scientific) and a non-reviewed Uniprot ([uniprot.org](http://uniprot.org)) *C. merolae* database. Search parameters included a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.8 Da. Peptides were searched with carbamidomethyl cysteine as a static modification and oxidized methionine and deamidated glutamine and asparagine as dynamic modifications.

*In-solution digestion and sample analysis.* We also analyzed the purified CmLSm complex in solution directly by mass spectrometry. 10  $\mu\text{g}$  of protein was denatured in 6 M urea for 10 min at RT. The urea was diluted to 1.4 M with 50 mM ammonium bicarbonate. 0.5  $\mu\text{g}$  of trypsin was added and incubated overnight at 37 C. The reaction was quenched with formic acid to 2% and 1/10 of the reaction was loaded directly onto a C18 reverse phase column for LC-MS/MS (LTQ Orbitrap Velos). A 75 min gradient was run at a flow rate of 300 nL/min: 20 min 95% Solvent A (0.1% formic acid) to 25% Solvent B (0.1% formic

acid, 100% acetonitrile), 40 min to 45% B, 10 min to 80% B. MS spectra were acquired in full ion scan mode from  $m/z$  400-1800 at a resolution of 30,000. In data-dependent MS/MS mode the 10 most intense precursor ions were fragmented by CID, followed by a dynamic exclusion of 20 s and an exclusion mass width of 10 ppm (Gingras et al. 2007; Trahan et al. 2016). All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.5.1). Mascot was set up to search the Refseq\_Cyanidioschyzon\_Merolae\_txid45157 database (5044 proteins) and the Refseq\_E\_Coli\_txid562\_20160930 database (1,224,352 proteins) assuming the digestion enzyme trypsin and one missed cleavage. Mascot was searched with a fragment tolerance 0.60 Da and a parent ion tolerance of 10.0 ppm. O+18 of pyrrolysine was specified in Mascot as a fixed modification. Oxidation of methionine was specified as a variable modification. Scaffold (version 4.3.4 Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 80% probability by the Peptide Prophet algorithm (Keller et al. 2002) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 95% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides that could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

*On-bead digestion and sample analysis.* A small fraction of the on-bead trypsin digested samples from the anti-LSm and rabbit IgG colPs and from the 2'OMe oligo pull downs was injected onto a C18 column on a Tribrid Fusion mass spectrometer (Thermo

Scientific), and the same gradient was run as on the Velos. The same volume was injected for the controls as for the experimental samples. MS/MS for the 10 most intense precursor ions was collected. Mascot was set up to search the Refseq\_Cyanidioschyzon\_Merolae\_txid45157 database, and peptide identifications were accepted if they exceeded specific database search engine thresholds. Mascot identifications required ion scores must be greater than both the associated identity scores and 30, 30, 25 and 25 for singly, doubly, triply and quadruply charged peptides respectively. The total spectral counts for each protein identified from the duplicate IP-MS experiments were averaged and only proteins with at least two unique peptides and that were at least twofold more abundant in the anti-LSm colP than in the IgG colP were reported.

**Immunofluorescence microscopy.** We grew *Cyanidioschyzon merolae* cells in MA2 media (Minoda et al. 2004) at 42 C and aerated with 5% CO<sub>2</sub> to an OD<sub>750</sub> of 1.4 under constant illumination, then subjected the culture to a 12 h dark/12 h light cycle. We collected a 15 mL aliquot 10 h into the second light cycle to isolate cells in interphase. We centrifuged the cells at 2000 x g for 5 min at room temperature, then resuspended the pellet in 50 µL of the supernatant. We mixed the cells with 1 mL of pre-chilled fixing solution (2% w/v paraformaldehyde, 0.2% v/v DMSO, 0.1 mM NaOH in methanol) and incubated at -20 C for 10 min. We centrifuged the cells at 800 x g for 1 min at 4 C (same for all subsequent centrifugation), removed the supernatant and washed once in 400 µL chilled methanol. We centrifuged 25 µL of the fixed cells, removed the supernatant, and resuspended in 30 µL PBS, then repeated and resuspended in 30 µL PBS containing 0.02% Triton-X100. After

incubation on ice for 15 min, we centrifuged the cells, washed once in PBS, then repeated and resuspended the cells in 30  $\mu$ L 5% BSA in PBS. We incubated on ice for 40 min, then washed the cells twice in 30  $\mu$ L PBS. We used rabbit anti-LSm antiserum (see previous methods for preparation) at a concentration of 1:1000 in PBS. The cells were suspended in 30  $\mu$ L primary anti-serum solution and incubated on ice overnight, then washed once in 30  $\mu$ L PBS. We used donkey anti-rabbit Alexa 488 secondary antibody (Thermo Fisher Scientific) at a final concentration of 1:1000. The cells were resuspended in 30  $\mu$ L secondary antibody solution, incubated on ice for 1 h, then washed 3 times in 30  $\mu$ L PBS. We resuspended the cells in 30  $\mu$ L of a solution of PBS/DAPI (Thermo Fischer Scientific) at a final concentration of 1  $\mu$ g/ml, incubated the cells on ice for 10 min, then washed once in PBS. Control reactions were performed exactly as above, with the exception of incubating in 1X PBS instead of anti-LSm antiserum. We mounted 1  $\mu$ L of sample with 1  $\mu$ L of ProLong Gold anti-fade reagent (Thermo Fisher Scientific), and visualized the cells with an Olympus BX61 fluorescence microscope using the following Semrock (IDEX Health and Science) filters: FITC-3540C, TxRED-4040C, DAPI-5060C. We captured images with a 100x objective lens using CellSens Dimensions software (Olympus). To better discern signal location, we used CellSens software to capture images in a z-stack using the following parameters: 7 z-slices total, 3  $\mu$ m range, 0.5  $\mu$ m step size. Digital subtraction of the blue (DAPI) from the green (anti-LSm) channels was performed in Adobe Photoshop using channel operations.

**Electron microscopy.** We prepared negative stain specimens by adsorbing purified proteins to glow discharged carbon-coated copper grids and staining with 0.75% (w/v) uranyl formate. We took images of the specimens on a Tecnai Spirit transmission electron microscope (FEI) equipped with a LaB6 filament and operated at an accelerating voltage of 120 kV. We obtained micrographs at a nominal magnification of 98,000x with an FEI Eagle 4K charge-coupled device (CCD) camera at a defocus value of -1.0  $\mu\text{m}$ . We averaged 2 x 2 pixels from the electron micrographs for a final pixel size of  $\sim 2.4 \text{ \AA}$  at the specimen level. We interactively selected particle images from these micrographs using Boxer. We subjected the selected particles to reference-free alignment and K-means classification using SPIDER.

**Immunoelectron Microscopy.** Immunoelectron microscopy was performed as described previously (Yagisawa et al. 2007). A cell pellet of *C. merolae* was rapidly frozen in liquid propane chilled in liquid nitrogen, transferred to dried acetone at -80 C (Miyagishima et al. 1999), and embedded in LR White resin (London Resin Co., London, UK). Serial, 88 nm thin sections were cut, put on the nickel mesh, and immunostained (Yagisawa et al. 2007). Samples on the mesh were blocked with 5% BSA in PBS for 30 min at room temperature. Primary antibody reactions were performed for 2 hours at 37 C with purified anti-CmLSm antibodies at a 1:10 dilution in 5% BSA. Grids were then incubated with goat anti-rabbit IgG secondary antibodies conjugated with 10-nm colloidal gold (Aurion) at a dilution of 1:50 for 2 hours at 37 C. Samples were washed with PBS and distilled water and stained with 3% uranyl acetate. Samples were then examined with an electron microscope (JEM=1200EX; JEOL, Tokyo, Japan).

## SUPPLEMENTAL REFERENCES

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Ansari A, Schwer B. 1995. SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J* **14**: 4001–4009.
- Brown NP, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**: 380–381.
- Dunn EA. 2014. Investigation of free U6 small nuclear ribonucleoprotein structure and function. ed. S.D. Rader. University of British Columbia, Vancouver.
- Dunn EA, Rader SD. 2014. Preparation of yeast whole cell splicing extract. *Methods Mol Biol* **1126**: 123–135.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Gingras A-C, Gstaiger M, Raught B, Aebersold R. 2007. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**: 645–654.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.
- Harlow E, Lane D. 1988. *Antibodies*. CSHL Press.

- Keller A, Nesvizhskii AI, Kolker E, Aebersold R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**: 5383–5392.
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res* **43**: W580–4.
- Milojevic T, Sonnleitner E, Romeo A. 2013. False positive RNA binding activities after Ni-affinity purification from *Escherichia coli*. *RNA Biol.*
- Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K. 2004. Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, *Cyanidioschyzon merolae* 10D. *Plant Cell Physiol* **45**: 667–671.
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**: 4646–4658.
- Oeffinger M, Wei KE, Rogers R, DeGrasse JA, Chait BT, Aitchison JD, Rout MP. 2007. Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat Methods* **4**: 951–956.
- Price SR, Ito N, Oubridge C, Avis JM, Nagai K. 1995. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* **249**: 398–408.

- Scheich C, Kümmel D, Soumailakakis D, Heinemann U, Büsow K. 2007. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res* **35**: e43.
- Söding J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**: 951–960.
- Stark MR, Dunn EA, Dunn WSC, Grisdale CJ, Daniele AR, Halstead MRG, Fast NM, Rader SD. 2015. Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proc Natl Acad Sci USA* **112**: E1191–200.
- Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.
- Trahan C, Aguilar L-C, Oeffinger M. 2016. Single-Step Affinity Purification (ssAP) and Mass Spectrometry of Macromolecular Complexes in the Yeast *S. cerevisiae*. *Methods Mol Biol* **1361**: 265–287.
- Vallejo AN, Pogulis RJ, Pease LR. 2008. PCR Mutagenesis by Overlap Extension and Gene SOE. *CSH Protoc* **2008**: pdb.prot4861.
- Ward N, Moreno-Hagelsieb G. 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS ONE* **9**: e101850.
- Wu D, Muhlrud D, Bowler MW, Jiang S, Liu Z, Parker R, Song H. 2014. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. *Cell Res* **24**: 233–246.

Wu S, Zhang Y. 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**: 3375–3382.



# RNA

A PUBLICATION OF THE RNA SOCIETY

## The sole LS<sub>m</sub> complex in *Cyanidioschyzon merolae* associates with pre-mRNA splicing and mRNA degradation factors

Kirsten A. Reimer, Martha R. Stark, Lisbeth-Carolina Aguilar, et al.

RNA 2017 23: 952-967 originally published online March 21, 2017  
Access the most recent version at doi:[10.1261/rna.058487.116](https://doi.org/10.1261/rna.058487.116)

---

### Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2017/03/21/rna.058487.116.DC1>

### References

This article cites 63 articles, 32 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/23/6/952.full.html#ref-list-1>

### Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



Profiling plasma microRNAs in populations. Read expert advice.

EXIQON

---

To subscribe to RNA go to:  
<http://rnajournal.cshlp.org/subscriptions>

---