BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 2335–2345 2335
DOI: 10.1021/bi901767w

Engineered Rings of Mixed Yeast Lsm Proteins Show Differential Interactions with Translation Factors and U-Rich RNA[†]

Meghna Sobti,[‡,⊥] Liza Cubeddu,[§] Paul A. Haynes,[‡,∥] and Bridget C. Mabbutt*[,‡]

[‡]*Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales 2109, Australia,* [§]*School of Molecular and Microbial Biosciences, University of Sydney, Sydney, New South Wales 2006, Australia, and* [∥]*Australian Proteome Analysis Facility, Macquarie University, Sydney, New South Wales 2109, Australia.* [⊥]*Present address: Molecular Genetics Department, Victor Chang Cardiac Research Institute, Sydney, New South Wales 2010, Australia.*

Received October 14, 2009; Revised Manuscript Received January 11, 2010

ABSTRACT: The Lsm proteins organize as heteroheptameric ring assemblies capable of binding RNA substrates and ancillary protein factors. We have constructed simplified Lsm polyproteins that organize as multimeric ring structures as analogues of the functional Lsm complexes. Polyproteins Lsm[2+3], Lsm[4+1], and Lsm[5+6] incorporate natural sequence extensions as linker peptides between the core Lsm domains. In solution, the recombinant products organize as stable ring oligomers (75 Å wide, 20 Å pores) in discrete tetrameric and octameric forms. Following immobilization, the polyproteins successfully act as affinity pull-down ligands for proteins within yeast lysate, including native Lsm proteins. Interaction partners were consistent with current models of the mixed Lsm ring assembly in vivo but also suggest that dynamic rearrangements of Lsm protein complexes can occur. The Lsm polyprotein ring complexes were seen in gel shift assays to have a preference for U-rich RNA sequences, with tightest binding measured for Lsm[2+3] with U_{10}. Polyprotein rings containing truncated forms of Lsm1 and Lsm4 were found to associate with translation, initiation, and elongation protein factors in an RNA-dependent manner. Our findings suggest Lsm1 and/or Lsm4 can interact with translationally active mRNA.

Lsm (Sm-like) proteins are engaged in cellular RNA metabolism across the three domains of life (*1*, *2*). Once organized into oligomeric ring assemblies, these proteins play diverse roles in RNA processing steps such as splicing, post-transcriptional modification, and degradation. In eukaryotes, the specific composition and architecture of the Lsm complex determine its cellular location, RNA target, and function (*1*, *3*). A complex of seven Lsm proteins in the cell nucleus, Lsm[2−8], directly binds and stabilizes the 3′ poly(U) tract of U6 snRNA, facilitating U4/U6 annealing/stabilization and di- and tri-snRNP formation (*4*). The Lsm[2−8] protein complex also processes other types of RNA in the nucleus, including tRNA, snoRNA, and rRNA forms, and assists in the decay of mRNA (reviewed in ref *1*). This complex is homologous to the heptamer of Sm proteins known to assemble on U-rich sequences of snRNA (U1, U2, U4, and U5) as part of the biogenesis of the respective ribonucleoproteins (RNPs) for pre-mRNA splicing (*5*−*7*).

A closely related Lsm complex localized to the cytoplasm, Lsm[1−7], associates with deadenylated mRNA and promotes decapping as part of the 5′-to-3′ decay machinery (*8*, *9*). The Lsm[1−7] heptamer has been shown to directly bind uracil-rich sequences near the 3′ terminus of mRNA (*10*). There, in association with Pat1 and other protein factors, it recruits the decapping complex engaged with the 5′ end of the transcript (containing Dcp1/2 and Dhh1) (*11*, *12*) and promotes subsequent degradation by exonuclease XrnI (*13*). Under certain conditions, many of these decay factors, including Lsm[1−7], are found concentrated within granular foci known as processing bodies (P-bodies) (*14*). P-Body formation is thought to control the balance between RNA decay and translation processes (*8*).

Proteins of the L/Sm family all contain an OB-related fold consisting of a highly bent antiparallel β-sheet, often capped by an N-terminal α-helix (*15*−*20*). A loop, named loop L4, of variable length and sequence connects strands β3 and β4 of the five-stranded sheet (Figure 1A). Closed-ring assemblies of contiguous Lsm monomers (homomeric in prokaryotes, heteromeric in eukaryotes) are naturally formed through alignment of the β4 strand of one subunit against strand β5 of its neighbor. The resulting stable oligomers present a toroid face, the "proximal" or "helix" face, in which key RNA-binding side chains (from loops L3 and L5) are clustered near the central cavity (*15*, *16*, *19*, *21*). On the opposite side of the Lsm ring is the "loop" or "distal" face, from which the different L4 loop regions protrude (*15*).

Some details concerning the mode of RNA binding to Lsm protein rings are provided by crystal structures of homomeric archaeal and bacterial Lsm assemblies bound to short oligonucleotides (*16*, *19*, *21*, *22*). For the Hfq homologue from *Staphylococcus aureus*, each of six bases (A or U) stacks within clefts between monomers on the proximal face of a hexameric protein ring (*22*). This allows individual base stacking to engage aromatic residues from loop L3 regions of two neighboring monomers. Charged side chains nearby from loop L5 appear to tune the specificity of each RNA-binding pocket (*22*, *23*). Crystal

[†]This work was supported by Macquarie University research funds and a scholarship (M.S.). L.C. acknowledges a National Breast Cancer Foundation Fellowship.

*To whom correspondence should be addressed: Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. Telephone: 61-2-98508282. Fax: 61-2-98508313. E-mail: bridget.mabbutt@mq.edu.au.

Published on Web 01/28/2010

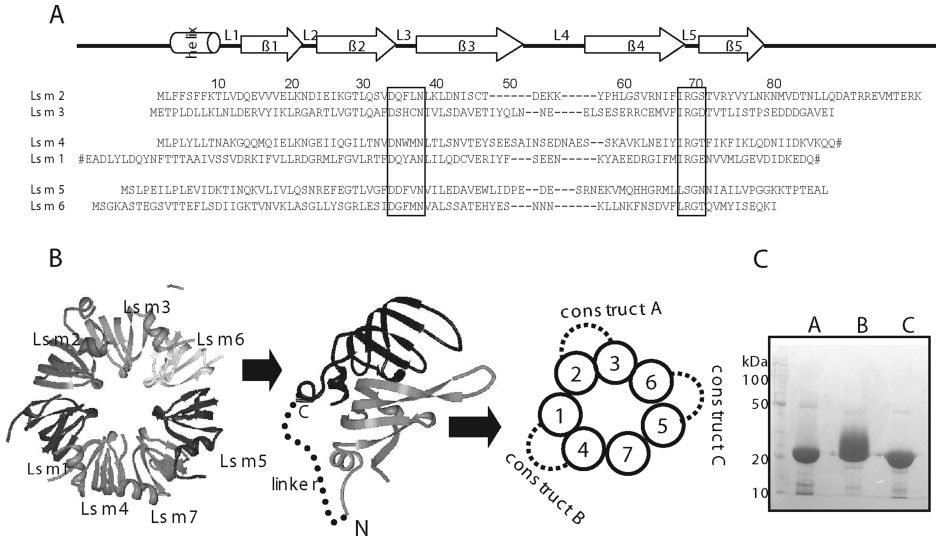

FIGURE 1: Design of polyprotein versions of Lsm ring assemblies. (A) Structure-based alignment of yeast Lsm sequences from the crystal structure of Lsm3 (*20*). Secondary structure elements and sequence numbering are from Lsm3; boxed residues have been implicated in RNA binding. Points at which native sequences were truncated for polyprotein constructs are denoted (#). (B) Model of the Lsm[1−7] complex (*20*), shown from the distal face. Construct A indicates polyprotein Lsm[2+3], construct B Lsm[4+1], and construct C Lsm[5+6]. (C) SDS−PAGE gels showing purified polyproteins obtained from constructs A (Lsm[2+3]), B (Lsm[4+1]), and C (Lsm[5+6]) stained with Coomassie dye.

structures of archaeal Lsm heptamers in complexes with U-rich oligonucleotides [Af-Sm1 with U_5 (*16*) and Pa-Sm1 with U_7 (*19*)] also show that the RNA bases pack within clefts on the proximal face lined by loop L3 and L5 residues. There is some evidence that the Hfq hexamer of *Escherichia coli* possesses a second RNA-binding site optimized for poly(A) binding on its distal ring face (*24*).

The Lsm complexes are prominent in RNA biology because of their ability to bind proteins as well as RNA within RNPs. Many Lsm proteins contain highly charged N- and/or C-terminal extensions beyond the L/Sm folded domain, which vary greatly in length. It is likely that these unstructured segments serve as points of recruitment for ancillary proteins to the Lsm ring assemblies (*20*), possibly in concert with the protruding L4 loops. The C-terminal regions of Lsm1 and Lsm4, for example, have been specifically implicated in both mRNA decay in yeast (*25*) and P-body assembly (*26*, *27*). Extensive protein interaction networks have now been mapped for the yeast Lsm protein complexes from two-hybrid screens (*28*) and mass spectrometry (MS) analysis of complexes isolated by immunoprecipitation (*29*, *30*), affinity purification, or tandem affinity purification (*31*, *32*). The heteroheptameric Lsm complexes are seen to readily recruit ancillary proteins: Lsm[2−8] interacts directly or indirectly with several splicing factors (Prp and Snu families) and Lsm[1−7] with general repression and decay proteins (such as Dcp1/Dcp2 and Pat1). Other interactions with proteins involved in pseudouridinylation, tRNA aminoacylation, translation, and ribosomal function have also been documented for both complexes (*33*).

Many cellular functions are not conducted by single protein components, rather by macromolecular complexes of multiple subunits with specific functions. These protein complexes are relatively sparse in the cell and purification yields insufficient for crystallographic study. Alternative approaches such as reconstitution of large complexes from recombinant protein subunits, as well as determination of crystal structures of individual subunits and functional domains, are required. The recent successful determination of the structure of U1 snRNP in association with U1 snRNA is an example of this, for which the complex was reconstituted from recombinant proteins overexpressed in *E. coli* (*7*). To pursue structural studies of related RNP complexes containing Lsm proteins, we have prepared recombinant versions of the core component proteins. It has been amply demonstrated that recombinant Lsm proteins organize immediately on expression as highly stable homomeric rings, usually hexamers (*6*, *22*) or heptamers (*15*−*19*). Thus, separate Lsm preparations cannot be readily mixed to form heteromeric protein assemblies for in vitro functional or structural study, unless denatured and refolded. Some success has been reported for human Lsm complexes, whereby specific pair (Lsm[2/3] and Lsm[4/8]) and triplet (Lsm[5/6/7]) combinations were prepared by coexpression via polycistronic vectors (*34*). The resulting Lsm[2/3] complex formed a relatively stable mixture of heptameric and octameric forms, with a ring-shaped morphology.

As an alternative avenue for the assembly of mixed Lsm complexes, we have engineered the yeast sequences into polyprotein combinations incorporating tail sequences as linkers (*35*, *36*). Having previously established that homomeric ring octamers of yeast Lsm3 possess binding epitopes for some natural Lsm partners (*20*), we have used the octameric ring scaffold as a design template for Lsm polyprotein dimers likely to self-associate as stable subcomplexes. The specific Lsm combinations we have pursued, named Lsm[2+3] and Lsm[4+1], are closely related to the SmD_1−SmD_2 and SmD_3−SmB/B′ dimer complexes, respectively. These specific Sm heteromeric dimers are known to engage in direct protein−protein interactions in vivo and to comprise discrete natural components of the assembly pathway of the Sm heptameric ring (*5*, *6*). Our related Lsm polyproteins now in turn provide a simplified set of ring complexes for probing the mechanism of Lsm assembly, specific RNA binding properties, and the protein interactome of the yeast Lsm family. Toward this end, we have utilized affinity experiments in which the prepared polyproteins act as bait for interacting protein partners within yeast lysate. In addition, the simplified complexes allow some detection of the specificity of individual Lsm subunits for RNA sequences.

Table 1: Primers Used in the Construction and Insertion of Yeast Lsm Constructs

encoded product	oligonucleotide primer sequence
Lsm2	5′ GCGCGGCAGCCATATGCTTTTCTTCTCCTTT 3′ (forward)
	5′ TGTCTCCATTTTTCTTTCAGTCATTACCTC 3′ (reverse)
Lsm3	5′ ACTGAAAGAAAAATGGAGACACCTTTGGATTTA 3′ (forward)
	5′ GTTAGCAGCCGGATCCTTATATCTCCACTGCGCC 3′ (reverse)
Lsm4	5′ GCGCGGCAGCCATATGATGCTACCTTTATATCTT 3′ (forward)
	5′ ATCGGCCTCTTGCTGCTTGACCTTGTC 3′ (reverse)
Lsm1	5′ AAGCAGCAAGACGCCGATTTATATCTC 3′ (forward)
	5′ GTTAGCAGCCGGATCCTTATTGATCTTCTTTATCGATGTC 3′ (reverse)
Lsm5	5′ GCGCGGCAGCCATATGATGAGTCTACCGGAGATT 3′ (forward)
	5′ CAAAAAATAATGAGTCTACCGGAGATT 3′ (reverse)
Lsm6	5′ ACGGAGGCGTTGATGTCCGGAAAAGCTTCTACA 3′ (forward)
	5′ GTTAGCAGCCGGATCCTTATATTTTTTGTTCACTGATATACAT 3′ (reverse)

EXPERIMENTAL PROCEDURES

Cloning of Saccharomyces cerevisiae LSM Genes. Genes encoding *LSM1*, *LSM3*, *LSM4*, *LSM5*, and *LSM6* were amplified from *S. cerevisiae* (strain BY4741) genomic DNA by PCR using appropriate primers. Intronless *LSM2* was a gift from G. Kornfeld (University of New South Wales, Sydney, Australia). Full-length genes were used except in two cases: *LSM4* was truncated at the 3′ end to remove 282 bp and *LSM1* at both 5′ and 3′ ends to remove 87 and 150 bp, respectively. Amino acid sequences of the gene products produced for this study are shown in Figure 1A.

To create selected Lsm polyprotein expression cassettes, genes were linked head to tail by partially overlapping primer-based PCR (*37*) for the following gene pairs: *LSM4* fused with *LSM1*, *LSM2* fused with *LSM3*, and *LSM5* fused with *LSM6*. Primers (see Table 1) were designed to create single genes of each pair containing 15−18 bp regions of neighbor overlap with a theoretical melting temperature of ≥50 °C. The initial annealing step utilized 50−100 ng of each gene of a pair with 200 μM dNTPs, 1 mM $MgCl_2$, and 1.25 units of KOD polymerase (Novagen) in a 50 μL reaction volume. No primers were added at this stage. Thermocycler conditions were as follows: one cycle of 94 °C for 3 min; 15 cycles of 94 °C for 30 s, 55 °C for 30−60 s, and 72 °C for 30−60 s; and one cycle of 72 °C for 5 min. Following purification with a spin column (Qiagen), 2.5−5 μL of material from the reaction mixture described above was combined with 5′ and 3′ end primers (each primer at 0.5 μM) containing 16−19 bp overlap regions with the target vector in the presence of 200 μM dNTPs, 1 mM $MgCl_2$, and 1.25 units of KOD polymerase in a 50 μL reaction volume. The following thermocycler conditions were used: one cycle of 94 °C for 3 min; five cycles of 94 °C for 30 s, 45 °C for 30 s, and 72 °C for 30−60 s; 25 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30−60 s; and one cycle of 72 °C for 5 min.

As a part of the construction of the Lsm polyproteins, appropriate peptide linkers were incorporated between each sequence pair to allow rings of subunits to form in an unstrained manner. Inspection of the crystal structures of octameric Lsm3 (*20*) and the Hfq hexamer (*38*) shows the N- and C-terminal segments of each structured subunit to lie on or near the toroid proximal face, but to be separated by up to ∼23 Å. We utilized long peptide linkers (16−27 residues) taken from the natural N- and C-terminally extended sequences of each Lsm protein pair to span folded subunits (as indicated, Figure 1B). For each construct, several variants were prepared with altered charges and lengths of linker peptide and/or termini.

The expression cassettes for *LSM*[*2+3*], *LSM*[*4+1*], and *LSM*[*5+6*] (constructs A−C in Figure 1B, respectively) were each cloned into pET15b (Novagen) using ligation-independent cloning (In-Fusion system, Clontech). Primers for cloning into pET15b were designed to supplement the cassettes with GCGCGGCAGCCATATG (5′ of the first gene) and GTTAGCAGCCGGATCCTTA (3′ of the last gene). The plasmids generated were transformed into *E. coli* strain BL21(DE3)/Rosetta (Novagen) for expression as hexa-His-tagged protein products. The vector introduces a thrombin cleavage sequence between the N-terminal tag sequence and the inserted protein product, i.e., MH_6SSGLVPRGSHM-. The single product *LSM3* was cloned and expressed as an N-terminally His-tagged protein as described previously (*20*).

Protein Expression and Purification. Transformed bacterial cells were grown at 37 °C in 2−4 L of LB broth containing ampicillin (100 μg/mL) (Amresco) and chloramphenicol (25 μg/mL) (Amresco). Expression of the recombinant protein was induced by addition of isopropyl β-D-thiogalactopyranoside (1 mM) to the culture at an optical density (600 nm reading) (OD_{600}) of 0.4−0.6. The culture was shaken at 25 °C for a further 6 h. The harvested cell pellet was lysed (French press) into buffer A [20 mM Tris-HCl (pH 8.0) and 200 mM NaCl] containing 10 mM imidazole and protease inhibitor cocktail (Sigma). Following filtration (0.22 μm), the lysate was applied to a 1 mL HisTrap column (GE Healthcare) pre-equilibrated with 10 mM imidazole in buffer A. Unbound proteins were removed with 40 mM imidazole in buffer A (20 column volumes), and affinity-tagged Lsm proteins were eluted into 500 mM imidazole in buffer A. The homogeneity of all isolated Lsm protein complexes was further enhanced by size exclusion chromatography (SEC) on a Superdex 200 10/300 GL column (GE Healthcare) running in buffer A. Under some salt conditions, a small proportion of a 50 kDa bacterial protein copurified with our Lsm complexes. This contaminant (<1%) was visible in some gels and identified as oligomeric forms of *E. coli* glycerol kinase.

Characterization of Lsm Protein Complexes. Purified proteins were analyzed by denaturing polyacrylamide gel electrophoresis (SDS−PAGE) using a 12% gel and a Tris/Tricine buffer system (*39*) and visualized with Coomassie R250. The size and heterogeneity of the Lsm oligomeric complexes in solution were determined by SEC and, for Lsm[2+3], multiangle light scattering (MALLS). Elution and void volumes of the SEC column were calibrated using protein standards (13−440 kDa) and blue dextran (GE Healthcare). MALLS analysis used a

three-angle miniDAWN detector (Wyatt Laboratories) in tandem with a refractometer (Wyatt Laboratories). Detectors were downstream of a liquid chromatography system (GE Healthcare) in which proteins were first separated using a Superdex 200 column (10/300 GL, GE Healthcare) operating at 0.5 mL/min in buffer A. Bovine serum albumin (Promega) was used to normalize MALLS output.

Matrix-assisted laser desorption ionization (MALDI) MS was conducted on a Bruker microflex instrument. Protein samples (0.5 μL) were applied to the sample plate at a rate of 20 pmol/μL in 0.1% formic acid, mixed with 0.5 μL of matrix solution, and air-dried. The matrix solution was 20 μg/μL sinapinic acid (Bruker Care) in 70% acetonitrile and 0.1% trifluoroacetic acid. The mass spectrometer was calibrated using Protein Calibration Standard II (Bruker Care) placed in the proximity of the sample. Data were processed and displayed using FlexAnalysis version 3.0 (Bruker Daltonics).

Circular dichroism (CD) spectra of protein complexes were recorded from 190 to 300 nm in buffer A. Data were collected at 22 °C on a Jasco J-180 spectropolarimeter using a 0.1 mm cuvette with a 2 s response time. For each experiment, 16 spectra were averaged to enhance the signal-to-noise ratio. Thermal denaturation of the protein complexes was monitored at 215 nm (or 220 nm for Lsm[5+6]), the wavelength at which folded and denatured samples exhibited the greatest intensity difference. Samples were heated over the range of 20−99 °C at a rate of 3 °C/min with a Peltier accessory. All spectra were analyzed using Jasco Spectra Manager version 1.52.

Transmission Electron Microscopy. Copper grids (200 μm mesh) were coated with a protein solution (10 μg/mL) for 5 s and stained with 1% uranyl acetate for 5 min. Micrographs were recorded on a Philips CM10 transmission electron microscope at an accelerated voltage of 100 kV and a magnification of 39000−105000×. Negatives were digitized using a rotating-drum scanner at 1200 dots per inch (dpi). Individual particle selection and particle size measurements were performed using the EMAN software suite (*40*).

Pull-Down Affinity Studies. Purified recombinant Lsm polyproteins (~15 mg) were immobilized via their affinity fusion tags to a HisTrap column (1 mL) equilibrated in buffer A. Yeast cell lysate from 1 L of culture (see below) was passed over the Lsm column at a rate of 0.5 mL/min. Nonassociating components were removed when the solution was washed with 75 mM imidazole in buffer A until a stable baseline was observed (25−30 column volumes). Mixtures of adhering proteins were subsequently batch eluted with 500 mM imidazole in buffer A. Aliquots (500 μL) of the eluant were precipitated with acetone (2 mL) and samples loaded (10 μL) on 12% SDS−PAGE gels. Following electrophoresis and staining with Coomassie dye, protein bands were excised and individual gel slices subjected to destaining, reduction (dithiothreitol), alkylation (iodoacetamide), and trypsin digestion using the reported methods (*41*). Tryptic peptides were extracted for MS analysis into 2% formic acid in 50% acetonitrile and concentrated under vacuum. All pull-down experiments were reproduced in triplicate and included a control run in the absence of bait protein. In addition to specific protein bands, unstained gel regions were also excised and digested with trypsin to ensure identification of proteins present at low concentrations.

Peptide extracts were analyzed by nano LC−MS/MS using a Surveyor HPLC pump (Thermo) modified for nanoliter flow rates and connected to a LCQ-Deca ion trap mass spectrometer (Thermo), as previously described (*42*). Reversed phase columns were packed in-house to ~7 cm × 100 μm using 5 mM Zorbax C-18 resin (100 Å, Agilent Technologies) in a fused-silica capillary with an integrated electrospray tip. A 1.8 kV electrospray voltage was applied via liquid junction upstream of the column. Samples were injected using a Surveyor autosampler (Thermo), and the column was washed (10 min) with 5% acetonitrile and 0.1% formic acid (1 μL/min). Peptides were subsequently eluted at 500 nL/min with a 0 to 50% gradient of 95% acetonitrile and 0.1% formic acid over 58 min, followed by a 50 to 95% gradient over 5 min. The column eluate was directed into the spectrometer nanospray ionization source, and spectra were scanned over the range of 400−1500 amu.

Automated peak recognition, dynamic exclusion, and tandem MS of the three most intense precursor ions at 40% normalization collision energy were performed using Xcalibur (Thermo), as previously outlined (*43*). Raw files were converted to mzXML format and analyzed using the XTandem algorithm within the global proteome machine (GPM) software interface (*44*). Program search settings included a maximum valid protein expectation value of 0.1, static mass modification of +57.022 for carbamidomethylated cysteine, potential residue mass modifications for oxidized methionine and threonine, and deamidation of asparagine and glutamine, trypsin enzyme specificity; spectrum parameters, including a fragment monoisotopic mass error of 0.5 Da and a parent monoisotopic mass error of 2.0 Da; and spectrum conditioning parameters of a 100.0 spectrum dynamic range, a total of 50 spectrum peaks, a minimum parent M + H of 400.0, and a minimum fragment m/z of 150.0. All MS/MS spectra were searched against the *E. coli* and yeast (*S. cerevisiae*) sequence databases (SwissProt).

The criteria used in this study for positive protein identification were as follows: (i) a minimum of two peptides identified and (ii) a protein expectation score value of log(e) < −10. Within these criteria, no protein matches were found by reverse database searching (*45*). Protein identifications were only considered valid when observed in at least two of the three replicates. Highly abundant yeast housekeeping proteins found in all gels, e.g., glycolysis enzymes (pyruvate kinase and glyceraldehyde-3-phosphate dehydrogenase), ribosomal proteins, and heat shock proteins (SSA2, Hsp40, Hsp70, and YDJ1), were treated as contaminants. Bacterial Hfq was not recovered as an interacting partner in any of the pull-down experiments. All results are tabulated in the Supporting Information.

Preparation of Yeast Cell Lysate. Cultures of *S. cerevisiae* (strain W203) were grown at 30 °C in yeast peptone dextrose medium for 6−8 h until the OD_{600} value reached 0.5−0.7 (i.e., at log phase). Cells were harvested, and the pellet was suspended in 5 mL (per gram of cell mass) of 50 mM Tris buffer (pH 7.5), 200 mM NaCl, 1 mM EDTA, 5 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride, 1 μg/mL leupeptin (Sigma-Aldrich), and 1 μg/mL pepstatin A (Sigma-Aldrich). Cells were lysed (French press), and the lysate was clarified and filtered (0.22 μm). For RNA-free experiments, the lysate preparation was further treated with 100 μg/mL RNase enzyme (New England Biolabs).

Electrophoretic Mobility Shift Assays (*EMSAs*). RNA oligonucleotides were labeled with ^{32}P at the 5′ ends according to standard procedures (*46*) and purified on a Sephadex G-25 spin column (Roche). The labeled RNA (20 fmol) was incubated with protein over a range of concentrations (0.1−20 μM in buffer A) at 22 °C for 30 min, as previously described (*18*). Samples were

Table 2: Structural Parameters of Engineered Yeast Lsm Polyprotein Complexes

protein	linker peptide length	M_R (kDa)	MALDI-MS (kDa)	native mass[a] (kDa)	subunit no.	T_m[b] (°C)
Lsm3	−	10.984	10.98	80	8	70
Lsm[2+3]	27	23.47	23.294	85 (II)	4	65
				185 (I)	8	−
Lsm[4+1]	17	23.05	22.791	100 (II)	4	55
				200 (I)	8	−
Lsm[5+6]	20	21.9	21.925	42	2	nd
				90	4	
				190 (I)	8	
				others (> 250)	16+	

[a]Values determined in Tris buffer (pH 8.0) and 200 mM NaCl, either by SEC (Lsm3, Lsm[4+1], and Lsm[5+6]) or by MALLS (Lsm[2+3]). Chromatography fractions I and II are indicated. [b]Melting temperatures were determined from the first derivative of the CD molar ellipticity change (215 nm) monitored in Tris buffer (pH 8.0) and 200 mM NaCl. The value for Lsm3 was taken from ref *49*.

fractionated on a 10% native PAGE gel at 20 mA for 40 min in Tris/borate buffer containing ethylenediaminetetraacetate (EDTA). Gels were exposed to a phosphorimage plate (GE Healthcare) overnight and visualized using Typhoon Phosphorimager software (GE Healthcare). Bands corresponding to bound and unbound RNA fractions were quantified using Image Gauge (Fuji). All EMSA experiments were performed in duplicate, and quantitation of binding reactions was achieved from an average of the two experiments. We determined apparent dissociation constant (K_D) values by plotting the fraction of bound RNA versus the protein concentration and fitting of mean data points to the following equation (*47*, *48*):

$$\text{fraction bound} = 1/(1 + K_D/[\text{protein}])$$

RESULTS

Oligomeric Complexes of Lsm Polyproteins. Prevailing models of heteromeric Lsm protein complexes present a heptamer ring with protomers arranged in a fashion analogous to that of the homologous Sm heptamer assembly (*1*). Thus, as depicted in Figure 1B, Lsm4 makes contact with adjacent Lsm1 and Lsm7 protomers, Lsm2 adjoins Lsm3 and Lsm1, and Lsm5 adjoins both Lsm6 and Lsm7. Some of these native intrasubunit interactions are likely to be replicated within simplified Lsm subcomplexes, provided they were also able to adopt a ring morphology. One means of achieving such quaternary structures is to produce single polyprotein chains comprising covalently linked dimers of Lsm sequences. With sufficiently flexible linker peptides incorporated, these polyproteins have the capacity to mimic the octameric and hexameric rings common to the Lsm family through assembly of four or three dimer units (*1*, *20*, *38*). Three specific polyproteins of interest are Lsm[2+3], Lsm[4+1], and Lsm[5+6], here named constructs A, B, and C, respectively (Figure 1B). These constructs directly relate to the stable Sm[D_1/D_2], Sm[D_3/B], and Sm[E/F] subcomplexes, respectively, previously produced by coexpression methods (*6*). Table 2 summarizes the physical properties of the recombinant products of Lsm polyprotein constructs utilizing sequences from yeast.

Polyprotein Lsm[2+3] (construct A) comprises full-length sequences of its component proteins, with an additional

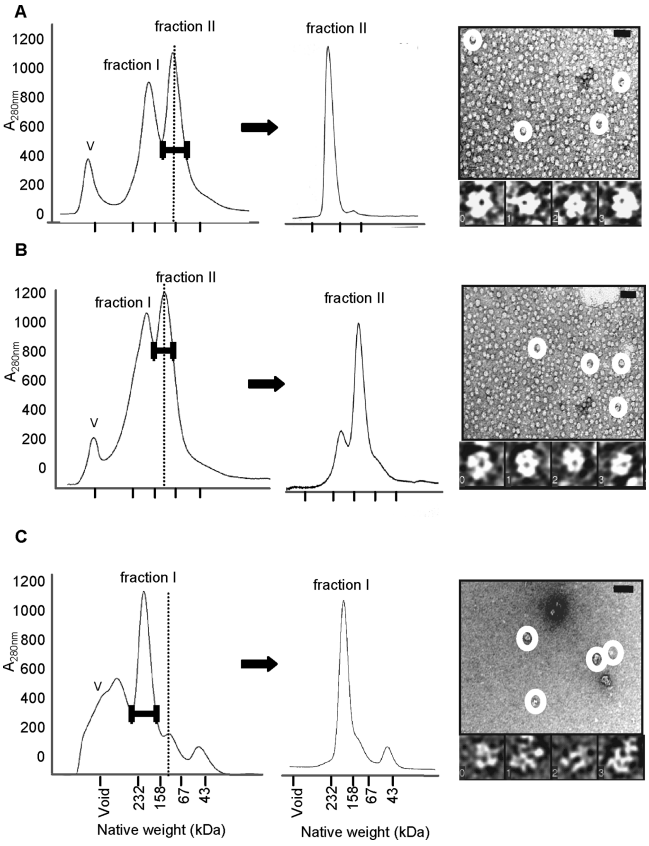

FIGURE 2: Biophysical characterization of yeast Lsm polyproteins. Data are shown for preparations (A) Lsm[2+3], (B) Lsm[4+1], and (C) Lsm[5+6]. Left panels show UV traces (280 nm) for SEC profiles on Superdex 200 in 20 mM Tris buffer (pH 8.0) with 200 mM NaCl (0.5 mL/min). The column was in-line with light scattering and refractometer detection for the determination of the native masses of protein species, for which elution volumes have been calibrated. The apparent native mass scale for all traces is indicated below panel C (with void volume, V), and dotted vertical lines indicate elution points for tetrameric forms of each distinct construct. Fraction I material corresponds to octamers of the polyproteins and fraction II to tetrameric forms. Protein samples collected for the second purification round (used in all further analysis) are indicated. Right panels show electron micrographs of the three polyproteins, tetramers of Lsm[2+3] and Lsm[4+1] (A and B) and octamers of Lsm[5+6] (C). The scale bar is 50 nm. Circled particles have been magnified at the bottom of each micrograph to view the morphology.

N-terminal hexa-His sequence. As detailed in Figure 1A, this construct creates a 25-residue flexible linker between two folded domains, incorporating the native tail sequence of Lsm2 (23 C-terminal residues) fused to the first two residues of Lsm3 so as to accommodate a nativelike ring architecture in an unstrained manner. Following purification, the construct A protein product yielded a single band on SDS−PAGE (~23 kDa in Figure 1C), and its molecular mass was confirmed by MALDI-MS. In a moderate salt solution at pH 8, Lsm[2+3] is found to exist in two discrete oligomeric forms. When separated via SEC in 200 mM NaCl (Figure 2A), these were determined by MALLS to be complexes of 185 and 85 kDa (fractions I and II, respectively). The two species are stable, do not interconvert when rechromatographed, and are consistent with octameric and tetrameric forms of the polyprotein. Given the oligomeric organizations previously seen for recombinant Lsm3 in both crystal and solution (*20*), the tetramer would correspond to a single ring containing eight folded Lsm domains and the Lsm[2+3] octamer to a pair of such rings.

Construct B consisted of a shortened sequence of Lsm4 (residues 1−84), affinity-tagged and linked to a truncated sequence of Lsm1 (residues 30−122). The Lsm[4+1] product thus contains a relatively short linker of 17 residues (five from the Lsm4 C-terminus, 12 from the Lsm1 N-terminus). Following purification, a 23 kDa product was verified by MS and yielded a single clear band on SDS−PAGE (Figure 1C). At pH 8, the Lsm[4+1] preparation contained at least two predominant species, determined to have masses of 200 kDa (fraction I) and 100 kDa (fraction II) (Figure 2B). Additional higher-molecular mass forms are also evident under the moderate salt conditions of the SEC runs (200 mM NaCl). Protein from fractions I and II, which likely comprise octameric and tetrameric Lsm[4+1] species, could be separated. In this case, the two oligomeric forms were somewhat labile and found to interconvert over time.

For construct C, full sequences of both Lsm components were included to encode a 22.1 kDa polyprotein. Lsm[5+6] thus incorporates a natural 20-residue linker, 10 C-terminal Lsm5 residues fused to 10 residues of Lsm6. On purification, a single 22 kDa protein product was verified (see SDS−PAGE in Figure 1C). The solution form of this polyprotein was relatively unstable, however, with no fewer than five fractions resolvable at pH 8 by SEC (Figure 2C), as well as large species present in the void. The predominant oligomeric form, estimated to be 190 kDa (fraction I), is consistent with an octamer of the Lsm[5+6] polyprotein. Later-eluting fractions indicate the additional presence of dimer and tetramer species. When separated, however, all oligomers of Lsm[5+6] readily interconvert over time, indicating that this polyprotein version does not here adopt any specific quaternary structure.

Evidence of the Ring Morphology of Lsm Polyproteins. Negatively stained Lsm samples were subjected to EM for visualization of the oligomeric structures adopted by the isolated polyproteins. After particle selection and class averaging, distinct ring morphology is clearly evident for isolated tetramers of Lsm[2+3] and Lsm[4+1], as depicted in Figure 2. These images confirm that the tetrameric complexes likely stack together according to nativelike subunit interactions. The widths measured from the EM micrographs for both Lsm[2+3] and Lsm[4+1] tetramers were 75 ± 2 Å, with a central pore of 20 ± 2 Å, parameters identical to those we determined previously for the octameric Lsm3 toroid (*20*). For the octamer of Lsm[2+3] (fraction I material), a micrograph scanned at a 15° tilt showed rings of these same dimensions, but coaxially stacked in pairs. These images of a stacked organization are consistent with dual packing of rings as seen for crystalline Lsm3 (*20*). Rings of comparable dimensions have also been reported previously from EM for human Lsm heteromeric complexes prepared by coexpression (i.e., for Lsm[2/3] and Lsm[5/6/7] complexes) (*34*). A reconstruction of a single particle of the Lsm[2+3] complex from 2298 individual images (not shown) yielded further depiction of the flattened ring form, but individual protein components could not be discerned.

In the case of octamers of polyprotein Lsm[5+6], intact ring particles are also evident in our EM micrographs, but mixed with smaller forms, likely due to dimer species (Figure 2C).

CD studies were used to monitor the stability of the separated tetrameric fractions of the polyprotein preparations. From the change in molar ellipticity at 215 nm, the characteristic minimum feature for folded Lsm samples, apparent melting temperatures were 65 °C for Lsm[2+3] and 55 °C for Lsm[4+1]. These polyprotein versions of ring complexes are thus only somewhat less robust than the single-component Lsm3 octamer ring, for which a melting temperature of 70 °C has previously been recorded (*49*) .

Dynamic Nature of Lsm Rings in Solution. The intermolecular forces that determine quaternary ring formation and stacking of Lsm complexes appear to be perturbed by solution conditions. We found the distribution of oligomeric species of each Lsm polyprotein to be affected by both salt and pH. For Lsm[2+3], single tetramer rings appeared to dissociate to smaller forms (likely dimers) as salt concentrations were lowered to 50 mM in SEC runs. At the same time, the proportion of higher-molecular mass species (possibly 10-mers) in solution was increased. CD data confirm that the polyprotein remained in a fully folded tertiary state over the range of salt conditions, clearly retaining the characteristic ellipticity minimum at 215 nm. Thus, simple dissociation of the Lsm[2+3] tetrameric ring complexes (pI 5.5) appears to occur at low ionic strengths, pointing to the need to promote hydrophobic interactions to favor formation of a single-ring Lsm assembly. At low ionic strengths, diverse oligomeric states appear possible, at least for Lsm[2+3].

For both Lsm[2+3] and Lsm[4+1] preparations, a pH-dependent oligomerization was also observed. At pH <7, a greater amount of each polyprotein eluted within the SEC void, indicating formation oligomers over 600 kDa in size. In addition, fraction I material (the stacked ring form) became predominant over the tetrameric species (fraction II) for each polyprotein. These larger complexes associated with acidic conditions may involve stacking interactions between proximal faces mediated by the uncleaved hexa-His tag, an arrangement previously observed between stacked rings of tagged Lsm3 (*20*).

Protein Interactions of Lsm Oligomeric Rings. Given that our prepared Lsm polyprotein complexes organize as stable rings of four, their component subunit pairs can serve as partial mimics of binding epitopes presented by native heteromeric rings of Lsm proteins. We therefore utilized our simplified tetrameric rings of Lsm[2+3] and Lsm[4+1] within yeast pull-down experiments to probe specific protein partner interactions. The recombinant protein complexes were immobilized on a Ni^{2+} affinity matrix to act as bait proteins for whole cell yeast lysate, as previously performed for His-tagged Lsm3 (*20*). Protein partners of the discrete polyprotein complexes were then identified following in-gel trypsin digestion and MS analysis of component peptides. All protein−protein interactions identified in this manner in triplicate are detailed in the Supporting Information and summarized in Table 3.

The SDS−PAGE gel of Figure 3A shows that when immobilized octameric Lsm3 is exposed to yeast lysate, protein complexes recovered from the matrix yield two predominant bands, as well as a few less abundant species. The intense lower band (9−12 kDa) was found to release Lsm3 peptides (assumed to derive from bait protein), as well as others from three specific partners: yeast Lsm6 (9.4 kDa), Lsm5 (10.4 kDa), and Lsm2 (11.2 kDa). The sharp gel band at 25 kDa contained Lsm3 exclusively, attributed to the Lsm3 dimer. A faint band at 50 kDa showed the presence of translation elongation factor 1 (EF-1α).

Within heteromeric Lsm complexes in vivo, Lsm2 and Lsm6 are presumed to be immediate neighbors of the Lsm3 protomer (illustrated in Figure 1B). The capture of these two partner proteins directly from lysate by the Lsm3 bait indicates the integrity of this recombinant assembly as a scaffold for nativelike

Table 3: Interaction Partners of Lsm Polyproteins

(A) Protein Partners Isolated from Yeast Lysate[a]

	Lsm proteins							RNA-binding proteins		translation initiation factors			translation elongation factors		
	Lsm2	Lsm3	Lsm6	Lsm5	Lsm7	Lsm4	Lsm1	Pab1	Npl3	Ded1	eIF4E	eIF4G	EF-1α	EF2	EF3
Lsm3	+ (9)		+ (12)	+ (8)									+ (6)		
Lsm[2+3]			+ (9)	+ (5)	+ (3)						+ (2)		+ (12)		
Lsm[4+1][b]	+ (5)	+ (2)				+ (5)	+ (18)	+ (11)	+ (5)	+ (6)	+ (3)	+ (5)	+ (21)	+ (4)	+ (7)

(B) Oligonucleotide Interactions

	oligonucleotide			
protein bait	A_2U_5	U_{10}	$(AU)_5$	A_{10}
Lsm3				
Lsm[2+3]	++	+++	++	+
Lsm[4+1][b]	+	+		

[a]This is a summary of interactions detected via peptide LC−MS/MS from gel-isolated protein complexes, as detailed in the Supporting Information. Numerals in parentheses denote the number of unique peptides verifying partner protein captured within each gel, averaged for the three triplicate runs. [b]Interactions of Lsm[4+1] with proteins other than Lsm4, Lsm1, and Lsm3 are likely mediated via bound RNA.

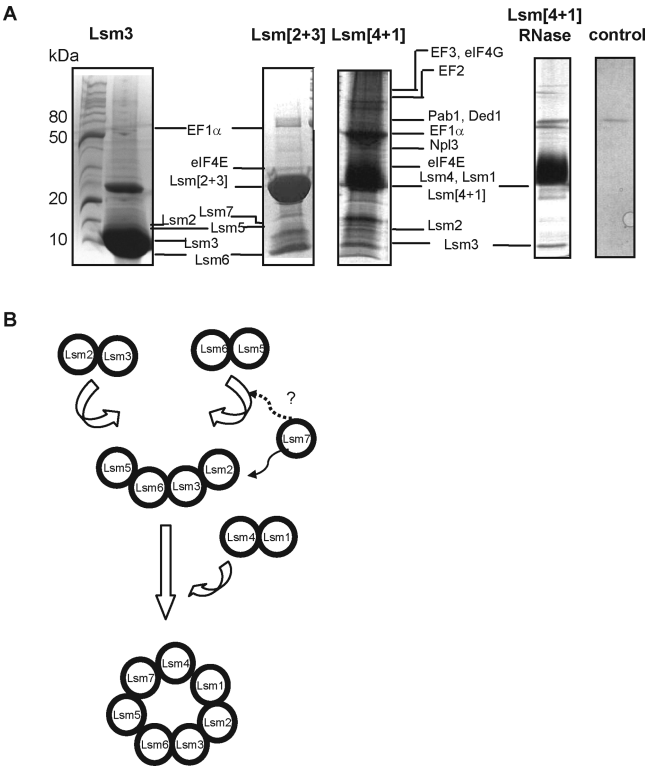

FIGURE 3: (A) SDS gels for detecting protein complexes pulled down from yeast lysate by N-terminally His_6-tagged Lsm proteins and polyproteins. Protein bands were visualized with Coomassie dye and labeled according to their identification by MS from released trypsin peptide fragments. The control lane shows the experiment conducted in the absence of any bait Lsm protein, with only a low-abundance bacterial contaminant eluted from the Ni-based matrix. (B) Proposed mechanism of heteromeric Lsm ring assembly based on the order of recruitment of Lsm component protomers observed in this work, as well as the proposed assembly order of the related Sm heptamer ring (*5*).

Lsm protein interactions, as previously reported (*20*). Lsm3 has been found elsewhere to interact with Lsm2, Lsm6, and Lsm5 in yeast systems (*32*, *50*) but has not been known to interact with EF-1α.

Using Lsm[2+3] (tetrameric form) as a bait protein, a new component has been introduced into the immobilized Lsm scaffold. This resulted in two extra proteins being recovered from yeast lysate, in addition to those noted for Lsm3. Within the lower region of the gel, Lsm7 (13 kDa) was detected along with Lsm5 and Lsm6, the known ring neighbors. A large band near 25 kDa (Figure 3A) was found to contain translation initiation factor eIF4E (24.3 kDa), as well as bait protein Lsm[2+3]. Although eIF4E has previously been identified as a protein partner for Lsm7 by TAP tagging (*31*), no such interaction has been documented previously for Lsm2 or Lsm3. Note that EF-1α again reappeared in this particular set of pull-down experiments (band at 50 kDa).

A large number of protein partners were recovered when yeast pull-down experiments were conducted with immobilized Lsm[4+1] (tetramer ring), despite the fact that this polyprotein comprises significantly truncated versions of the Lsm4 and Lsm1 sequences. Protein interaction partners identified from the resulting gels included Lsm2 and Lsm3 (Figure 3A). Importantly, peptides recovered (intense band at ~25 kDa) were consistent with sequences derived from the entire length of Lsm4 and Lsm1 (Supporting Information). These included peptides Lsm4:75−90 and Lsm1:18−59, segments lacking from the components of the engineered polyprotein. Thus, as well as indicating bait material (22.6 kDa), this band also reveals adhesion of native yeast Lsm4 and Lsm1 protein (21.2 and 20.3 kDa, respectively). Lsm4 and Lsm1 have been recruited directly from yeast lysate, in addition to just two other Lsm partners (Lsm2 and Lsm3).

Other proteins were also observed to bind to Lsm[4+1] in this affinity experiment: translation elongation factors EF-1α (50 kDa), EF2 (93.2 kDa), and EF3 (116 kDa); translation initiation factors eIF4E (24.3 kDa), eIF4G (107 kDa), and Ded1 (65.5 kDa); and mRNA poly A binding protein Pab1 (64.3 kDa) and mRNA binding protein Npl3 (45.4 kDa). However, when the pull-down experiment was repeated using yeast lysate pretreated with RNase, only the direct Lsm neighbor interactions were preserved (Figure 3A), including Lsm3, Lsm1, and Lsm4. It is thus likely that the secondary groups of partner proteins captured with Lsm[4+1] were engaged with the ring complex through bound RNA. Lsm4 and Lsm1 have not previously been seen to

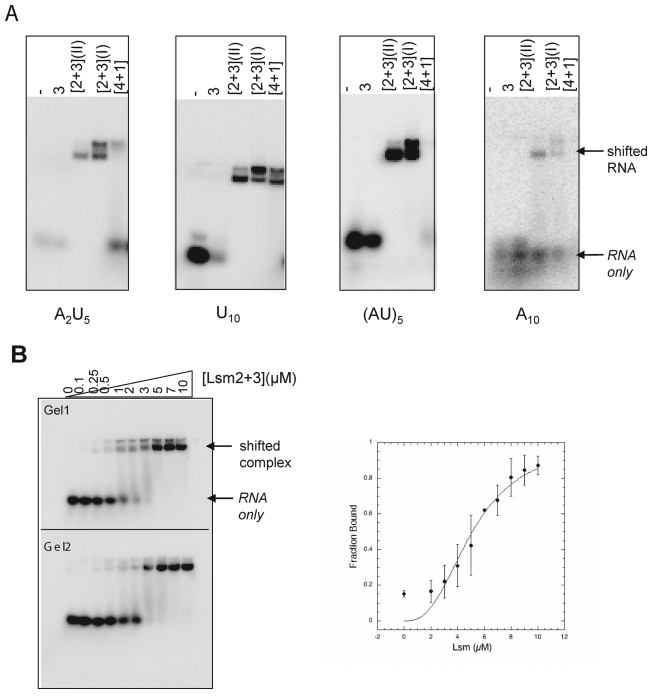

FIGURE 4: RNA interactions observed for Lsm polyproteins probed with ^{32}P-labeled sequences. (A) Electromobility shift assays are visualized on native PAGE gels following incubation of RNA with Lsm3, Lsm[2+3] fraction II, Lsm[2+3] fraction I, and Lsm[4+1]. Lanes are labeled according to Lsm sample, with the RNA control at the right (−). (B) Binding of A_2U_5 at a range of concentrations (0.1−10.0 μM) of Lsm[2+3] (fraction II). The fraction of RNA bound is plotted to determine the apparent K_D, in this case 5100 nM. Vertical bars show the standard deviation.

interact with translation initiation and elongation factors, although affinity experiments have documented Lsm4−Pab1, Lsm4−eIF4G, and Lsm1−Pab1 contacts (*30*).

RNA Binding Properties. To investigate the RNA binding properties of the simplified Lsm protein rings presented by our polypeptides, we conducted electrophoretic mobility shift assays (EMSAs) with ^{32}P-labeled RNA oligonucleotides. The four sequences investigated were A_2U_5, U_{10}, $(AU)_5$, and A_{10}, i.e., U- and A-rich sequences for which members of the L/Sm protein family are known to have some affinity (*10*, *16*, *17*, *19*, *22*, *23*, *51*).

Markedly different binding responses were observed for the Lsm polyprotein samples tested, comprising separated fractions of Lsm[2+3] and Lsm[4+1], as well as octameric Lsm3. Most surprisingly, the Lsm3 octamer ring (second lane of each panel of Figure 4A) did not show any detectable interaction with any of the oligonucleotides tested. In marked contrast, the Lsm[2+3] polyprotein (adjacent lanes) showed some degree of binding across the set of oligonucleotides. Although both forms of Lsm[2+3] bound this set of oligonucleotides, two separate species were resolved on the gels when octameric Lsm[2+3] bound to RNA. The Lsm[4+1] tetramer ring is observed to bind only U_{10} and A_2U_5 sequences, the latter more weakly (lane 5 of Figure 4A). Binding of U_{10} to Lsm[4+1] also appeared to promote formation of the two oligomeric states.

The contrast seen between oligonucleotide affinities of Lsm[2+3] and Lsm3 assemblies indicates that (i) the Lsm2 sequence might exclusively contain specific binding determinants, (ii) the Lsm2−Lsm3 interface is specifically required for RNA interactions, or (iii) the Lsm3 octamer may be induced in the presence of RNA to form a scaffold in which the octameric rings stack via proximal faces, occluding RNA-binding pockets.

We determined apparent K_D values for some of the Lsm−RNA interactions by conducting binding reactions over a range of protein concentrations (Figure 4B). For tetrameric Lsm[2+3] (fraction II material), K_D values were determined to be 5100 nM for A_2U_5, 4300 nM for $(AU)_5$, and 455 nM for U_{10}. A slight shift seen at high protein concentrations with A_{10} was too weak to be quantified for this specific complex. From these data, it is apparent that uracil-rich sequences have a 10-fold higher affinity for Lsm[2+3] than do sequences containing adenosine. For concentrations of Lsm[4+1] up to 20 μM, no binding to the same U_{10} or A_2U_5 RNA was detectable, signifying weaker RNA interactions for this truncated version of the protein complex. These values contrast with apparent K_D measurements in the nanomolar range found for intact Lsm[1−7] assemblies with more complex RNAs (*52*).

DISCUSSION

Oligomeric States of Mixed Lsm Rings. Past work with recombinant versions of Lsm proteins has demonstrated that these molecules self-associate to form hexamers (*6*, *22*), heptamers (*15*−*17*), and octamers (*20*) in vitro. Assembly of mixed rings was first demonstrated in Sm proteins where co-expressed Sm[D_1/D_2] and Sm[D_3/B] species assembled to form hexamers (*6*). The strategy was later extended to human Lsm proteins (*34*), where monomers, dimers, and trimers were used to reconstitute Lsm heptamers in vitro. This suggested that different Lsm combinations can be brought into ring structures in various combinations. We are adopting a polyprotein approach as a path to determining the structures of mixed rings of yeast Lsm protein assemblies.

The compositions of polyproteins devised to date for this work, Lsm[2+3], Lsm[4+1], and Lsm[5+6], were based on the specific dimer and trimer combinations thought to be stable intermediates in the formation of mixed heptameric Lsm proteins in vivo. Our laboratory complexes of Lsm[2+3] and Lsm[4+1], the latter containing truncated sequences, were highly soluble and seen to adopt stable ring structures. For these polyproteins, the most common oligomers observed were tetramers and a larger octamer form, likely to be stacking of two tetramer rings. In this study, we found the Lsm[5+6] polyprotein does not favor one specific oligomeric form. Although production of the yeast Lsm[6+5+7] polyprotein was attempted in expectation of a more robust protein species, the resulting recombinant product was insoluble.

Interestingly, the assemblies adopted by the recombinant Lsm polyproteins were seen to be affected by solution conditions. The dependence on ionic strength indicated that single-ring Lsm forms require conditions that promote intrasubunit hydrophobic interactions. Thus, certain solution conditions, such as high concentrations of nucleic acid, may exist in a cellular context which promote Lsm protein complexes to exist in smaller and less discrete oligomeric forms, and even to be somewhat dynamic. This may explain why discrete subsets of Lsm proteins are able to act as lysate binding partners with the recombinant Lsm rings.

Lsm−Lsm Interactions. The pull-down studies using tagged Lsm polyproteins as bait complexes have allowed us to probe specific recruitment and intersubunit interactions. The distinct interactions observed between Lsm proteins conformed to the accepted arrangement for the mixed ring group, based on our

current understanding of the related Sm heptamer. Thus, Lsm3 recruits Lsm2 and Lsm6, in conjunction with Lsm5. Incorporation of Lsm2 into the affinity bait alongside Lsm3 encourages the additional recruitment of Lsm7 to this grouping. When polyprotein Lsm[4+1] is exposed to yeast lysate, proteins Lsm2 and Lsm3 specifically bind, as do yeast-derived versions of Lsm1 and Lsm4. Most importantly, not every Lsm protein is recruited from lysate by these simplified complexes; if intact heptameric Lsm rings were recovered from yeast, all seven Lsm proteins would have been identified in our experiments.

The propensity of certain subgroups of Lsm proteins to be specifically captured from lysate allows us to also postulate the existence of other oligomeric combinations of these proteins in vivo. Two scenarios can thus be proposed with regard to the mechanism of Lsm protein recruitment in our affinity experiments. Our results are consistent with immobilized Lsm3 recruiting discrete Lsm[2/3] and Lsm[5/6] complexes directly from lysate. Alternatively, a dynamic nature of Lsm combinations raises the possibility that the Lsm ring structures may be capable of expanding or contracting to allow subunit addition or interchange. Thus, the Lsm3 octamer may reorganize to subsume other Lsm protomers from solution.

The specific affinities observed for our recombinant subcomplexes point toward a possible mode of assembly of the full mixed Lsm heptamer in vivo, for which a possible scenario is depicted in Figure 3B, strongly based on steps proposed previously for Sm proteins (*53*). Given that recombinant Lsm[2+3] did not recruit Lsm4 or Lsm1 (and vice versa for Lsm[4+1]), then Lsm5 and Lsm6 (possibly with Lsm7) must be recruited relatively early. Our study indicates that the resulting subcomplex (Lsm[2/3/5/6/7]) would subsequently recruit Lsm4 and Lsm1 components to form an active Lsm heptamer.

Possible Role of Lsm Proteins in Translation Repression. The processes of mRNA decapping and translation are highly intertwined, and there is a key transition step whereby eukaryotic mRNAs exit translation and are instead sequestered in an mRNP state that becomes accumulated into P-bodies (*54*). In yeast, several proteins involved in decapping (such as Pat1 and Dhh1) are known to have dual roles, functioning not only in mRNA decay but also in the general and active mechanism of translational repression (*55*). In mammals, Lsm4 and Lsm1 are among several proteins implicated in P-body formation through increasing the pool of translationally repressed mRNAs (*56*−*60*). There is as yet, however, no direct evidence in yeast of a role for the Lsm proteins in this mechanism of translation repression.

In this study, polyprotein Lsm[4+1] was seen to engage RNA as well as a set of RNA-binding proteins (Pab1 and Npl3), translation initiation factors (Ded1, eIF4E, and eIF4G), and translation elongation factors (EF1α, EF2, and EF3) (Table 3). The is an important observation, as it identifies an interaction between specific Lsm proteins and an RNP complex organized for translation initiation and elongation. We note this is consistent with the finding that Lsm1 cofractionates with polysomes (along with Pat1), indicating its association with translating ribosomes (*61*). Taken together, these studies suggest that yeast Lsm1 and/or Lsm4 can interact with a translationally active mRNA, pointing to a new role for these proteins. This is in contrast to the general belief that Lsm[1−7] is recruited onto mRNA only after it exits translation, destined for degradation (*62*).

Our results allow us to propose that some ring components of yeast Lsm assemblies are able to play a role in converting a translationally active mRNA to a repressed P-body mRNP, along with Pat1 and Dhh1. Recent results show that individual core P-body mRNPs aggregate into larger assemblies through interactions utilizing the Lsm4 Q/N-rich unstructured C-terminal extension (*27*). Thus, two functions may reside within Lsm4, the N-terminal folded domain having a role in translation repression and the C-terminus being involved in the buildup of large P-bodies.

While Lsm[1−7] has been shown to interact with Pat1 and associated proteins of mRNA decay (*12*, *32*, *63*, *64*), none of these protein factors were recruited by our Lsm subcomplexes. There is also a general lack of interaction observed here with splicing machinery. Since our Lsm[4+1] construct lacks most of the C-terminal residues of the wild-type sequences (103 deleted from Lsm4 and 51 deleted from Lsm1), we can concur that it is the strongly charged C-terminal segments extending well beyond the Sm fold that are required for recruitment of mRNA decay factors. Our polyprotein complexes do, however, retain the folded domains that organize into coherent affinity sites for some of the components known to be associated with Lsm biology.

RNA Binding by the Lsm Rings. At this stage, the RNA interaction sites within Lsm ring assemblies are not extensively well-defined. From structures of Lsm assemblies bound with short oligonucleotides studied by crystallography and cryo-EM, it appears that RNA winds around the central pore of the ring and likely also passes through it. Additional RNA binding sites have been documented for some Lsm proteins: the bacterial protein Hfq is shown to have a distinct second RNA binding site at the distal face (*24*), and an "external" RNA binding site was defined in the crystal structure of Pa-Sm1 with bound nucleotides.

Our simplified mixed Lsm complexes prepared in this work provide a first opportunity to probe binding and specificity for different Lsm protein ring components. Surprisingly, octameric ring protein Lsm3 did not bind any of the RNA oligonucleotides tested. However, when present in combination with its ring neighbor Lsm2, detectable interactions with a number of oligonucleotides were observed, most strongly with purely U-rich sequences. From the known structures of Lsm−RNA complexes, highly conserved residues of loops L3 and L5 are generally implicated in RNA binding: Asp and Asn in loop L3 and Arg and Gly in loop L5 (*6*, *15*). In addition, a highly conserved aromatic residue in loop L3 is responsible for stacking of the bases of the nucleotide (Figure 5). Sequences of the components are homologous in Lsm2 and Lsm3 proteins: D-S/Q-H/F-C/L-N (loop 3) and I-R-G-D/S (loop 5). The crystal structure of octameric Lsm3 and the derived model for the Lsm[2+3] complex show that all binding features are present within the subunit clefts on both complexes (Figure 5C,D). Moreover, the conserved residues cluster correctly around the intersubunit cleft on the proximal face to be able to capture RNA. Interacting side chains available for RNA binding include Phe35 from Lsm2 and His36 from Lsm3, as well relevant hydrogen bonding groups. Consideration of the local chemical environment of these binding residues in Lsm2 and Lsm3 shows no immediate differences to rationalize the functional difference (Figure 5).

The polyprotein complex of Lsm[4+1] interacted comparatively weakly with the four RNA oligonucleotides tested in this study. However, yeast pull-down experiments showed this same Lsm ring is able to recruit an RNP assembly in an RNA-dependent process. It is thus likely that these particular Lsm

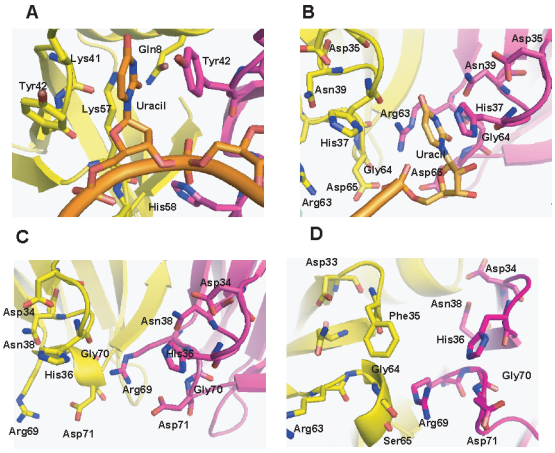

FIGURE 5: Comparison of RNA binding active sites of Lsm proteins. (A) Sa-Hfq [Protein Data Bank (PDB) entry 1KQ2]. (B) AF-Sm1 (PDB entry 1I5L). (C) *S. cerevisiae* Lsm3 (PDB entry 3BW1). (D) Lsm[2+3] model (*20*) (Lsm2 colored yellow and Lsm3 colored pink). Side chains making contact with RNA are labeled.

proteins are optimized for RNA sequences other than those tested here.

Previous work has shown that the human Lsm[2−8] complex reconstituted from coexpressed components is able to bind with U6 snRNA in vitro (*34*). However, when the components themselves were tested (i.e., Lsm[2/3], Lsm[4/8], and Lsm[5/6/7]), none interacted with the snRNA. Kambach and workers have pointed to the need for strong RNA target discrimination on the basis of the presence or absence of a single specific Lsm subunit. Recent UV cross-linking studies with Lsm proteins and U6 snRNA have revealed that the Lsm2 component of the intact Lsm[2−8] ring forms the interface between the snRNA and Prp24 (*65*). The observations made in this study of differential RNA binding for the yeast Lsm polyprotein complexes, in spite of similar oligomeric forms, confirm a degree of target discrimination by different Lsm proteins.

ACKNOWLEDGMENT

We are grateful to Geoff Kornfeld (University of New South Wales) for yeast genetic material and intronless Lsm2, Junior T'eo (Macquarie University) for advice on overlap PCR, Jacquie Matthews and Daniel Ryan (University of Sydney) for access to MALLS instrumentation, Debra Birch (Macquarie University) for electron microscopy, Nishen Naidoo (Macquarie University) for coordinates of Lsm ring models, Karlie Nelson (Macquarie University) for peptide MS−MS, and Pia Hønnerup Jensen for MALDI-MS data. We thank Martin Reijns and Paul Curmi for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Data and parameters for the MS analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Beggs, J. D. (2005) Lsm proteins and RNA processing. *Biochem. Soc. Trans. 33*, 433–438.
2. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S., and Seraphin, B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J. 18*, 3451–3462.
3. He, W., and Parker, R. (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Curr. Opin. Cell Biol. 12*, 346–350.
4. Verdone, L., Galardi, S., Page, D., and Beggs, J. D. (2004) Lsm proteins promote regeneration of pre-mRNA splicing activity. *Curr. Biol. 14*, 1487–1491.
5. Nagai, K., Muto, Y., Pomeranz Krummel, D. A., Kambach, C., Ignjatovic, T., Walke, S., and Kuglstatter, A. (2001) Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. *Biochem. Soc. Trans. 29*, 15–26.
6. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J., and Nagai, K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell 96*, 375–387.
7. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J., and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature 458*, 475–480.
8. Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007) The highways and byways of mRNA decay. *Nat. Rev. 8*, 113–126.
9. Boeck, R., Lapeyre, B., Brown, C. E., and Sachs, A. B. (1998) Capped mRNA degradation intermediates accumulate in the yeast spb8-2 mutant. *Mol. Cell. Biol. 18*, 5062–5072.
10. Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007) The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA 13*, 998–1016.
11. Tharun, S., and Parker, R. (2001) Targeting an mRNA for decapping: Displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol. Cell 8*, 1075–1083.
12. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000) Yeast Sm-like proteins function in mRNA decapping and decay. *Nature 404*, 515–518.
13. Teixeira, D., and Parker, R. (2007) Analysis of P-body assembly in *Saccharomyces cerevisiae*. *Mol. Biol. Cell 18*, 2274–2287.
14. Sheth, U., and Parker, R. (2003) Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science 300*, 805–808.
15. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M., and Mabbutt, B. C. (2001) Crystal structure of a heptameric Sm-like protein complex from archaea: Implications for the structure and evolution of snRNPs. *J. Mol. Biol. 309*, 915–923.
16. Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J. 20*, 2293–2303.
17. Mura, C., Cascio, D., Sawaya, M. R., and Eisenberg, D. S. (2001) The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proc. Natl. Acad. Sci. U.S.A. 98*, 5532–5537.
18. Collins, B. M., Cubeddu, L., Naidoo, N., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M., and Mabbutt, B. C. (2003) Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF. *J. Biol. Chem. 278*, 17291–17298.
19. Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003) Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *J. Biol. Chem. 278*, 1239–1247.
20. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R., Curmi, P. M., and Mabbutt, B. C. (2008) Crystal structure of Lsm3 octamer from *Saccharomyces cerevisiae*: Implications for Lsm ring organisation and recruitment. *J. Mol. Biol. 377*, 1357–1371.
21. Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M., and Eisenberg, D. (2003) The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). *Protein Sci. 12*, 832–847.
22. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P., and Brennan, R. G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: Aa bacterial Sm-like protein. *EMBO J. 21*, 3546–3556.
23. Brennan, R. G., and Link, T. M. (2007) Hfq structure, function and ligand binding. *Curr. Opin. Microbiol. 10*, 125–133.
24. Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D., and Feig, A. L. (2004) *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat. Struct. Mol. Biol. 11*, 1206–1214.
25. Tharun, S., Muhlrad, D., Chowdhury, A., and Parker, R. (2005) Mutations in the *Saccharomyces cerevisiae* LSM1 gene that affect mRNA decapping and 3′ end protection. *Genetics 170*, 33–46.
26. Decker, C. J., Teixeira, D., and Parker, R. (2007) Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J. Cell Biol. 179*, 437–449.

27. Reijns, M. A., Alexander, R. D., Spiller, M. P., and Beggs, J. D. (2008) A role for Q/N-rich aggregation-prone regions in P-body localization. *J. Cell Sci. 121*, 2463–2472.
28. Fromont-Racine, M., Mayes, A. E., Brunet-Simon, A., Rain, J. C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J. D., and Legrain, P. (2000) Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast 17*, 95–110.
29. Neubauer, G. (2005) The analysis of multiprotein complexes: The yeast and the human spliceosome as case studies. *Methods Enzymol. 405*, 236–263.
30. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature 440*, 631–636.
31. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature 415*, 141–147.
32. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature 440*, 637–643.
33. Wilusz, C. J., and Wilusz, J. (2005) Eukaryotic Lsm proteins: Lessons from bacteria. *Nat. Struct. Mol. Biol. 12*, 1031–1036.
34. Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K., and Kambach, C. (2005) Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J. Biol. Chem. 280*, 16066–16075.
35. Farr, G. W., Furtak, K., Rowland, M. B., Ranson, N. A., Saibil, H. R., Kirchhausen, T., and Horwich, A. L. (2000) Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell 100*, 561–573.
36. Heddle, J. G., Yokoyama, T., Yamashita, I., Park, S. Y., and Tame, J. R. (2006) Rounding up: Engineering 12-membered rings from the cyclic 11-mer TRAP. *Structure 14*, 925–933.
37. Te'o, V. S., Cziferszky, A. E., Bergquist, P. L., and Nevalainen, K. M. (2000) Codon optimization of xylanase gene xynB from the thermophilic bacterium *Dictyoglomus thermophilum* for expression in the filamentous fungus *Trichoderma reesei*. *FEMS Microbiol. Lett. 190*, 13–19.
38. Sauter, C., Basquin, J., and Suck, D. (2003) Sm-like proteins in Eubacteria: The crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res. 31*, 4091–4098.
39. Schagger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem. 166*, 368–379.
40. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol. 128*, 82–97.
41. Andon, N. L., Hollingworth, S., Koller, A., Greenland, A. J., Yates, J. R., III, and Haynes, P. A. (2002) Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. *Proteomics 2*, 1156–1168.
42. Breci, L., Hattrup, E., Keeler, M., Letarte, J., Johnson, R., and Haynes, P. A. (2005) Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics 5*, 2018–2028.
43. Cooper, B., Eckert, D., Andon, N. L., Yates, J. R., and Haynes, P. A. (2003) Investigative proteomics: Identification of an unknown plant virus from infected plants using mass spectrometry. *J. Am. Soc. Mass Spectrom. 14*, 736–741.
44. Craig, R., and Beavis, R. C. (2004) TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics 20*, 1466–1467.
45. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J. Proteome Res. 2*, 43–50.
46. Sambrook, J., and Russell, D. W. (2000) Molecular cloning: A laboratory manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
47. Garner, M. M., and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: Application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res. 9*, 3047–3060.
48. Jelinska, C., Conroy, M. J., Craven, C. J., Hounslow, A. M., Bullough, P. A., Waltho, J. P., Taylor, G. L., and White, M. F. (2005) Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure 13*, 963–971.
49. Naidoo, N. (2006) Molecular structures of the RNA-binding Lsm proteins. Ph.D. Thesis, Macquarie University, Sydney.
50. Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M., and Seraphin, B. (2000) A Sm-like protein complex that participates in mRNA degradation. *EMBO J. 19*, 1661–1671.
51. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, 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.
52. Chowdhury, A., and Tharun, S. (2008) lsm1 mutations impairing the ability of the Lsm1p-7p-Pat1p complex to preferentially bind to oligoadenylated RNA affect mRNA decay in vivo. *RNA 14*, 2149–2158.
53. Raker, V. A., Hartmuth, K., Kastner, B., and Luhrmann, R. (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol. 19*, 6554–6565.
54. Coller, J., and Parker, R. (2004) Eukaryotic mRNA decapping. *Annu. Rev. Biochem. 73*, 861–890.
55. Coller, J., and Parker, R. (2005) General translational repression by activators of mRNA decapping. *Cell 122*, 875–886.
56. Andrei, M. A., Ingelfinger, D., Heintzmann, R., Achsel, T., Rivera-Pomar, R., and Luhrmann, R. (2005) A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA 11*, 717–727.
57. Ferraiuolo, M. A., Basak, S., Dostie, J., Murray, E. L., Schoenberg, D. R., and Sonenberg, N. (2005) A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay. *J. Cell Biol. 170*, 913–924.
58. Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J. C., Fritzler, M. J., and Chan, E. K. (2005) Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell Biol. 7*, 1267–1274.
59. Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M. J., Scheuner, D., Kaufman, R. J., Golan, D. E., and Anderson, P. (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol. 169*, 871–884.
60. Pauley, K. M., Eystathioy, T., Jakymiw, A., Hamel, J. C., Fritzler, M. J., and Chan, E. K. (2006) Formation of GW bodies is a consequence of microRNA genesis. *EMBO Rep. 7*, 904–910.
61. Bonnerot, C., Boeck, R., and Lapeyre, B. (2000) The two proteins Pat1p (Mrt1p) and Spb8p interact in vivo, are required for mRNA decay, and are functionally linked to Pab1p. *Mol. Cell. Biol. 20*, 5939–5946.
62. Parker, R., and Sheth, U. (2007) P bodies and the control of mRNA translation and degradation. *Mol. Cell 25*, 635–646.
63. Collins, S. R., Kemmeren, P., Zhao, X. C., Greenblatt, J. F., Spencer, F., Holstege, F. C., Weissman, J. S., and Krogan, N. J. (2007) Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics 6*, 439–450.
64. Tarassov, K., Messier, V., Landry, C. R., Radinovic, S., Serna Molina, M. M., Shames, I., Malitskaya, Y., Vogel, J., Bussey, H., and Michnick, S. W. (2008) An in vivo map of the yeast protein interactome. *Science 320*, 1465–1470.
65. Karaduman, R., Dube, P., Stark, H., Fabrizio, P., Kastner, B., and Luhrmann, R. (2008) Structure of yeast U6 snRNPs: Arrangement of Prp24p and the LSm complex as revealed by electron microscopy. *RNA 14*, 2528–2537.