MOLECULAR AND CELLULAR BIOLOGY, Nov. 2005, p. 10989–11004 Vol. 25, No. 24
0270-7306/05/$08.00+0 doi:10.1128/MCB.25.24.10989–11004.2005

Specific Sequence Features, Recognized by the SMN Complex, Identify snRNAs and Determine Their Fate as snRNPs

Tracey J. Golembe, Jeongsik Yong, and Gideon Dreyfuss*

Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Received 22 June 2005/Returned for modification 27 July 2005/Accepted 19 September 2005

The survival of motor neurons (SMN) complex is essential for the biogenesis of spliceosomal small nuclear ribonucleoproteins (snRNPs) as it binds to and delivers Sm proteins for assembly of Sm cores on the abundant small nuclear RNAs (snRNAs). Using the conserved snRNAs encoded by the lymphotropic *Herpesvirus saimiri* (HVS), we determined the specific sequence and structural features of RNAs for binding to the SMN complex and for Sm core assembly. We show that the minimal SMN complex-binding domain in snRNAs, except U1, is comprised of an Sm site (AUUUUUG) and an adjacent 3′ stem-loop. The adenosine and the first and third uridines of the Sm site are particularly critical for binding of the SMN complex, which directly contacts the backbone phosphates of these uridines. The specific sequence of the adjacent stem (7 to 12 base pairs)-loop (4 to 17 nucleotides) is not important for SMN complex binding, but it must be located within a short distance of the 3′ end of the RNA for an Sm core to assemble. Importantly, these defining characteristics are discerned by the SMN complex and not by the Sm proteins, which can bind to and assemble on an Sm site sequence alone. These findings demonstrate that the SMN complex is the identifier, as well as assembler, of the abundant class of snRNAs in cells because it is able to recognize an snRNP code that they contain.

Pre-mRNA splicing takes place in the nucleus of eukaryotic cells and is mediated by the spliceosome. The major components of the spliceosome are the small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U5, and U4/U6, each of which is comprised of one U snRNA molecule, a common core of seven Sm proteins (B/B′, D1, D2, D3, E, F, and G), and several snRNP-specific proteins (31, 32, 66). The biogenesis of snRNPs occurs in the cytoplasm shortly after the nuclear export of nascent snRNAs and requires the assembly of the Sm proteins into a seven-membered ring (25, 63) on a consensus sequence ($PuAU_{4-6}GPu$) known as the Sm site of the U snRNA (2, 48). After formation of the Sm core, the 7-methyl guanosine cap of the snRNA is hypermethylated to become a 2,2,7-trimethyl guanosine cap (35, 57). A properly assembled Sm core, cap hypermethylation, and 3′ end processing are required for the translocation of the mature snRNPs into the nucleus, where they function in splicing (10, 11, 20, 36, 37, 66).

The process of bringing the protein and RNA components together during U snRNP assembly in the cytoplasm is mediated by and dependent upon the survival of motor neurons (SMN) protein complex (3, 9, 29, 30, 40, 43, 55, 56, 67–69). Reduced levels of SMN due to a genetic defect cause spinal muscular atrophy, a severe neuromuscular disease that is characterized by degeneration of motor neurons in the spinal cord (6, 7, 21). SMN, as an oligomeric protein, is part of a large multiprotein complex that contains Gemin2 (30), the DEAD box RNA helicase Gemin3 (4), Gemin4 (5), Gemin5 (19), Gemin6 (52), and Gemin7 (1). Although the function of the SMN complex in snRNP assembly is its best-characterized activity, it most likely functions in the assembly, metabolism, and transport of various other RNPs, including snoRNPs, miRNPs, and the machineries that carry out transcription and pre-mRNA splicing (3, 12, 18, 24, 39, 44, 45, 49, 51, 53–55).

Purified snRNP total proteins, a preparation referred to as TPs, readily assemble an Sm core on a minimal Sm sequence oligonucleotide in vitro without ATP hydrolysis or other non-snRNP factors (59, 60, 64). However, in cell extracts the biogenesis of U snRNPs requires ATP hydrolysis (26, 40, 56), suggesting that snRNP proteins are not free to randomly associate with any uridine-rich RNA sequences in cells. Rather, it is the SMN complex that actively brings Sm proteins to U snRNAs, acting as a crucial specificity factor to ensure that highly stable Sm cores are only assembled on the correct snRNAs (56, 67, 69). Several components of the SMN complex bind directly to the Sm proteins, including the binding of SMN to the RG-rich C-terminal domains of the Sm proteins B, D1, and D3 (1, 1a, 3–5, 12, 19, 30, 52, 53). This interaction is enhanced by the symmetric dimethylarginine modification of specific arginines by the 20S methylosome that contains an arginine methyltransferase (JBP1/PRMT5) (13–15, 41). The SMN complex also binds directly and with sequence specificity to the Sm site containing U snRNAs. For U1 snRNA, the SMN complex-binding domain is contained within stem-loop 1 (68). For the other major U snRNAs, U2, U4, and U5, the minimal SMN complex-binding domains are closer to their 3′ ends and contain the Sm site and the 3′ stem-loop. These SMN complex-binding domains are necessary and sufficient for SMN complex binding and SMN-dependent assembly of Sm cores (67, 69). Previous studies have suggested that the SMN complex contains at least two separate high-affinity RNA binding domains—one for U1 snRNA and the other for U2, U4, and U5 snRNAs (67, 69).

* Corresponding author. Mailing address: Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 328 CRB, 415 Curie Blvd., Philadelphia, PA 19104-6148. Phone: (215) 898-0398. Fax: (215) 573-2000. E-mail: gdreyfuss@hhmi.upenn.edu.

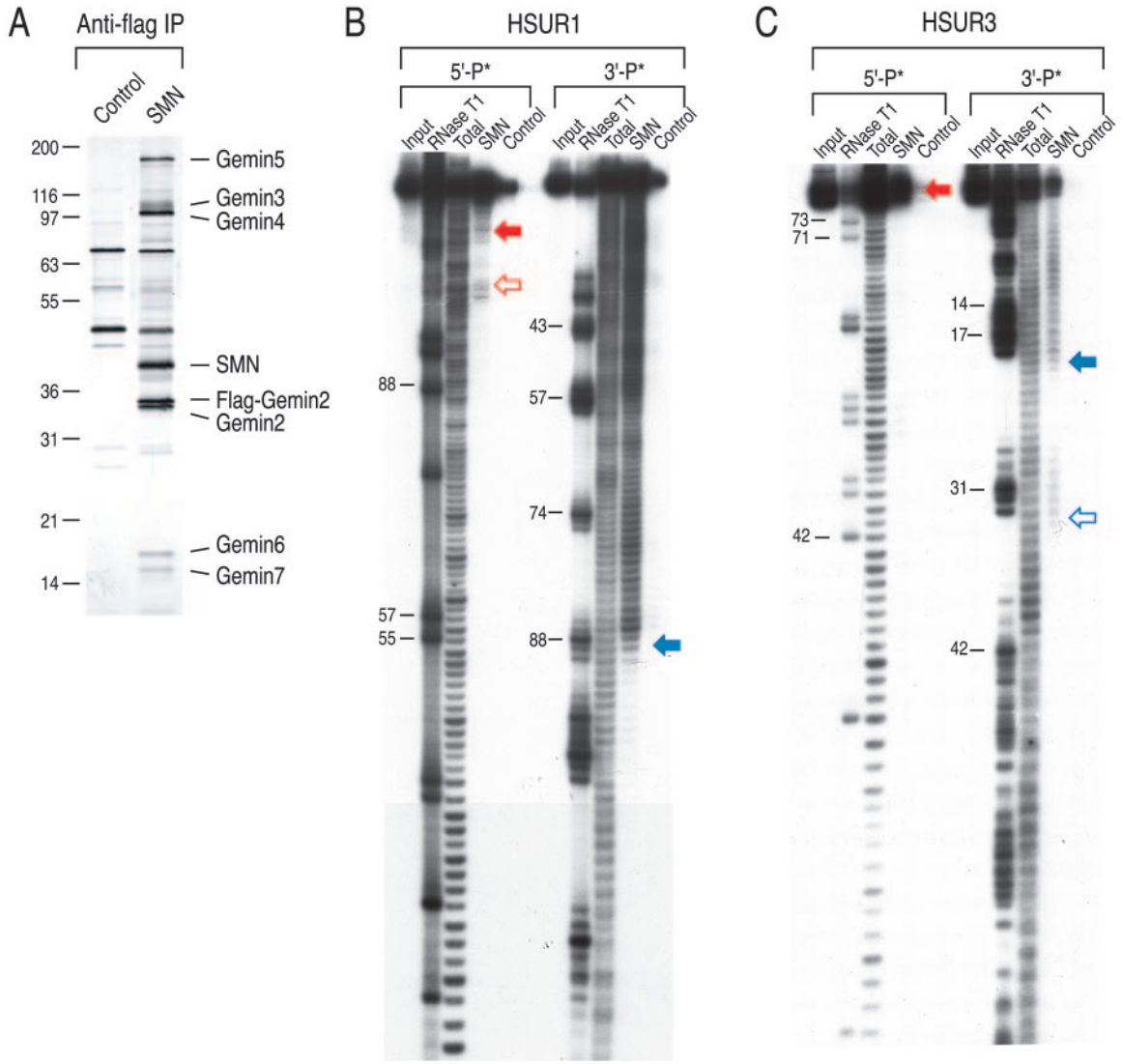


FIG. 1. Mapping and direct binding of the minimal SMN complex-binding domains of the HSURs. (A) Native SMN complexes (SMN) were purified under high-salt conditions from stable cell lines expressing flag-Gemin2 (as described in Materials and Methods) and were analyzed by electrophoresis on 12.5% sodium dodecyl sulfate-polyacrylamide gels and silver staining. Immunoprecipitation using anti-flag antibody from the parental HeLa cell line was used as a control (Control). Core components of the SMN complex are labeled based upon molecular weight and Western blotting. Under these conditions, there is no detectable association of SMN complex with Sm proteins. (B) The SMN complex-binding domain of HSUR1. The 5′ (5′-P*)- and 3′ (3′-P*)-end-labeled HSUR1 was subjected to limited alkaline hydrolysis. The resulting hydrolyzed RNA ladders were incubated with purified SMN complex (SMN) or nonspecific proteins purified from HeLa cells (Control). The RNA fragments bound to the SMN complex were isolated and analyzed by electrophoresis on 7 M urea–8% acrylamide gels. The full-length RNA was digested with RNase T1 to provide a size marker. The solid red and blue arrows indicate the largest region that includes the SMN complex-binding domain, and open arrows indicate the smallest possible binding regions. For HSUR1, the open red arrow most likely indicates nonspecific degradation also seen in the input lane. Total represents 5% of input. (C) The SMN complex-binding domain of HSUR3. The same experiment as described in panel B was performed using 5′- and 3′-end-labeled HSUR3. The solid red and blue arrows indicate the largest region that includes the SMN complex-binding domain, and open arrows indicate the smallest possible binding regions. (D) The 5′ end deletion mutants of HSUR1 were transcribed in the presence of [^{32}P]UTP and incubated with flag-purified SMN complex or nonspecific HeLa control proteins (control) for 1 h at 4°C. Bound RNAs were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. Total represents 10% of input. (E) The 5′ end deletion mutants of HSUR3 were transcribed in the presence of [^{32}P]UTP and subjected to SMN complex binding as described in panel D. (F) The 5′-end deletion mutants of HSUR4 were transcribed in the presence of [^{32}P]UTP and subjected to SMN complex binding as described in panel D. Control not shown. (G) The 5′-end deletion mutants of HSUR5 were transcribed in the presence of [^{32}P]UTP and subjected to SMN complex binding as described in panel D. Control not shown.

Recently, we have shown that the *Herpesvirus saimiri* (HVS)-encoded small nuclear RNAs (HSURs 1 to 7) (46) also use the SMN complex to assemble Sm cores (16). HVS strain 11, the prototype gamma 2 herpesvirus, causes acute leukemias and T-cell lymphomas in some New World primates (8). The HSURs bind the SMN complex with very high affinity to the U4 snRNA-type binding site and can effectively compete with host snRNAs for snRNP assembly (16). Because the Sm sites

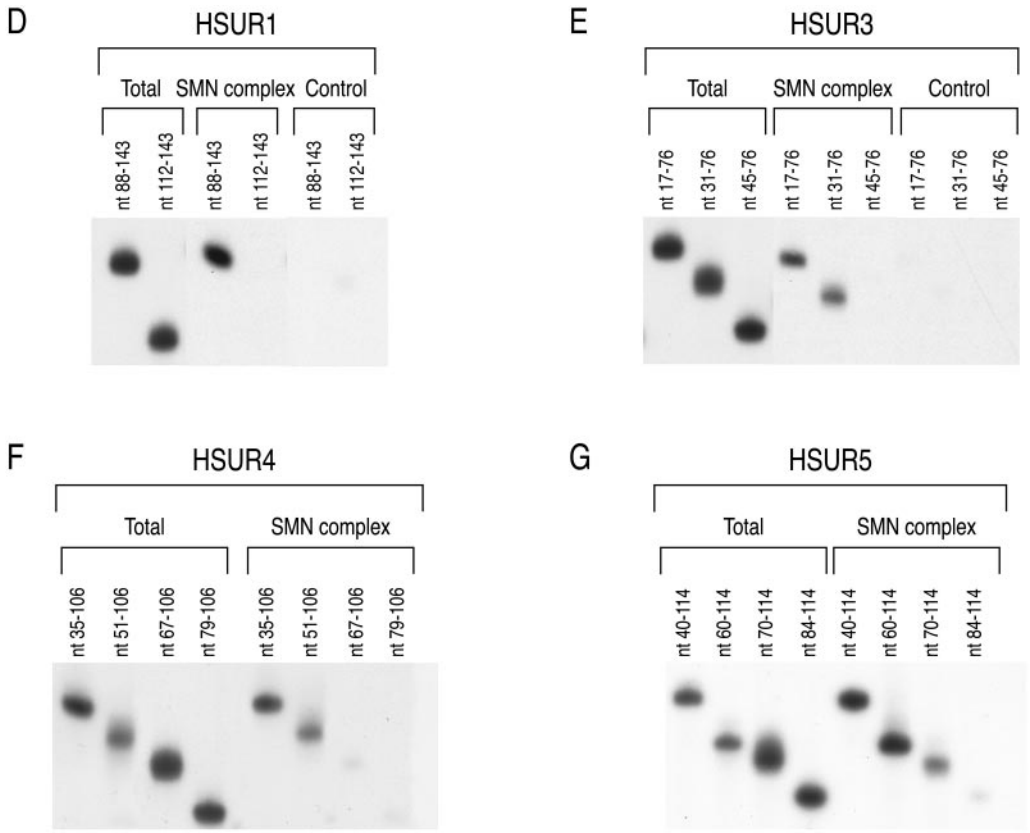


FIG. 1—*Continued*.

and predicted secondary structures of the HSURs are simple, conserved, and have apparently evolved to closely resemble those of U2, U4, and U5 snRNAs (27, 28), it seems likely that their RNA motifs required for SMN complex binding and Sm core assembly might also be conserved. Therefore, we used the HSURs as a tool to identify the minimal RNA elements required for SMN complex binding and Sm core assembly.

Here we describe the essential features in an RNA that distinguish it and define it as an snRNA on which an Sm core will assemble, and we demonstrate that the SMN complex decodes these specific sequence features of an snRNA to determine its fate as a snRNP.

MATERIALS AND METHODS

Plasmids for in vitro transcription. Plasmids for in vitro transcription of HSURs 1, 3, 4, and 5 are described elsewhere (16). Construction of all deletion mutants of HSUR 1, 3, 4, and 5 cDNAs and the HSUR5-60 with a destabilized terminal stem (des-stem) cDNA was carried out by PCR according to Imai et al. (22). The HSUR4 and HSUR5 stem, swap, and flip constructs were synthesized as DNA oligonucleotides (Dharmacon) that were cloned into pGem-3Z. Fifteen and 70 nucleotides (nt) were added to HSUR5-60 RNA by digestion of carrier vector with restriction enzymes (HindIII and MseI, respectively) prior to runoff transcription.

Labeling of RNAs. In vitro transcription and [^{32}P]UTP labeling of RNAs were carried out as described previously (68). [^{32}P]UTP-labeled RNAs were purified by electrophoresis on 7 M urea–6% polyacrylamide gels and precipitated with ethanol. RNAs were resuspended in deionized distilled water. The 5′ or 3′ end labeling of HSURs was carried out as described elsewhere (70).

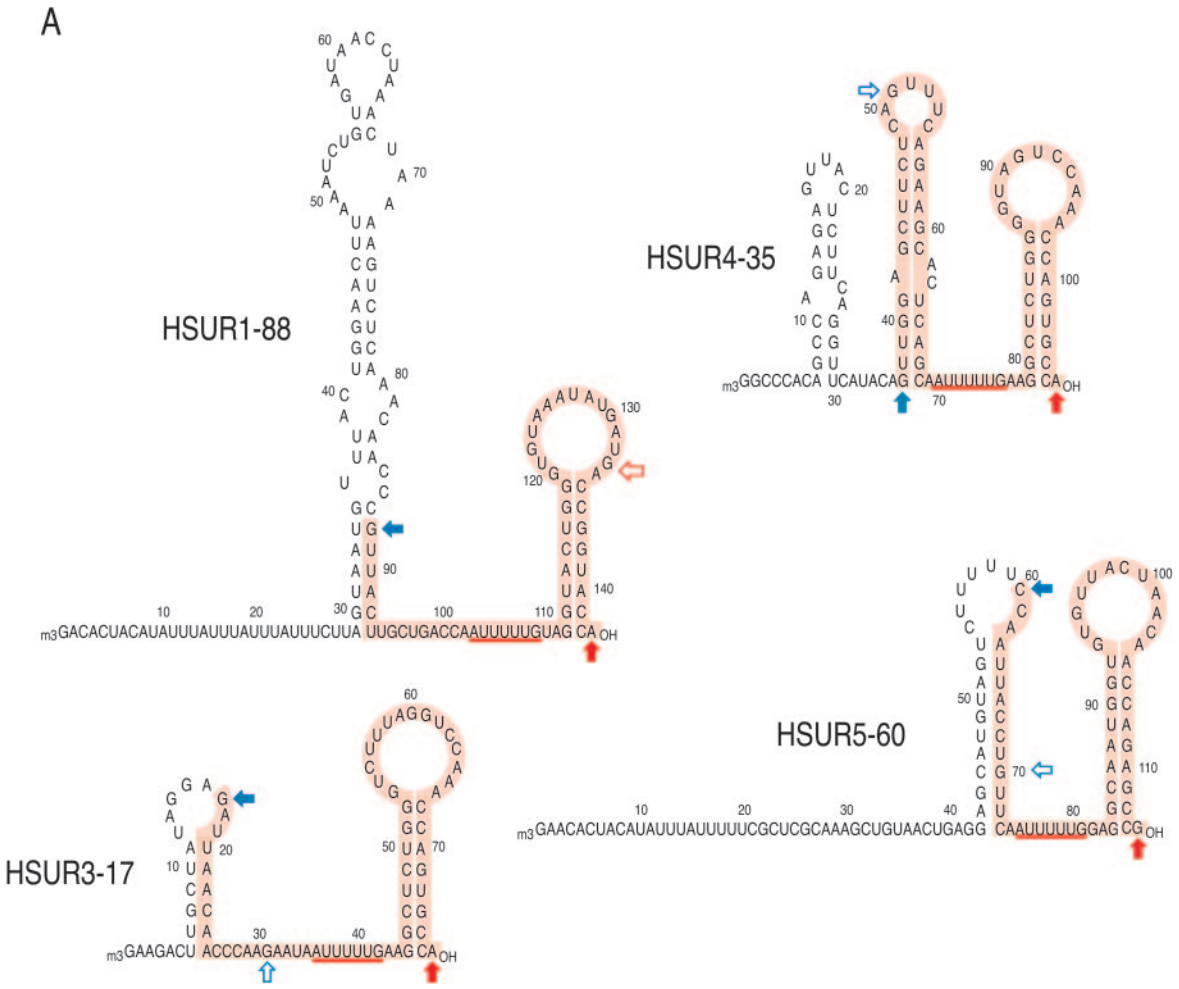
Limited alkaline hydrolysis and minimal binding analysis. Limited alkaline hydrolysis was carried out as described previously (70).

Purification of native SMN complex. The SMN complex was purified from flag-Gemin2 HeLa Tet-ON cells as described previously (67). The parental HeLa cell line served as a negative control. For purification of SMN complex under low-salt conditions, SMN complex or control bound to anti-flag beads (Sigma) was washed extensively with RSB-100 (10 mM Tris-HCl [pH 7.5], 100 mM NaCl, 2.5 mM $MgCl_2$) containing 0.02% NP-40. For complex purification under more stringent conditions, three additional washes were performed for 15 min each at 4°C with 10 bead volumes of RSB-500 containing 0.02% NP-40. The bound proteins were either equilibrated with 10 bead volumes of RSB-100 containing 0.01% NP-40 for binding experiments or eluted for 1 h at 4°C with 3× Flag peptide (Sigma) at a final concentration of 0.5 mg/ml for analysis by silver staining or Western blotting. Proteins were resolved on sodium dodecyl sulfate–12.5% polyacrylamide gels or NuPAGE Novex Bis-Tris precast gradient 4 to 12% mini-gels (Invitrogen). The following mouse monoclonal antibodies were used for Western blotting: 2B1 (anti-SMN), 12H12 (anti-Gemin3), 17D10 (anti-Gemin4), 10G11 (anti-Gemin5), Y12 (anti-Sm), and 3C2 (hnRNP K). A rabbit polyclonal antibody was used to detect Gemin6.

In vitro binding of RNAs. In vitro binding experiments were performed as previously described (67). The bound RNAs were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels.

Preparation of HeLa cell cytoplasmic extracts. HeLa cell cytoplasmic extracts competent for snRNP assembly were prepared as described previously (67) and immunodepletion of SMN complex was performed as described by Golembe et al. using anti-SMN monoclonal antibody (2B1) (16).

Assay for assembly of snRNPs. In vitro Sm core assembly and electrophoretic mobility shift assays were carried out as described previously (56). To quantitate the Sm core assembly, assembly reaction products were incubated with 4 μg of anti-Sm (Y12) monoclonal antibody or control antibody (SP2/0) conjugated to protein A-Sepharose CL-4B (Amersham) in RSB-500 (10 mM Tris-HCl [pH 7.5], 500 mM NaCl, 2.5 mM $MgCl_2$) containing 0.1% NP-40 and 2 mg/ml heparin for 1 h at 4°C. Subsequently, the beads were washed five times with binding buffer and treated with proteinase K, and the immunoprecipitated RNAs were purified


A
HSUR1-88
HSUR4-35
HSUR5-60
HSUR3-17
B
HSUR3
HSUR3-17
HSUR4
HSUR4-35
HSUR5
HSUR5-60
-
CE
ΔSMN
Sm core
Free RNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

by phenol-chloroform extraction and ethanol precipitation. RNAs were analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography.

RNase T1 digestion. HSUR5-60 RNAs were labeled at the 5′ end with [γ-^{32}P]ATP as described previously (70), and 20,000 cpm of each RNA was then incubated with RNase T1 (Ambion) according to the manufacturer's protocol. Digested RNAs were treated with proteinase K, and the RNAs were purified by phenol-chloroform extraction and ethanol precipitation and analyzed by electrophoresis on 7 M urea–12% polyacrylamide gels and autoradiography.

Phosphorothioate interference mapping. HSUR4-35 and HSUR5-60 RNAs were transcribed using T7 polymerase in the presence of adenosine, uridine, guanosine, or cytidine (A, U, G, or C) α-thiotriphosphates (Glen Research) as described by Ryder et al. (61). The RNAs were purified by electrophoresis on 7 M urea–6% polyacrylamide gels, precipitated with ethanol, and labeled at the 5′ end with [γ-^{32}P]ATP as described previously (70). A total of 100,000 cpm of each RNA was incubated with high-salt-washed SMN complex or nonspecific purified HeLa cell proteins immobilized on anti-flag beads (Sigma) in RSB-100 containing 0.01% NP-40 for 1 h at 4°C in the presence of 1 μM tRNA. For the interference experiments using purified snRNP TPs, 500,000 cpm of HSUR5-60 was incubated with purified TPs for 1 h at 30°C in the same buffer. RNAs with assembled Sm cores were immunopurified using Y12 antibody as described for snRNP assembly. Subsequently, the beads were washed five times with binding buffer and treated with proteinase K, and the RNAs were purified by phenol-chloroform extraction and ethanol precipitation. Precipitated RNAs were resuspended in 10 ml of loading buffer and cleaved at sites of phosphorothioate incorporation by addition of 1/10 volume of 2 mM I_2 in ethanol. Equivalent cpm counts of RNA not subjected to SMN complex binding (input) were cleaved with I_2 and run in parallel to normalize for phosphorothioate incorporation. The RNA fragments were resolved on either 7 M urea–10% polyacrylamide or 7 M urea–12% polyacrylamide sequencing gels for HSUR4-35 and HSUR5-60, respectively, and band intensities were quantified by PhosphorImager analysis. The interference value, kappa (κ), at each position was calculated as described by McConnell et al. (38) and is the ratio of phosphorothioate incorporated over the phosphorothioate bound in the particle. According to Ryder et al. (61), κ values of >2 indicate interferences, and values of <0.5 indicate enhancements; however, because the experiments presented here utilize a novel experimental approach, these values are used only as a guide.

Purification of native snRNP TPs. Native snRNP TPs were prepared as described previously (64). HeLa U snRNPs were immunopurified using anti-2,2,7-trimethylguanosine (mouse) agarose conjugate (Calbiochem). Proteins were resolved on NuPAGE Novex Bis-Tris precast 4 to 12% mini-gels (Invitrogen) and analyzed by silver staining or Western blotting.

***Xenopus* oocyte microinjections.** Injections were carried out as described in Fischer et al. (9). Briefly, oocytes were harvested and incubated for 2 h in modified Barth's solution containing 0.2% collagenase type II (Sigma). Defolliculated stage V and VI oocytes were collected and used the next day for microinjection. In a typical injection experiment, 20 nl of ^{32}P-labeled RNAs (usually $\sim 1 \times 10^6$ cpm/ml for each RNA) were injected into the cytoplasm of oocytes. After a 1.5-h incubation, total extract was made and subjected to immunoprecipitation as described for snRNP assembly.

RESULTS

The SMN complex binds to specific and conserved domains of the HSURs. We have previously mapped the SMN complex-binding domains of U2, U4, and U5 snRNAs and found that the SMN complex binds to regions close to the 3′ ends of these RNAs, including the Sm site and the 3′ stem-loop (67). To investigate if this minimal binding domain is conserved among other Sm site-containing RNAs, we performed limited alkaline hydrolysis on the HSURs and assayed their binding to the SMN complex. The 5′- or 3′-end-labeled HSUR1 or HSUR3 was subjected to partial alkaline hydrolysis, and the resulting hydrolyzed RNA ladders were incubated with high-salt-purified SMN complexes. Bound RNAs were purified and analyzed by denaturing polyacrylamide gel electrophoresis (Fig. 1). This method allows a rough delineation of the regions of the RNAs that may be necessary or dispensable for binding to the SMN complex. As seen in Fig. 1B, the domain of HSUR1 that is required for SMN complex binding contained, at most, the region between nt 88 and the 3′ end (nt 143). Although it appears that some additional nucleotides at the 3′ end may be dispensable for binding, these nucleotides are more likely the result of nonspecific degradation that can also be seen in the input lane at higher exposures. Efficient binding of the SMN complex to HSUR3 required nucleotides from position 17 through the 3′ end, although an even smaller region, beginning with nt 31, may be sufficient for weaker binding (Fig. 1C). To confirm that the regions delineated by these mapping experiments are sufficient for SMN complex binding, the RNA fragments mapped in Fig. 1B and C were produced by transcription in vitro in the presence of [^{32}P]UTP and tested for direct binding to purified SMN complex or nonspecific HeLa control proteins. Figure 1D and E show that the minimal regions of HSUR1 and HSUR3, respectively, were sufficient for binding to the SMN complex, and none of the RNAs bound to the negative control proteins. Thus, similar to U2, U4, and U5 snRNAs, the regions of HSUR1 and HSUR3 necessary for SMN complex binding include both the Sm site and the 3′ terminal stem-loop.

To define the SMN complex-binding domains of HSUR4 and HSUR5, using the alkaline hydrolysis data as a guide, we produced a series of 5′ end deletion mutants. The RNA fragments were transcribed in the presence of [^{32}P]UTP and incubated with purified SMN complex or with nonspecific proteins as a control, and bound RNAs were purified. Among the deletion mutants of HSUR4, the region from nt 35 to the 3′ end (nt 106) bound efficiently to the SMN complex, while a smaller fragment, beginning with nt 51 bound more weakly (Fig. 1F). In the case of HSUR5, the region from nt 60 to the 3′ end (nt 114) efficiently bound to the SMN complex, but the fragment that begins with nt 70 is an even shorter domain that mediated weak binding (Fig. 1G). None of the RNAs showed any binding to the negative control (data not shown). Similar to HSURs 1 and 3, the minimal regions of HSURs 4 and 5 that are necessary for SMN complex binding also contain the Sm sites and the 3′ terminal stem-loops. For HSURs 1 to 4, the 3′ stem-loop alone was tested for binding but was not sufficient (Fig. 1D, E, F, and G). Overall, the minimal SMN complex-binding domains of HSURs 1 to 4, as indicated in pink in

FIG. 2. The minimal SMN complex-binding domains are sufficient for SMN-dependent Sm core assembly. (A) The secondary structures of HSUR1, HSUR3, HSUR4, and HSUR5 and their SMN complex-binding domains (highlighted in pink). Solid blue and red arrows designate the maximum 5′ and 3′ end boundaries of the SMN complex-binding domains, and open blue arrows indicate smaller domains that mediate weak binding to the SMN complex. The open red arrow indicates nonspecific degradation. (B) [^{32}P]UTP-labeled HSUR3, HSUR3-17, HSUR4, HSUR4-35, HSUR5, and HSUR5-60 were incubated with buffer (−), HeLa mock-depleted cytoplasmic extracts (CE), or SMN complex-depleted HeLa extracts (ΔSMN) for 1 h at 30°C. The assembly reaction products were analyzed by electrophoresis on 6% native polyacrylamide gels and autoradiography. Sm cores and free RNAs are each indicated by brackets.

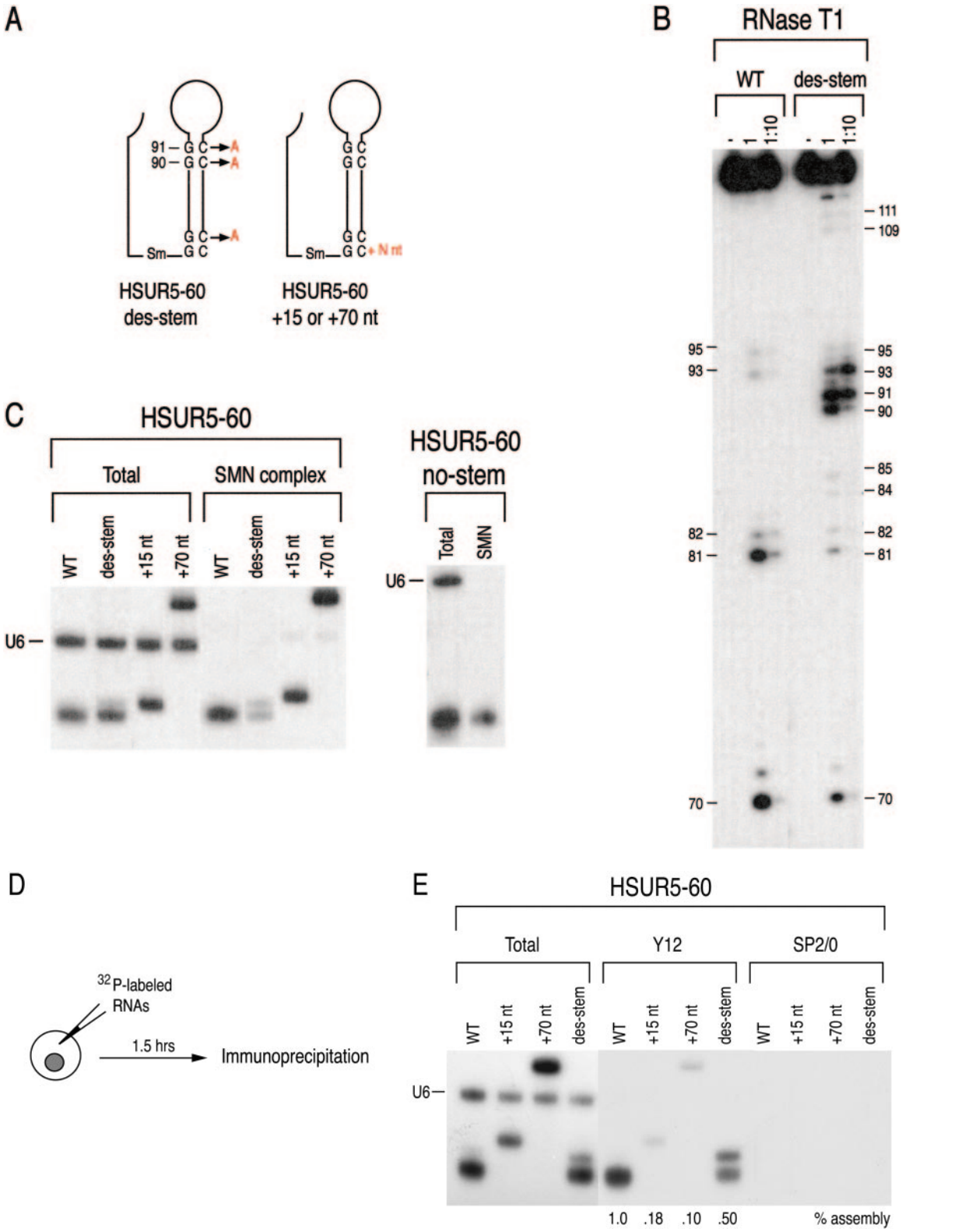


FIG. 3. A terminal stem-loop is necessary for SMN complex binding in vitro and Sm core assembly in vivo. (A) Illustration of HSUR5-60 RNAs used in this experiment. Site-directed mutagenesis was used to destabilize the terminal stem of HSUR5-60 (des-stem) and either 15 or 70 nt were added to the 3′ end of HSUR5-60 (+15 nt and +70 nt). (B) HSUR5-60 wild type (WT) or des-stem RNAs were labeled at the 5′ end with [γ-^{32}P]ATP and incubated with serial dilutions of RNase T1 for 15 min at 25°C. The digested RNAs were purified and analyzed by electrophoresis on 7 M urea–12% polyacrylamide gels and autoradiography. (C) HSUR5-60 (WT), des-stem, +15-nt, and +70-nt RNAs or an HSUR5-60 RNA with a complete deletion of the 3′ stem-loop (no-stem) were [^{32}P]UTP labeled, mixed with U6 as a negative control, and incubated with purified SMN complex for 1 h at 4°C. Bound RNAs were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiog-

Fig. 2A, resemble those for U2, U4, and U5 snRNAs and highlight the overall conservation of SMN complex-binding domains among Sm site-containing snRNAs.

The minimal SMN complex-binding domains are sufficient for SMN-dependent Sm core assembly. To investigate if the minimal SMN complex-binding domains are sufficient for Sm core assembly, we incubated [^{32}P]UTP-labeled HSUR3-17, HSUR4-35 and HSUR5-60 (Fig. 2A) in HeLa cytoplasmic extracts and analyzed the snRNP assembly reaction products by electrophoresis on native polyacrylamide gels. The full-length RNAs (HSUR3, HSUR4, and HSUR5) were also labeled and tested in parallel as a positive control. As seen in Fig. 2B, the minimal regions of all three HSURs assembled Sm cores in vitro (lanes 5, 11, and 17) that migrate at a size similar to that of the full-length RNA counterparts (lanes 2, 8, and 14) and supershift upon the addition of anti-Sm (Y12) monoclonal antibody (data not shown). The shorter minimal domains, whose 5′ ends are indicated by an open arrow (Fig. 2A), in addition to HSUR1-88, did not efficiently assemble Sm cores (data not shown). For HSUR5, the band that migrates slightly faster than the Sm core band most likely consists of HSUR5 complexed with the HuR protein, which binds to AUUUA repeats at the 5′ ends of HSURs 1, 2, and 5 (47) (Fig. 2B, lanes 14 and 15).

To examine the role of the SMN complex in the assembly of Sm cores on the minimal HSURs, cytoplasmic extracts were immunodepleted of the SMN complex prior to the assembly reaction. As previously demonstrated for full-length RNAs, the removal of the SMN complex resulted in loss of Sm core assembly on minimal RNAs (lanes 3, 6, 9, 12, 15, and 18), even though immunodepletion of SMN did not significantly reduce the amount of Sm proteins in the extracts (16, 56). The minimal domains of HSUR4 (HSUR4-35) and HSUR5 (HSUR5-60) assembled Sm cores in an SMN-dependent manner, although they appear to have reduced efficiency compared to the corresponding full-length RNAs (Fig. 2B, compare lanes 8 and 11 and 14 and 17). Overall, these experiments indicate that the minimal SMN complex-binding domains are sufficient for Sm core assembly.

A terminal 3′ stem-loop is required for SMN complex binding and Sm core assembly. Because the presence of a 3′ stem-loop is a common feature of the minimal SMN complex-binding domains of both the HSURs (Fig. 2A) and most of the U snRNAs (67), we wished to assess the importance of this structural feature for SMN complex binding and Sm core assembly. Toward this end, we carried out site-directed mutagenesis to destabilize the terminal stem in HSUR5-60 (Fig. 3A, des-stem). Limited RNase T1 digestion, which cleaves 3′ of guanosines in single-stranded RNAs, confirmed the linearization of the stem structure (Fig. 3B). In addition, to determine if the stem-loop must be at the 3′ end of the RNA, constructs were made that included either 15 or 70 additional nucleotides at the 3′ end of HSUR5-60 (Fig. 3A, +15 nt and +70 nt). The RNAs were then labeled with [^{32}P]UTP, mixed with labeled U6 snRNA as a negative control, and tested for binding to purified SMN complex or nonspecific immunopurified HeLa cell proteins as a negative control. As Fig. 3C shows, disruption of the secondary structure of the stem-loop strongly reduced SMN complex binding. In addition, a construct that completely removed the 3′ stem-loop (Fig. 3C, no-stem) also showed a reduction in SMN complex binding (Fig. 3C). However, the addition of nucleotides to the 3′ end of HSUR5-60 did not impair SMN complex binding (Fig. 3C), presumably because the SMN binding site itself (Fig. 2A) has not been altered. None of the RNAs tested bound to the negative control (data not shown).

To investigate if the structure and/or the position of the stem-loop affect the assembly of an Sm core in vivo, the des-stem RNA and the +15-nt and +70-nt RNAs were ^{32}P-labeled and microinjected into the cytoplasm of *Xenopus* oocytes as illustrated in Fig. 3D. After 1.5 h, immunoprecipitation of assembled Sm cores with Y12 antibody revealed that the des-stem RNA is 50% less efficient in Sm core assembly (Fig. 3E), consistent with its decreased binding to the SMN complex in vitro (Fig. 3C). However, despite the efficient binding of the +15 nt and +70 nt HSUR5-60 mutants to the SMN complex (Fig. 3C), these RNAs are also significantly reduced in Sm core assembly, by 80% and 90%, respectively (Fig. 3E). We further studied the assembly of the Sm cores on these RNAs in vitro with purified SMN complex or purified snRNP TPs (Fig. 4). In agreement with the findings that TPs can readily assemble Sm cores on Sm site-containing oligonucleotides alone (59), they efficiently assembled Sm cores on both the des-stem RNA and the +15-nt and +70-nt RNAs (Fig. 4A). To assess SMN-mediated Sm core assembly on the same RNAs, SMN complexes were purified at low-salt conditions so that endogenous Sm proteins remained associated with them (Fig. 4D). SMN-Sm protein complexes are necessary and sufficient for Sm core assembly (16, 56). The silver-stained gels demonstrate that purified snRNP proteins (TPs) and the SMN complexes used in these experiments contained similar amounts of Sm proteins (Fig. 4D and E). In contrast to TPs, assembly of Sm cores was drastically reduced on the des-stem RNA, +15-nt and +70-nt RNAs, and no-stem RNA, compared to the wild-type (Fig. 4B and C). The assembly of the des-stem RNA is decreased by 95% in vitro and 50% in vivo, while the assembly of the +15-nt and +70-nt RNAs is decreased by 50% and 75% in vitro and 82% and 90% in vivo (Fig. 3E and 4B). Thus, the overall level of assembly of all three RNAs is decreased in comparison to wild type both in vitro and in vivo. Therefore, unlike TPs, the SMN complex requires a stem-loop 3′ of an Sm site for binding, and it will only assemble an Sm core efficiently if the stem-loop is located at the 3′ end of the RNA.

raphy. Total represents 10% of input. (D) Experimental strategy used for *Xenopus* oocyte microinjection. (E) HSUR5-60 (WT), des-stem (the lower band), or +15 nt and +70 nt RNAs were [^{32}P]UTP labeled, mixed with labeled U6, and injected into the cytoplasm of oocytes as shown in panel D. After incubation for 1.5 h, the oocytes were homogenized, and immunoprecipitations were carried out with either anti-Sm (Y12) or control nonimmune (SP2/0) antibodies. RNAs were purified and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels. Total represents 10% of the fractions used for each immunoprecipitation. The levels of assembly were quantitated using imaging software and normalized to the wild type.

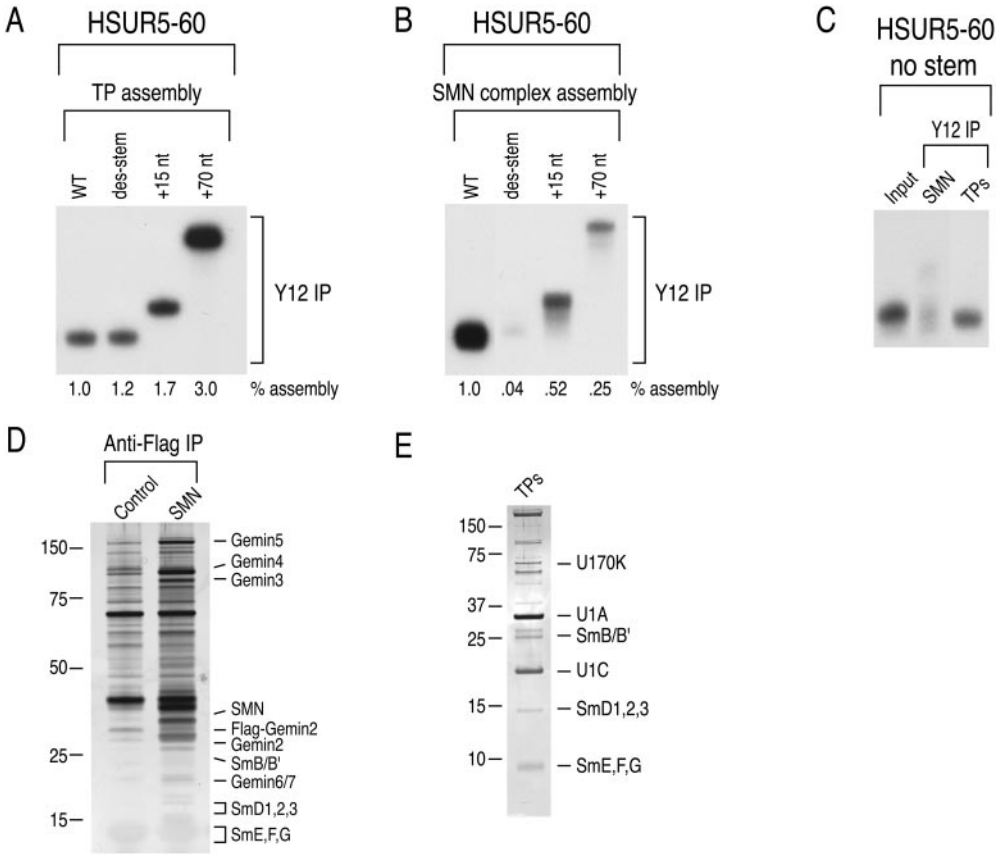


FIG. 4. A terminal stem-loop is required for SMN-dependent Sm core assembly. (A) A total of 10,000 cpm of HSUR5-60 (WT), des-stem, or +15-nt and +70-nt RNAs (Fig. 3) were incubated with HeLa snRNP TPs for 1 h at 30°C and then subjected to immunoprecipitation with anti-Sm (Y12) antibody. RNAs with assembled Sm cores were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. The levels of assembly were quantitated using imaging software and normalized to the wild type, and the input values were normalized prior to assembly. The composition of TPs is shown in panel E. (B) The RNAs used in panel A were incubated with low-salt-purified SMN complex for 1 h at 30°C and then subjected to immunoprecipitation with anti-Sm (Y12) antibody. RNAs with assembled Sm cores were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. The levels of assembly were quantitated using imaging software and normalized to the wild type, and the input values were normalized prior to assembly. The composition of SMN complex is shown in panel D. (C) HSUR5-60 no-stem (no 3′ stem-loop) RNA was incubated with low-salt-purified SMN complex (SMN) or TPs for 1 h at 30°C and then subjected to immunoprecipitation with anti-Sm (Y12) antibody. RNAs with assembled Sm cores were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. Input represents 10% of total. The experiments shown in panels B and C were performed in parallel. (D) Native SMN complexes (SMN) or nonspecific proteins (Control) were purified from flag-Gemin2 cells or the parental HeLa cells, respectively, under low-salt conditions as described in Materials and Methods. Flag-purified proteins were eluted with 3× Flag peptide, resolved by electrophoresis on 4 to 12% gradient polyacrylamide gels, and analyzed by silver staining. Under these conditions, all seven Sm proteins copurify with the SMN complex. (E) Native snRNP TPs were purified from HeLa cells as described in Materials and Methods and were analyzed by electrophoresis on 4 to 12% gradient polyacrylamide gels and by silver staining.

The sequence of the 3′ stem-loop is not critical for SMN complex binding and Sm core assembly. To determine if the specific sequence of the stem or the loop is important for SMN complex binding and SMN-mediated Sm core assembly, we constructed several mutants of HSUR4-35 and HSUR5-60 that have alterations in the sequences, but not the lengths, of the terminal stem-loop. To maintain the base-pairing properties of the stem and the size of the loop, only the directionality of the sequences was changed, so that the stem is either flipped upside-down (flip), the 5′ face of the stem is swapped with the 3′ face of the stem (swap), or the sequence of the loop is reversed (loop), as illustrated in Fig. 5A. These RNAs were then transcribed in vitro with [^{32}P]UTP, mixed with labeled U6 snRNA as a negative control, and incubated with purified SMN complexes or with the corresponding HeLa-immunopurified control fractions. Interestingly, most of the sequence alterations in the stem or the loop of HSUR4-35 and HSUR5-60 did not reduce the binding to the SMN complex (Fig. 5B). Only one construct, HSUR5-60 swap, had some reduction in SMN complex binding (Fig. 5B). None of the RNAs showed background binding to the negative control (data not shown).

To determine the effects of stem-loop mutations on Sm core assembly, the ^{32}P-labeled RNAs were microinjected into the cytoplasm of *Xenopus* oocytes, as described in Fig. 3D. Immunoprecipitation with Y12 antibody from total extract revealed that all of the stem-loop mutants of both HSUR4-35 and HSUR5-60 efficiently assembled Sm cores in vivo (Fig. 5C and D, respectively). As a majority of the sequence changes do not affect SMN complex binding or Sm core assembly, it seems unlikely that the SMN complex recognizes a specific sequence

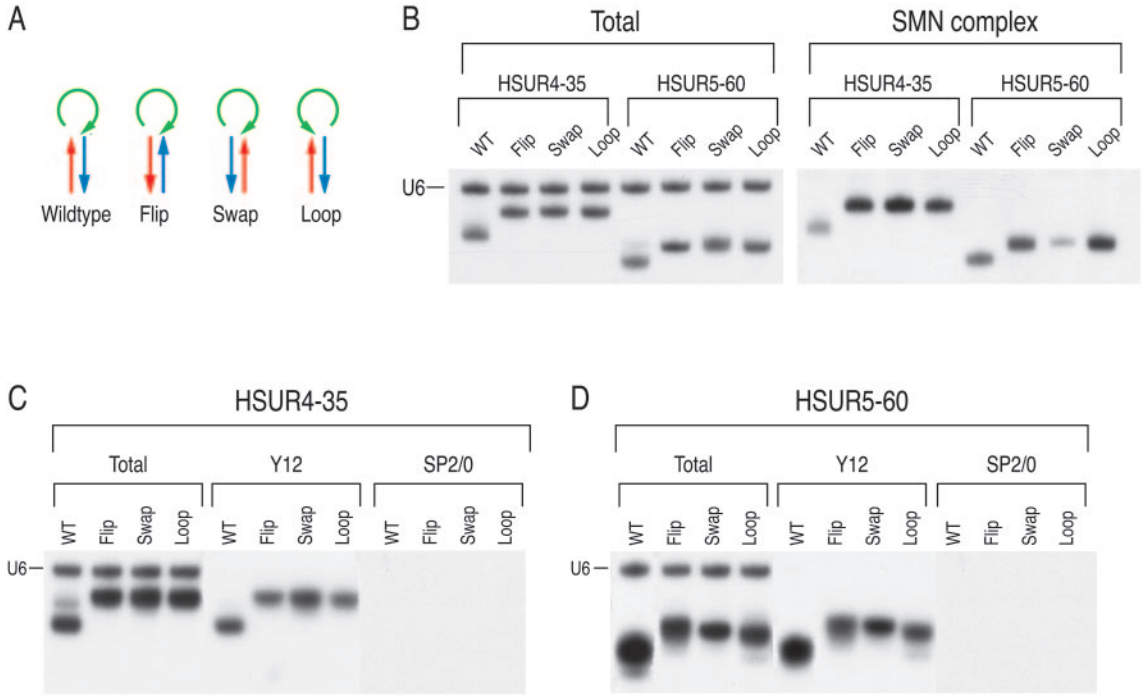


FIG. 5. The sequence of the 3′ stem-loop is not critical for SMN complex binding in vitro or Sm core assembly in vivo. (A) Illustration of the constructs used in this experiment. (B) The wild-type (WT), flip, swap, and loop constructs of HSUR4-35 and HSUR5-60, as illustrated in panel A, were labeled with [^{32}P]UTP, mixed with U6 as a negative control, and incubated with purified SMN complex for 1 h at 4°C. Bound RNAs were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. The wild type has a different electrophoretic mobility than the flip, swap, and loop constructs due to digestion with a different restriction enzyme prior to runoff transcription. Total represents 10% of input. (C) The wild-type, flip, swap, and loop constructs of HSUR4-35 were labeled with [^{32}P]UTP, mixed with labeled U6, and injected into the cytoplasm of *Xenopus* oocytes. After a 1.5-h incubation, the oocytes were homogenized, and immunoprecipitations were carried out with either anti-Sm (Y12) or control nonimmune (SP2/0) antibodies. RNAs were purified and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels. Total represents 10% of the fractions used for each immunoprecipitation. (D) The wild-type, flip, swap, and loop constructs of HSUR5-60 were subjected to the same experiment as described in panel C.

in either the stem or the loop of these RNAs. Rather, the stem-loop appears to be an important structural element required for SMN complex binding.

Specific nucleotides of the Sm site are essential for SMN complex binding and Sm core assembly. In addition to the 3′ stem-loop, the Sm site is a conserved feature among the minimal SMN complex-binding domains (Fig. 2A) (67). To determine if the conserved sequence of the Sm site is important for SMN complex binding, we produced mutations in the Sm site of HSUR5-60 (AUUUUUG) that changed uridines 1 through 5 to a cytosine one at a time, in addition to the adenosine and guanosine, which were changed to a cytosine and adenosine, respectively. The HSUR5-60 mutants were then labeled with [^{32}P]UTP, mixed with U6 as a negative control, and bound to purified SMN complexes or nonspecific control proteins. As seen in Fig. 6A, substitution of the first or third uridines of the Sm site significantly reduced SMN complex binding by 80% and 90%, respectively, and a change in the second, fourth, or fifth uridines decreased SMN complex binding by 38 to 55%. Substitution of the Sm site adenosine also drastically reduced SMN complex binding, whereas substitution of the guanosine had no detectable effect (Fig. 6A). None of the Sm site mutants bound to the negative control (Fig. 6A). To differentiate between Sm site mutations that interfere with Sm protein binding versus mutations that affect SMN-dependent Sm core assembly, the Sm site mutants were incubated with either TPs or low-salt-purified SMN complex (Fig. 6B and C). Interestingly, TPs readily assembled an Sm core when any one of the Sm site positions was changed but did not assemble on a mutant that contains a double substitution in both the first and third uridines of the Sm site (Fig. 6C and data not shown). In contrast, a change to the first and/or third uridines of the Sm site caused a dramatic reduction in SMN-mediated Sm core assembly (Fig. 6B). Interestingly, when the fourth uridine is changed to a cytosine, the Sm site becomes identical to that of a subset of U1 snRNAs. Similar to the reduction in SMN complex binding observed for this substitution (Fig. 6A), the SMN-mediated Sm core assembly on this RNA was also reduced by at least one-half (data not shown). While the overall sequence of the Sm site appears to be important, these data suggest that the adenosine and first and third uridines of the Sm site are critical for SMN complex binding and SMN-mediated Sm core assembly.

The SMN complex contacts the snRNA backbone at the first and third uridines of the Sm site. Because the sequence of the Sm site appears to be important for SMN complex binding and Sm core assembly, we asked if the SMN complex directly contacts this region of the snRNA. To identify potential interactions of the SMN complex with the RNA backbone, we carried out phosphorothioate interference mapping. Phosphorothio-

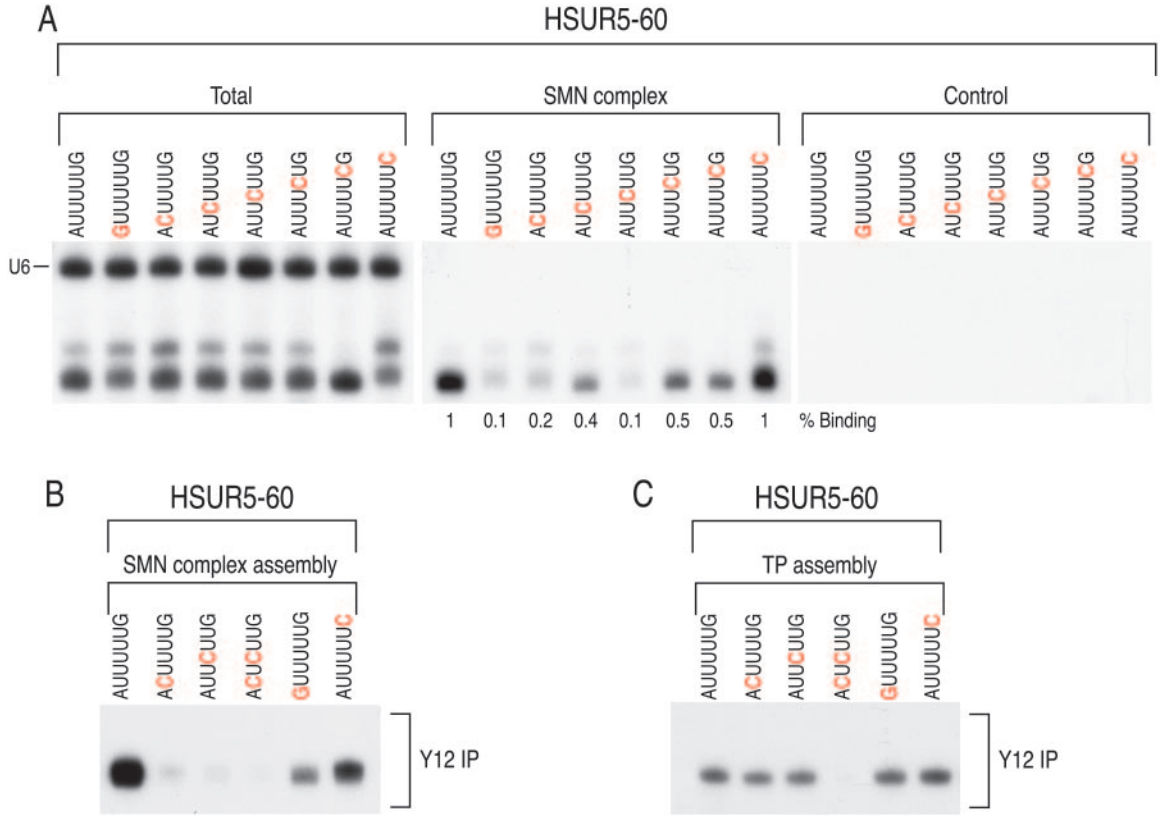


FIG. 6. The first and third uridines of the Sm site are essential for SMN complex binding (and Sm core assembly in vitro). (A) Wild-type HSUR5-60 (AUUUUUG) or HSUR5-60 in which each position of the Sm site was changed one at a time to either a guanosine or a cytosine (Sm site substitutions are indicated in red) was [^{32}P]UTP labeled, mixed with U6 as a negative control, and incubated with purified SMN complex (SMN complex) or control HeLa proteins (Control) for 1 h at 4°C. Bound RNAs were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and by autoradiography. Total represents 10% of input. The levels of binding were quantitated using Photoshop and normalized to the wild type. (B) A subset of the RNAs used in panel A were incubated with low-salt-purified SMN complex (Fig. 4D) for 1 h at 30°C and then subjected to immunoprecipitation with anti-Sm (Y12) antibody. RNAs with assembled Sm cores were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography. (C) The RNAs used in panel B were incubated with TPs (Fig. 4E) for 1 h at 30°C and then subjected to immunoprecipitation with anti-Sm (Y12) antibody. RNAs with assembled Sm cores were isolated and analyzed by electrophoresis on 7 M urea–8% polyacrylamide gels and autoradiography.

ate interference mapping utilizes the random incorporation of α-phosphorothioate-tagged nucleotide analogs into the phosphate backbone of the RNA molecule. After exposure to a functional assay, the tagged RNAs are recovered, and the phosphorothioate linkages are cleaved upon addition of iodine, producing a ladder on a sequencing gel (61). At a position where the substitution interferes with protein binding, a band of lower intensity will be present on the gel that is quantitatively represented as an increased kappa (κ, or interference) value. Likewise, a reduction in the κ value indicates a position of enhancement (61). To identify positions in the HSURs where phosphorothioate substitution affects SMN complex binding, HSUR5-60 and HSUR4-35 were transcribed in the presence of the adenosine (A), uridine (U), guanosine (G), or cytidine (C) α-thiotriphosphates, labeled at the 5′ end with [γ-^{32}P]ATP and incubated with high-salt-washed immunopurified SMN complexes or nonspecifically immunopurified proteins as a control. The purified SMN complex appears to be free of Sm proteins, as their association is undetectable by both silver stain and Western blotting (Fig. 1A; see also Fig. 8D) (52, 56, 67, 68). Bound RNAs were then purified, cleaved with iodine, and analyzed by denaturing polyacrylamide gel electrophoresis.

Figure 7A and B show the sequencing ladders for HSUR5-60 and HSUR4-35, respectively. Although the full-length HSUR5-60 and HSUR4-35 RNAs were used, only the regions contained within nucleotides 70 to 114 and 52 to 92, respectively, were analyzed for interference values. As a control for phosphorothioate incorporation and cleavage efficiency, an equivalent amount of the total RNAs was treated with iodine and loaded in parallel (Fig. 7A and B). In addition, total and SMN complex-bound RNAs not treated with iodine were analyzed for nonspecific cleavage, which was undetectable (data not shown). For both HSUR5-60 and HSUR4-35, strong interferences occurred at the first and third uridine positions of the Sm site (Fig. 7A [U76 and U78] and B[U71 and U73]), in addition to U89 in the terminal loop and C65 in the 5′ stem-loop of HSUR4-35, respectively (Fig. 7B). The graphs shown in Fig. 7C and D display the numerical κ values for the uridine positions of the two RNAs, which are increased for the first and third Sm site uridines. The kappa values that we consider significant range from 1.4 to 2.2; rather than the absolute values of the

interferences, it is the relative differences in the values among the nucleotide positions that suggest an interaction with the SMN complex. Even though the Sm site adenosine also appears to be important for SMN complex binding (Fig. 6C), no specific interference was observed at this position or at any other A, C, or G position within the regions analyzed that has not been specifically mentioned (Fig. 7A and B and data not shown). Contacts with nucleotides in the 5′ stem (C65) and the 3′ loop (U89) of HSUR4-35 are also seen, but these contacts are not conserved in HSUR5-60, suggesting that they are not likely to be critical, especially specific nucleotides within the 3′ stem-loop (Fig. 5). Because the HSURs share a common binding site on the SMN complex with U4 snRNA (16), a similar analysis of the minimal SMN complex-binding domain of U4 snRNA (67) was performed and also revealed an increase in the κ values for the first and third Sm site uridines (data not shown). It is possible that the SMN complex also makes contacts at positions outside of the analyzed region or in a different manner that does not result in interference with phosphate backbone interactions. However, these results clearly indicate that SMN complex recognition of snRNAs does involve direct contact of the phosphate backbone of specific Sm site uridines.

Sm proteins alone do not contact the phosphate backbone of Sm site uridines. To rule out interference that might be caused by Sm proteins, if any, rather than SMN complex, purified snRNP TPs that contain all seven Sm proteins in addition to U1 snRNP-specific proteins (Fig. 4E) were incubated with phosphorothioate-tagged HSUR5-60. RNAs that acquired an Sm core were immunoprecipitated with Y12, cleaved with iodine, and analyzed by denaturing polyacrylamide gel electrophoresis. Figure 8A shows the resulting sequencing ladder for the region of HSUR5-60 that includes nt 70 through the 3′ end. Interestingly, the calculated interference (κ) values revealed that the TPs do not contact the phosphate backbone at any position within the Sm site (Fig. 8B). Rather, interferences were numerically determined to occur only at positions outside of the Sm site, including U92, C86, and G109 within the region analyzed (Fig. 8C and data not shown). Importantly, the pattern of interference revealed by TPs is completely different from that of the SMN complex (compare Fig. 7 and 8), and the TPs used in this experiment do not contain any obvious traces of SMN complex components (Fig. 8D). This experiment demonstrates that direct contact of the Sm site is due to SMN complex, rather than Sm protein, interactions with the snRNA.

DISCUSSION

Previous experiments indicated that the process of snRNP assembly, particularly the formation of the highly stable Sm core, is not a self-assembly process as had been widely believed, but, rather, is an active process mediated by the SMN complex (3, 9, 29, 30, 40, 42, 43, 55, 56, 67–69). Perhaps the key reason for cells to employ the SMN complex for snRNP assembly is to prevent the potentially promiscuous Sm proteins from forming Sm cores indiscriminately and to ensure that Sm cores only assemble on the correct RNAs (56, 67). Subsequent experiments delineated the general domains in the major snRNAs (U1, U2, U4, and U5) that contain the specific binding sites for the SMN complex (67, 68). However, other than the short Sm site, there is no extensive sequence similarity among the SMN complex-binding domains of these snRNAs; therefore, how the SMN complex might distinguish them from other cellular RNAs remained unclear. The HSURs, the snRNAs encoded by HVS (46), provided an attractive system to address this question because they use the SMN complex to assemble Sm cores, bind the SMN complex with high affinity, have considerable sequence conservation among them, and bear a striking resemblance to the overall structure of several of the major snRNAs (16, 27, 28). By systematic mutagenesis of the HSURs and using several experimental approaches, we determined the critical RNA sequence features that confer binding to the SMN complex and assembly of an Sm core. These experiments revealed a surprisingly simple structural configuration, comprised of an Sm site (AUUUUUG) and a 3′ terminal stem-loop that is critical for Sm core assembly in vitro and in vivo. These structural features, illustrated in Fig. 9, are remarkably independent of RNA sequence. This motif constitutes a snRNP code that is recognized by direct binding to the SMN complex and triggers formation of the Sm core on the Sm site.

Several lines of evidence indicate that it is the SMN complex itself, and not the Sm proteins on their own, that is responsible for deciphering this snRNP code. First, the purified SMN complex (high-salt washed) used in the RNA-binding experiments contained no detectable Sm proteins. This is assessed both by silver staining and immunoblotting (Fig. 1A and 8D) and by RNP gels that showed no detectable Sm core formation from these complexes (data not shown). Second, the profile of interactions of the SMN complex with the RNAs, assayed by phosphorothioate interference mapping, is significantly different from that of Sm proteins alone. The SMN complex contacts the backbone phosphates of the first and third uridines (Fig. 7), while TPs alone contact only the backbone of the 3′ stem-loop but not Sm site uridines (Fig. 8). These interference data are consistent with the previous report that TPs interact with the backbone of the 3′ stem-loop of U1 snRNA but not at nucleotide positions within the Sm site (38). Therefore, it is unlikely that the data we present here represent Sm protein, rather than SMN complex-dependent, interactions with snRNAs. UV cross-linking experiments have shown that Sm proteins, particularly, SmG and SmB/B′, also contact the first and third uridines in vitro (65). Although we cannot entirely rule out the possibility that the high-salt-purified SMN complex contains trace amounts of Sm proteins and that, therefore, they might play a role in the recognition of the snRNAs, it is clear that if the Sm proteins do play a role in the binding to snRNAs, their properties must be changed and controlled by the SMN complex. Furthermore, the assembly activity of purified TPs is much more promiscuous than that of purified SMN complex (Fig. 4 and 6), despite comparable concentrations of Sm proteins in each preparation (Fig. 4D and E). In vivo, Sm proteins are not free but are associated with a number of protein complexes, including the 6S pICln-containing complex and the 20S methylosome (14, 15, 41, 58). Because only Sm proteins that are carried by the SMN complex are competent for Sm core assembly (56), it makes sense that the SMN complex might recognize the very same site upon which Sm proteins assemble. This stringency would explain how the SMN complex ensures that Sm cores assemble only on the correct RNA targets (56). At this time, it is not known which protein component(s) of the SMN complex are responsible for binding RNA.

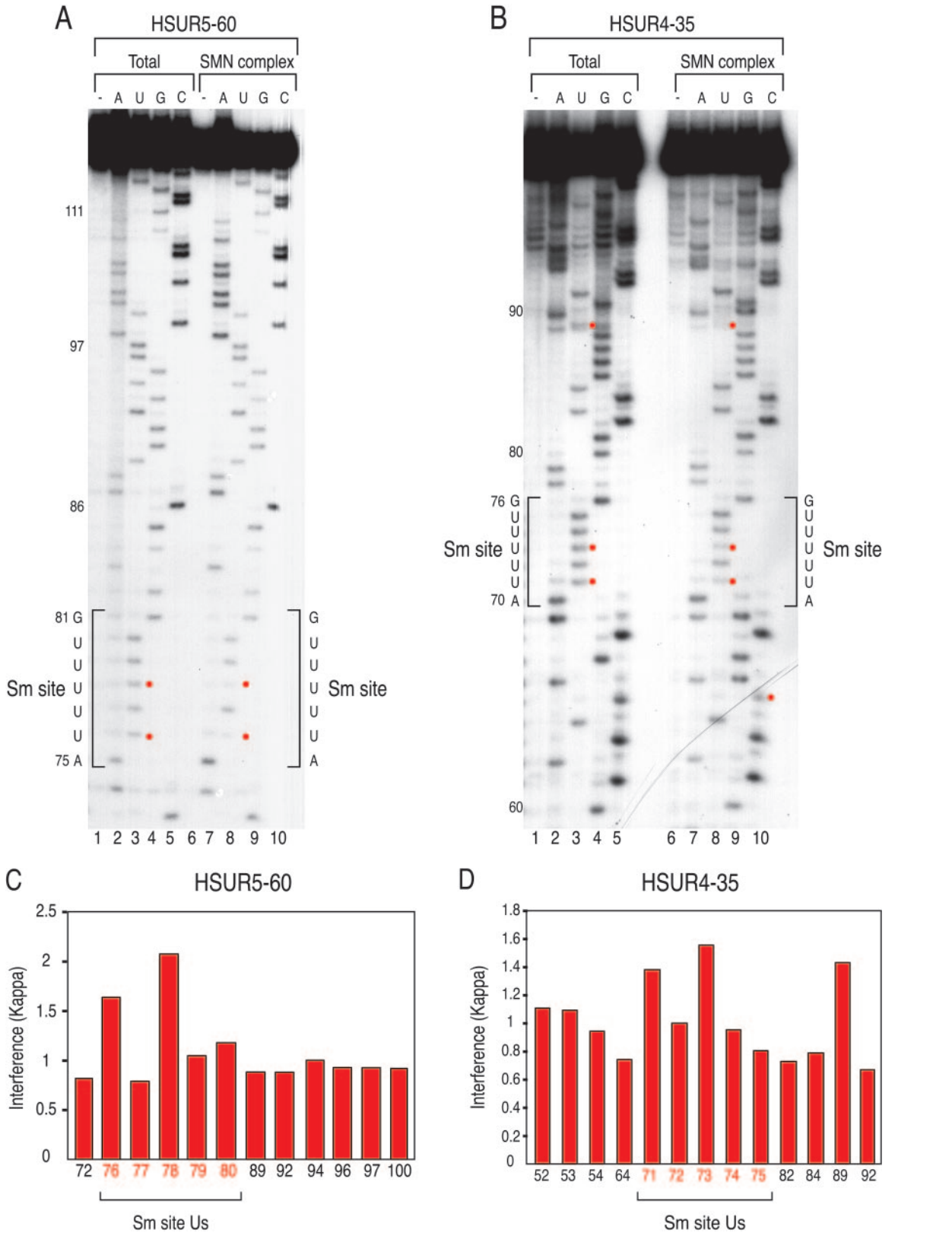


FIG. 7. The SMN complex contacts the phosphate backbone of the first and third uridines of the Sm site.(A) Phosphorothioate interference of HSUR5-60. HSUR5-60 transcribed in the presence of A, U, G, or C α-thiotriphosphate or no analog (−) was 5′ end labeled with [γ-^{32}P]ATP and incubated with purified SMN complex for 1 h at 4°C. Bound RNAs were isolated and cleaved with iodine, and the resulting fragments were resolved on a 7 M urea–12% polyacrylamide sequencing gel. Unbound (Total) RNA was also cleaved with iodine to correct for phosphorothioate incorporation and cleavage efficiency. The Sm site is indicated by brackets. Solid red circles to the right of U76 and U78 indicate positions where

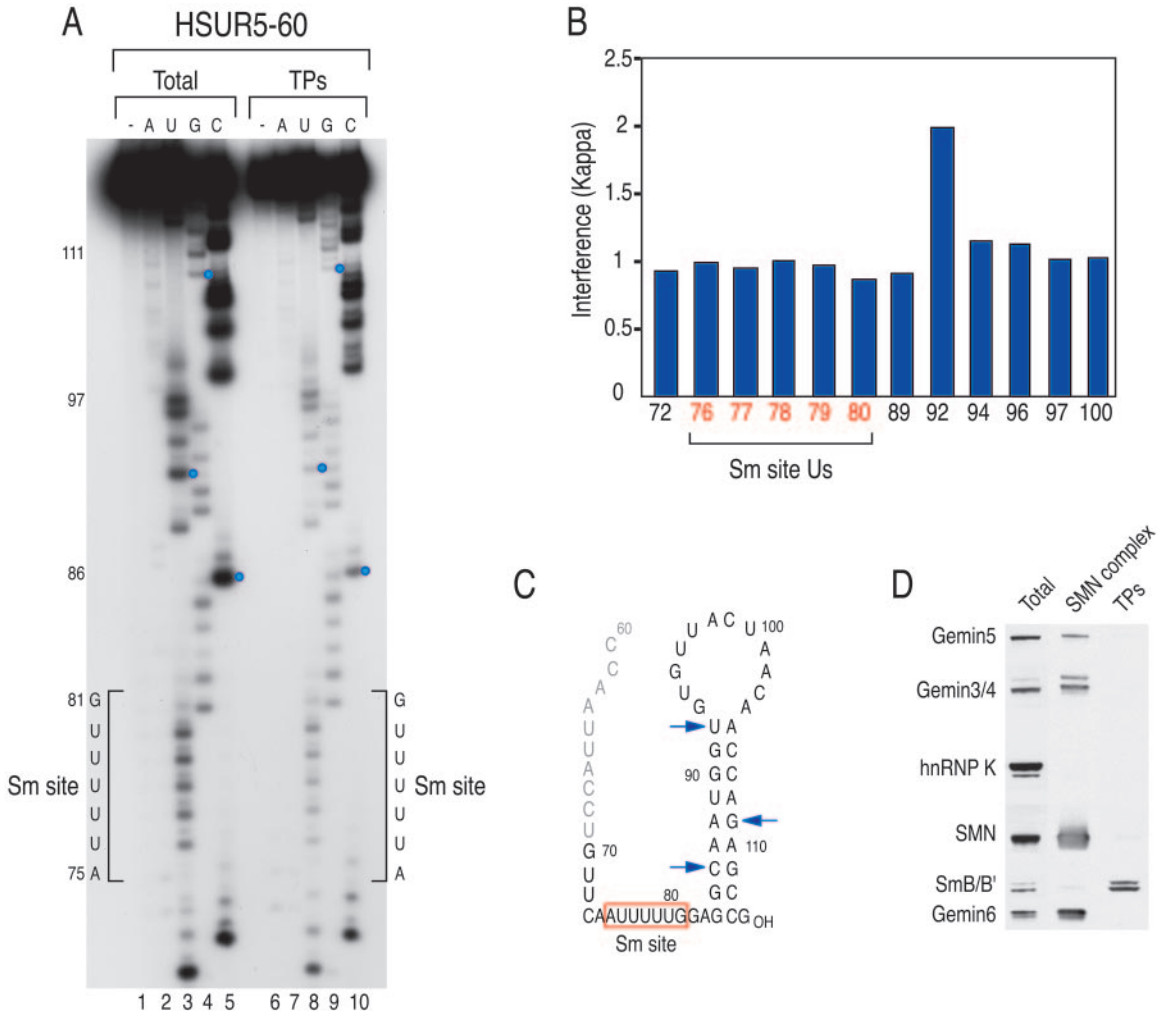


FIG. 8. Sm proteins do not cause phosphorothioate interference at Sm site uridines. (A) HSUR5-60 transcribed in the presence A, U, G, or C α-thiotriphosphate or no analog (−) was 5′ end labeled with [γ-^{32}P]ATP and incubated with purified TPs (Fig. 4E) for 1 h at 30°C. Bound RNAs were isolated by immunoprecipitation with anti-Sm (Y12) antibody and cleaved with iodine, and the resulting fragments were resolved on a 7 M urea–12% polyacrylamide sequencing gel. Unbound (Total) RNA was also cleaved with iodine to correct for phosphorothioate incorporation and cleavage efficiency. The Sm site is indicated by brackets. Solid blue circles to the right of the specified bands indicate positions where the phosphorothioate substitution interferes with TP binding and assembly. The adenosines (A) can be seen with increased exposure. (B) Kappa values for phosphorothioate interference at HSUR5-60 uridines. Sm site uridines are numbered in red. An increased κ value is interpreted as an interference. (C) Secondary structure of the region of HSUR5-60 used for phosphorothioate interference. Blue arrows indicate positions where phosphorothioate substitution interferes with TP binding and assembly. (D) HeLa cell extracts (Total), high-salt-purified SMN complex (SMN complex), and snRNP total proteins (TPs) were run on 4 to 12% gradient polyacrylamide gels and Western blotted for components of the SMN complex.

The strong resemblance of the minimal SMN complex-binding domains of the HSURs to the domains in the major Sm site-containing U snRNAs, U2, U4, and U5 (previously shown to contain the SMN-binding sites [67]), argues that the features we define here are of general significance for snRNAs. The clear exception is U1 snRNA, which contains a distinctly different motif for binding to the SMN complex. In U1, the high-affinity binding site for the SMN complex is in stem-loop 1 and does not require the Sm site (67–69). Furthermore, the Sm site sequence of U1, AAUUU(C/G)UGG, is different from and not interchangeable with that of U5 snRNA (23). Consequently, U1 snRNA binds to the SMN complex through a

the phosphorothioate substitution interferes with SMN complex binding. (B) Phosphorothioate interference of HSUR4-35. The same experiment as described in panel A was performed using 5′-end-labeled HSUR4-35. The cleaved RNAs were resolved on a 7 M urea–10% polyacrylamide sequencing gel. Solid red circles to the right of U71, U73, U89, and C65 indicate positions where the phosphorothioate substitution interferes with SMN complex binding. (C) Kappa values for phosphorothioate interference at HSUR5-60 uridines. Sm site uridines are numbered in red. An increased κ value is interpreted as an interference. (D) Kappa values for phosphorothioate interference at HSUR4-35 uridines. Sm site uridines are numbered in red. An increased κ value is interpreted as an interference.

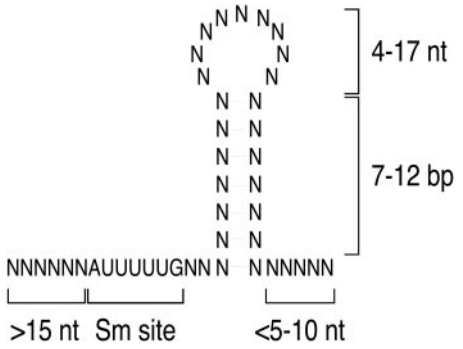


FIG. 9. Model for the critical RNA sequence features that confer binding to the SMN complex and assembly of an Sm core. The details of these features are described in the text.

sequence at its 5′, rather than 3′, end and at an independent binding site on the SMN complex, while the other U snRNAs and the HSURs share a second binding site (16, 67). Indeed, competition experiments suggest that there are at least two snRNA binding sites on the SMN complex, one for U1 snRNA and another for all the other snRNAs and the HSURs. Unlike Sm proteins, which promiscuously assemble an Sm core on single-stranded uridine-rich sequences in vitro (59), SMN complex recognition of an snRNA requires more. Consistent with previous assembly experiments that suggested that the Sm site might cooperate specifically with other elements of snRNAs for snRNP assembly (23), the data presented here suggest that the SMN complex recognizes the Sm site of an snRNA when it is presented within the context of a 3′ stem-loop. Upon destabilization or removal of the stem-loop, SMN complex binding and SMN-mediated Sm core assembly are dramatically reduced (Fig. 3 and 4). However, in contrast to the critical uridines of the Sm site, the sequence of the stem-loop seems to be relatively unimportant (Fig. 5). This makes sense considering that the stem-loop sequences of the U snRNAs and HSURs do not appear to have extensive sequence conservation both within and between the two groups. Furthermore, five variants of U5 snRNA exist within cells that have various changes throughout their stem-loop sequences that do not affect their assembly into snRNPs (62). Via the assembly and selection of Sm cores on randomized RNAs in *Xenopus* oocytes, Grimm et al. (17) identified the motif AAUUUUUGG, located near the 3′ stem of the carrier RNA, as the predominant sequence that confers Sm core assembly. As we now know that it is the SMN complex that both selects RNAs and mediates the assembly of Sm cores on them, these findings support the conclusion that sequences and structures unrelated to those defined here can function efficiently in SMN binding. Figure 9 summarizes the ranges of lengths of the 3′ end stem and loop that exist among all U snRNAs and HSURs. Less is known about the requirement for sequences located 5′ of the Sm site. Mapping data from both U snRNAs (67) and HSURs (this study) suggest that at least ~15 nt are required 5′ of the Sm site. Although there is no sequence or length conservation among the minimal 5′ end regions, we cannot rule out that a specific sequence or that individual nucleotides specific to each RNA may be important for SMN complex binding. Although the SMN complex appears to contact the phosphate backbone of a cytosine in the stem 5′ of the Sm site and a uridine in the 3′ loop of HSUR4-35 (Fig. 7), these nucleotides are not conserved in HSUR5-60 and likely do not play the same role in the recognition by the SMN complex as do Sm site uridines. It is also possible that the 5′ end may be required to maintain the overall secondary structure of the snRNA, to present a single-stranded Sm site, or to stabilize the Sm protein ring. In addition, there appears to be an upper limit on the length of the sequence just 3′ of the SMN complex-binding element. Masuyama et al. (34) have recently shown that the length of the RNA determines the export pathway it utilizes. The addition of ~300 nt to U1 snRNA shunted it to the mRNA export pathway, possibly influencing its activity and fate in the cytoplasm.

Binding to the SMN complex, while necessary, is not sufficient for Sm core assembly. The SMN complex binds to an RNA that has the right configuration of an Sm site sequence and an adjacent 3′ stem-loop. This conclusion is further supported by experiments with a U1 snRNA construct (U1 Swap) that is identical to wild-type U1, except that the positions of stem-loops 1 and 4 were swapped, so that stem-loop 1 is placed 3′ to the Sm site (U1 Swap) (67). While U1 Swap contains all the sequence elements required for SMN complex binding and Sm core assembly, including a high-affinity SMN complex binding site and an Sm site, the SMN complex discerns that the position of the Sm site relative to the 3′ end of the RNA has been altered and, subsequently, will bind to but not assemble an Sm core on the RNA (67). It is currently not known if the distance between the Sm site and the 3′ stem-loop is an important prerequisite in the assembly of an Sm core. Among all of the U snRNAs and HSURs, however, this length is maintained somewhere between 1 and 5 nt. In addition, if the RNA contains more than about 15 single-stranded nucleotides at the 3′ end, the SMN complex will not assemble an Sm core on it, in marked contrast to the behavior of TPs (Fig. 3 and 4). Thus, binding and assembly are two distinct steps within the snRNP biogenesis pathway, and the SMN complex has a built-in surveillance capacity that aborts the assembly reaction if the RNA does not have all the correct features. It is not clear whether the aborted RNAs accumulate in the cytoplasm, are degraded, or are shortened via an exonucleolytic mechanism to become substrates for Sm core assembly. At least after a 1-h incubation, the elongated RNAs appear to remain full-length (Fig. 3). Other studies have reported that U1 snRNA molecules with heterogenous 3′ ends do not efficiently assemble Sm cores and become poor substrates for normal 3′-end processing in *Xenopus* oocytes (50). In vivo, U snRNAs are transcribed as precursors that have been reported to contain as many as 6 to 8 nt more than the mature molecule (33), and Sm core assembly seems to be required for the 3′ end processing (71).

The discovery that the SMN complex decodes the critical sequence features of snRNAs through direct binding offers a mechanism to explain how Sm cores are only assembled on the correct RNAs. These findings suggest a structural algorithm that can be used to predict additional snRNAs, if such exist, and possibly motor neuron-specific RNA substrates for the SMN complex. Viewed in the more general context of how cells distinguish among the different classes of RNAs, our studies reveal a clear signature in the major snRNAs and demonstrate that the SMN complex performs the task of iden-

tifying that signature as well as carrying out the assembly of these RNAs into the corresponding RNPs.

ACKNOWLEDGMENTS

We are grateful to Joan A. Steitz for providing plasmids. We thank the members of our laboratory, especially Amelie Gubitz, Dan Battle, and Lili Wan, for helpful discussions and comments on the manuscript. We are also grateful to Gina Daly for secretarial assistance.

This work was supported by the Association Française Contre les Myopathies (AFM) and by a grant from the National Institutes of Health. T.J.G. is a Predoctoral Fellow of the Howard Hughes Medical Institute. G.D. is an Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. **Baccon, J., L. Pellizzoni, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2002. Identification and characterization of Gemin7, a novel component of the survival of motor neuron complex. J. Biol. Chem. **277:**31957–31962.

1a. **Brahms, H., Lo. Meheus, V. de Brabandere, U. Fischer, and R. Luhrmann.** 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA **7:**1531–1542.

2. **Branlant, C., A. Krol, J. P. Ebel, E. Lazar, B. Haendler, and M. Jacob.** 1982. U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J. **1:**1259–1265.

3. **Buhler, D., V. Raker, R. Luhrmann, and U. Fischer.** 1999. Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. Hum. Mol. Genet. **8:**2351–2357.

4. **Charroux, B., L. Pellizzoni, R. A. Perkinson, A. Shevchenko, M. Mann, and G. Dreyfuss.** 1999. Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. J. Cell Biol. **147:**1181–1194.

5. **Charroux, B., L. Pellizzoni, R. A. Perkinson, J. Yong, A. Shevchenko, M. Mann, and G. Dreyfuss.** 2000. Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli. J. Cell Biol. **148:**1177–1186.

6. **Cifuentes-Diaz, C., T. Frugier, and J. Melki.** 2002. Spinal muscular atrophy. Semin. Pediatr. Neurol. **9:**145–150.

7. **Crawford, T. O., and C. A. Pardo.** 1996. The neurobiology of childhood spinal muscular atrophy. Neurobiol. Dis. **3:**97–110.

8. **Fickenscher, H., and B. Fleckenstein.** 2001. Herpesvirus saimiri. Philos. Trans. R Soc. Lond. B **356:**545–567.

9. **Fischer, U., Q. Liu, and G. Dreyfuss.** 1997. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell **90:**1023–1029.

10. **Fischer, U., and R. Luhrmann.** 1990. An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. Science **249:**786–790.

11. **Fischer, U., V. Sumpter, M. Sekine, T. Satoh, and R. Luhrmann.** 1993. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. EMBO J. **12:**573–583.

12. **Friesen, W. J., and G. Dreyfuss.** 2000. Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). J. Biol. Chem. **275:**26370–26375.

13. **Friesen, W. J., S. Massenet, S. Paushkin, A. Wyce, and G. Dreyfuss.** 2001. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol. Cell **7:**1111–1117.

14. **Friesen, W. J., S. Paushkin, A. Wyce, S. Massenet, G. S. Pesiridis, G. Van Duyne, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2001. The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol. Cell. Biol. **21:**8289–8300.

15. **Friesen, W. J., A. Wyce, S. Paushkin, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2002. A novel WD repeat protein component of the methylosome binds Sm proteins. J. Biol. Chem. **277:**8243–8247.

16. **Golembe, T. J., J. Yong, D. J. Battle, W. Feng, L. Wan, and G. Dreyfuss.** 2005. Lymphotropic *Herpesvirus saimiri* uses the SMN complex to assemble Sm cores on its small RNAs. Mol. Cell. Biol. **25:**602–611.

17. **Grimm, C., E. Lund, and J. E. Dahlberg.** 1997. In vivo selection of RNAs that localize in the nucleus. EMBO J. **16:**793–806.

18. **Gubitz, A. K., W. Feng, and G. Dreyfuss.** 2004. The SMN complex. Exp. Cell Res. **296:**51–56.

19. **Gubitz, A. K., Z. Mourelatos, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2002. Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. J. Biol. Chem. **277:**5631–5636.

20. **Hamm, J., N. A. Dathan, D. Scherly, and I. W. Mattaj.** 1990. Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. EMBO J. **9:**1237–1244.

21. **Iannaccone, S. T., S. A. Smith, and L. R. Simard.** 2004. Spinal muscular atrophy. Curr. Neurol. Neurosci. Rep. **4:**74–80.

22. **Imai, Y., Y. Matsushima, T. Sugimura, and M. Terada.** 1991. A simple and rapid method for generating a deletion by PCR. Nucleic Acids Res. **19:**2785.

23. **Jarmolowski, A., and I. W. Mattaj.** 1993. The determinants for Sm protein binding to *Xenopus* U1 and U5 snRNAs are complex and non-identical. EMBO J. **12:**223–232.

24. **Jones, K. W., K. Gorzynski, C. M. Hales, U. Fischer, F. Badbanchi, R. M. Terns, and M. P. Terns.** 2001. Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. J. Biol. Chem. **276:**38645–38651.

25. **Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li, and K. Nagai.** 1999. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell **96:**375–387.

26. **Kleinschmidt, A. M., J. R. Patton, and T. Pederson.** 1989. U2 small nuclear RNP assembly in vitro. Nucleic Acids Res. **17:**4817–4828.

27. **Lee, S. I., S. C. Murthy, J. J. Trimble, R. C. Desrosiers, and J. A. Steitz.** 1988. Four novel U RNAs are encoded by a herpesvirus. Cell **54:**599–607.

28. **Lee, S. I., and J. A. Steitz.** 1990. *Herpesvirus saimiri* U RNAs are expressed and assembled into ribonucleoprotein particles in the absence of other viral genes. J. Virol. **64:**3905–3915.

29. **Liu, Q., and G. Dreyfuss.** 1996. A novel nuclear structure containing the survival of motor neurons protein. EMBO J. **15:**3555–3565.

30. **Liu, Q., U. Fischer, F. Wang, and G. Dreyfuss.** 1997. The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell **90:**1013–1021.

31. **Luhrmann, R.** 1990. Functions of U-snRNPs. Mol. Biol. Rep. **14:**183–192.

32. **Luhrmann, R., B. Kastner, and M. Bach.** 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim. Biophys. Acta **1087:**265–292.

33. **Madore, S. J., E. D. Wieben, and T. Pederson.** 1984. Intracellular site of U1 small nuclear RNA processing and ribonucleoprotein assembly. J. Cell Biol. **98:**188–192.

34. **Masuyama, K., I. Taniguchi, N. Kataoka, and M. Ohno.** 2004. RNA length defines RNA export pathway. Genes Dev. **18:**2074–2085.

35. **Mattaj, I. W.** 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell **46:**905–911.

36. **Mattaj, I. W., W. Boelens, E. Izaurralde, A. Jarmolowski, and C. Kambach.** 1993. Nucleocytoplasmic transport and snRNP assembly. Mol. Biol. Rep. **18:**79–83.

37. **Mattaj, I. W., and E. M. De Robertis.** 1985. Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell **40:**111–118.

38. **McConnell, T. S., R. P. Lokken, and J. A. Steitz.** 2003. Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. RNA **9:**193–201.

39. **Meister, G., D. Buhler, B. Laggerbauer, M. Zobawa, F. Lottspeich, and U. Fischer.** 2000. Characterization of a nuclear 20S complex containing the survival of motor neurons (SMN) protein and a specific subset of spliceosomal Sm proteins. Hum. Mol. Genet. **9:**1977–1986.

40. **Meister, G., D. Buhler, R. Pillai, F. Lottspeich, and U. Fischer.** 2001. A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat. Cell. Biol. **3:**945–949.

41. **Meister, G., C. Eggert, D. Buhler, H. Brahms, C. Kambach, and U. Fischer.** 2001. Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr. Biol. **11:**1990–1994.

42. **Meister, G., C. Eggert, and U. Fischer.** 2002. SMN-mediated assembly of RNPs: a complex story. Trends Cell Biol. **12:**472–478.

43. **Meister, G., and U. Fischer.** 2002. Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. EMBO J. **21:**5853–5863.

44. **Mourelatos, Z., L. Abel, J. Yong, N. Kataoka, and G. Dreyfuss.** 2001. SMN interacts with a novel family of hnRNP and spliceosomal proteins. EMBO J. **20:**5443–5452.

45. **Mourelatos, Z., J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. **16:**720–728.

46. **Murthy, S., J. Kamine, and R. C. Desrosiers.** 1986. Viral-encoded small RNAs in herpes virus saimiri induced tumors. EMBO J. **5:**1625–1632.

47. **Myer, V. E., S. I. Lee, and J. A. Steitz.** 1992. Viral small nuclear ribonucleoproteins bind a protein implicated in messenger RNA destabilization. Proc. Natl. Acad. Sci. USA **89:**1296–1300.

48. **Nagai, K., Y. Muto, D. A. Pomeranz Krummel, C. Kambach, T. Ignjatovic, S. Walke, and A. Kuglstatter.** 2001. Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture. Biochem. Soc. Trans. **29:**15–26.

49. **Narayanan, U., T. Achsel, R. Luhrmann, and A. G. Matera.** 2004. Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. Mol. Cell **16:**223–234.

50. **Neuman de Vegvar, H. E., and J. E. Dahlberg.** 1990. Nucleocytoplasmic transport and processing of small nuclear RNA precursors. Mol. Cell. Biol. **10:**3365–3375.

51. **Pellizzoni, L., J. Baccon, B. Charroux, and G. Dreyfuss.** 2001. The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Curr. Biol. **11:**1079–1088.

52. **Pellizzoni, L., J. Baccon, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2002.

Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. J. Biol. Chem. **277:**7540–7545.
53. **Pellizzoni, L., B. Charroux, and G. Dreyfuss.** 1999. SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. Proc. Natl. Acad. Sci. USA **96:**11167–11172.
54. **Pellizzoni, L., B. Charroux, J. Rappsilber, M. Mann, and G. Dreyfuss.** 2001. A functional interaction between the survival motor neuron complex and RNA polymerase II. J. Cell Biol. **152:**75–85.
55. **Pellizzoni, L., N. Kataoka, B. Charroux, and G. Dreyfuss.** 1998. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell **95:**615–624.
56. **Pellizzoni, L., J. Yong, and G. Dreyfuss.** 2002. Essential role for the SMN complex in the specificity of snRNP assembly. Science **298:**1775–1779.
57. **Plessel, G., U. Fischer, and R. Luhrmann.** 1994. m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. Mol. Cell. Biol. **14:**4160–4172.
58. **Pu, W. T., G. B. Krapivinsky, L. Krapivinsky, and D. E. Clapham.** 1999. pICln inhibits snRNP biogenesis by binding core spliceosomal proteins. Mol. Cell. Biol. **19:**4113–4120.
59. **Raker, V. A., K. Hartmuth, B. Kastner, and R. Luhrmann.** 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.
60. **Raker, V. A., G. Plessel, and R. Luhrmann.** 1996. The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J. **15:**2256–2269.
61. **Ryder, S. P., L. Ortoleva-Donnelly, A. B. Kosek, and S. A. Strobel.** 2000. Chemical probing of RNA by nucleotide analog interference mapping. Methods Enzymol. **317:**92–109.
62. **Sontheimer, E. J., and J. A. Steitz.** 1992. Three novel functional variants of human U5 small nuclear RNA. Mol. Cell. Biol. **12:**734–746.
63. **Stark, H., P. Dube, R. Luhrmann, and B. Kastner.** 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature **409:**539–542.
64. **Sumpter, V., A. Kahrs, U. Fischer, U. Kornstadt, and R. Luhrmann.** 1992. In vitro reconstitution of U1 and U2 snRNPs from isolated proteins and snRNA. Mol. Biol. Rep. **16:**229–240.
65. **Urlaub, H., V. A. Raker, S. Kostka, and R. Luhrmann.** 2001. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. EMBO J. **20:**187–196.
66. **Will, C. L., and R. Luhrmann.** 2001. Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. **13:**290–301.
67. **Yong, J., T. J. Golembe, D. J. Battle, L. Pellizzoni, and G. Dreyfuss.** 2004. snRNAs contain specific SMN-binding domains that are essential for snRNP assembly. Mol. Cell. Biol. **24:**2747–2756.
68. **Yong, J., L. Pellizzoni, and G. Dreyfuss.** 2002. Sequence-specific interaction of U1 snRNA with the SMN complex. EMBO J. **21:**1188–1196.
69. **Yong, J., L. Wan, and G. Dreyfuss.** 2004. Why do cells need an assembly machine for RNA-protein complexes? Trends Cell Biol. **14:**226–232.
70. **Zhang, A., K. M. Wassarman, J. Ortega, A. C. Steven, and G. Storz.** 2002. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol. Cell **9:**11–22.
71. **Zieve, G. W.** 1987. Cytoplasmic maturation of the snRNAs. J. Cell Physiol. **131:**247–254.