THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 7, Issue of February 15, pp. 5631–5636, 2002

Printed in U.S.A.

Gemin5, a Novel WD Repeat Protein Component of the SMN Complex That Binds Sm Proteins*

Received for publication, October 1, 2001, and in revised form, November 13, 2001
Published, JBC Papers in Press, November 19, 2001, DOI 10.1074/jbc.M109448200

Amelie K. Gubitz‡, Zissimos Mourelatos‡§, Linda Abel‡, Juri Rappsilber¶‖, Matthias Mann¶, and Gideon Dreyfuss‡ ‡‡**

From the ‡Howard Hughes Medical Institute, Department of Biochemistry and Biophysics and §Department of Pathology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148 and ¶Protein Interaction Laboratory, Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark

The survival of motor neurons (SMN) protein is the product of the disease gene of spinal muscular atrophy and is found both in the cytoplasm and the nucleus, where it is concentrated in gems. SMN is part of a multi-protein complex that includes Gemin2, Gemin3, and Gemin4. The SMN complex plays an important role in the cytoplasmic assembly of small nuclear ribonucleoproteins (snRNPs) and likely other RNPs in pre-mRNA splicing and in the assembly of transcriptosomes. Here, we report the identification of an additional component of the SMN complex, a novel WD repeat protein termed Gemin5. Gemin5 binds SMN directly and is a component of the SMN complex. Furthermore, Gemin5 interacts with several of the snRNP core proteins including SmB, SmD1, SmD2, SmD3, and SmE, suggesting that it participates in the activities of the SMN complex in snRNP assembly. Immunolocalization studies demonstrate that Gemin5 is found in the cytoplasm and in the nucleus, where it colocalizes with SMN in gems. The presence of 13 WD repeat domains in the amino-terminal half of Gemin5 and a coiled-coil motif near its carboxyl terminus indicate that it may form a large heteromeric complex and engage in multiple interactions.

Spinal muscular atrophy (SMA)[1] is a common neuromuscular disorder characterized by the degeneration of motor neurons in the spinal cord. Spinal muscular atrophy is an autosomal recessive disease and is the leading genetic cause of infant mortality (1, 2). The survival of motor neurons (*SMN*) gene was identified as the disease gene of spinal muscular atrophy (3), and it is present as two inverted copies on human chromosome 5, telomeric *SMN* (*SMN1*) and centromeric *SMN* (*SMN2*; Refs. 4 and 5). More than 98% of spinal muscular atrophy patients have complete or partial deletions in the telomeric copy of the *SMN* gene, and some carry loss of function mutations (3). The centromeric copy fails to protect from the disease due to a single nucleotide difference that modifies the splicing pattern, primarily generating a functionally defective form of the protein that lacks the carboxyl terminus encoded by exon 7 (6). The SMN protein is ubiquitously expressed in metazoans, and its function is essential in a range of species including human, mouse, chicken, *Caenorhabditis elegans* and *Schizosaccharomyces pombe* (3, 7–12).

In the nucleus and cytoplasm, SMN in oligomeric form is part of a multi-protein complex that contains Gemin2, Gemin3 (a DEAD box putative RNA helicase), and Gemin4 (13–16). The designation of the components of this complex as "gemins" is based on their colocalization with SMN in gems, nuclear bodies similar in size and number to Cajal (coiled) bodies (17). Cajal bodies and gems are distinct and independent nuclear structures, but there is a dynamic functional relationship between them, and they are often associated (18–20). Thus, although gems and Cajal bodies colocalize in several types of cultured cells and adult tissues, they are separate in fetal tissues and other, often rapidly proliferating, cells in culture (17, 20–22). Furthermore, whereas snRNP-rich Cajal bodies fail to assemble in coilin knock-out mice, gems are still present in embryonic fibroblasts derived from these animals (23).

SMN binds directly to Gemin2 and Gemin3, whereas Gemin4 is indirectly associated with SMN via a strong interaction with Gemin3 (16). In addition to these intrinsic components of the SMN complex, it also interacts with several other protein substrates, including Sm and Sm-like (Lsm) proteins, RNA helicase A, fibrillarin, and GAR1 (13, 24–27). Furthermore, SMN has been shown to associate with a novel zinc finger protein called ZPR1 (28).

Studies in *Xenopus laevis* oocytes and mammalian cells revealed an important role for the SMN complex in the cytoplasmic biogenesis of small nuclear ribonucleoproteins (snRNPs; Refs. 19, 24, 29, 30). Furthermore, SMN is involved in pre-mRNA splicing, most likely by ensuring the regeneration of functionally active snRNPs or other components of the spliceosome (19, 31) and in the assembly of the polymerase II transcription/processing machinery (27). The identification of all the constituents of the SMN complex is an essential step toward understanding the precise molecular mechanism of SMN function in these processes.

Here we report the identification by nanoelectrospray mass spectrometry (32, 33), the molecular cloning, and the characterization of a novel component of the SMN complex that we term Gemin5. Gemin5 is a novel, large tryptophan-aspartic acid (WD) repeat protein that binds to SMN both *in vivo* and *in vitro*. In HeLa cells, Gemin5, like the other components of the SMN complex, is found in the cytoplasm, the nucleoplasm, and in nuclear foci where it colocalizes with SMN.

* This work was supported by National Institutes of Health Grants NS 40575 (to G. D.) and K08 NS02199 (to Z. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ A Marie Curie Fellow.

** Supported by the Danish National Research Foundation of the Center of Experimental Bioinformatics.

‡‡ An Investigator of the Howard Hughes Medical Institute. To whom correspondence should be addressed. Tel.: 215-898-0398; Fax: 215-573-2000; E-mail: gdreyfuss@hhmi.upenn.edu.

[1] The abbreviations used are: SMA, spinal muscular atrophy; snRNP, small nuclear ribonucleoprotein; GST, glutathione *S*-transferase; PBS, phosphate-buffered saline.

MATERIALS AND METHODS

DNA Constructs—cDNA corresponding to the open reading frame of Gemin5 was amplified from a HeLa cDNA library using sequence-specific primers and *Pfu* Turbo (Stratagene). The full-length cDNA was subcloned into the pcDNA3.1D/V5-His TOPO vector (Invitrogen). For amino-terminal GST fusion, a partial Gemin5 cDNA encoding amino acids 43–1508 was cloned into the pGex-5x-1 vector (Amersham Biosciences, Inc.). Furthermore, a partial Gemin5 cDNA encoding amino acids 694–1508 was cloned into the pET28b vector (Novagen). The plasmid constructs for the Sm proteins, His-SMN, and His-SmB were previously described (13, 14).

Antibodies—The following antibodies were used in this study: mouse IgG1 monoclonal anti-SMN (2B1; Ref. 17); rabbit polyclonal anti-SMN/exon7 (13), mouse IgG1 monoclonal anti-Gemin2 (2E17; Ref. 13); mouse IgG1 monoclonal anti-Gemin3 (11G9 and 12H12; Ref. 14); mouse IgG1 monoclonal anti-Gemin4 (17D10)[2]; mouse IgG1 monoclonal anti-heterogeneous nuclear RNP A1 (4B10; Ref. 34); mouse IgG2a monoclonal anti-V5 (Invitrogen); mouse IgG2b monoclonal anti-T7 tag (Novagen); peroxidase-conjugated goat anti-mouse IgG plus IgM (Jackson ImmunoResearch Laboratories); Alexa Fluor 488 goat anti-rabbit IgG; and Alexa Fluor 594 goat anti-mouse IgG (Molecular Probes). Anti-Gemin5 antiserum was produced in mice as follows; His-Gemin5 (amino acids 694–1508) was expressed from a pET bacterial expression system (Novagen) in the *Escherichia coli* strain BL21(DE3) and purified by electroelution from the insoluble fraction. Polyclonal antiserum was prepared by immunizing Balb/C mice with this purified, carboxyl-terminal half of Gemin5.

Protein Microsequencing by Mass Spectrometry—Subsequent to immunoprecipitation from total HeLa cell extract using anti-Gemin3 (11G9) and anti-SMN (2B1) mouse monoclonal antibodies, a protein band of ~170 kDa was excised from a Coomassie Blue-stained SDS-PAGE. The protein was reduced, alkylated, and digested with trypsin in-gel (32). Tryptic peptides in the supernatant were desalted and concentrated on an OligoR2 (Perseptive) micro column packed in a GELoader tip (Eppendorf). Mass mapping and sequencing was performed on a QSTAR Pulsar mass spectrometer (MDS Sciex, Canada) equipped with a nanoelectrospray ion source (MDS Proteomics, Denmark). Peptide sequence tags (35) were assigned in the spectra and searched against the EST division of GenBank™ using the Protein and Peptide Software Suite (PPSS) from MDS Proteomics.

Cell Culture and Transient Transfections—HeLa and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (Invitrogen). Transient transfections were performed using the Geneporter system as specified by the manufacturer (Gene Therapy Systems). Transfected cells were processed for coimmunoprecipitation experiments 24 h after transfection.

Immunoprecipitation and Immunoblotting—Immunoprecipitations were performed using total HEK293T cell lysate prepared in the presence of RSB-200 buffer (20 mM Tris-Cl, pH 7.4, 200 mM NaCl, 2.5 mM $MgCl_2$) containing 0.05% Nonidet P-40 and protein A-Sepharose CL-4B (Amersham Biosciences, Inc.)-bound anti-V5 antibody or GammaBind G-Sepharose (Amersham Biosciences, Inc.)-bound 2B1 antibody as previously described (34). Proteins were resolved on a 10% SDS-PAGE gel, and immunoblotting was conducted according to Liu *et al.* (13). Antibodies were diluted in PBS containing 3% bovine serum albumin as follows; anti-V5 1:1000; anti-Gemin5 1:500; 2B1 1:1000; 2E17 1:500; 12H12 1:500; 17D10 1:500; peroxidase-conjugated goat anti-mouse IgG plus IgM 1:4000.

Immunofluorescence Microscopy—HeLa cells grown on glass coverslips were briefly washed in PBS, fixed in 2% formaldehyde PBS for 15 min at room temperature, and permeabilized in acetone for 1 min at −20 °C. Cells were blocked in PBS containing 3% bovine serum albumin for 30 min at room temperature. For double-labeling immunofluorescence experiments, incubations with anti-Gemin5 antiserum and anti-SMN/exon7 antibody were followed by incubations with the respective secondary antibodies (Alexa Fluor 594 goat anti-mouse IgG for anti-Gemin5 and Alexa Fluor 488 goat anti-rabbit IgG for anti-SMN/exon7). Incubations were carried out for 1 h at room temperature, and all antibodies were diluted in PBS containing 3% bovine serum albumin (anti-Gemin5 1:250; anti-SMN/exon7 1:500; Alexa Fluor 488 and 594 1:500). Laser confocal fluorescence microscopy was performed with a Leica TCS four-dimensional confocal microscope.

In Vitro Protein Binding Assays—GST-Gemin5 fusion protein was expressed from the pGex-5x-1 vector in the *E. coli* strain BL21 and purified using glutathione-Sepharose beads (Amersham Biosciences, Inc.) as specified by the manufacturer. All proteins tested were produced by *in vitro*-coupled transcription and translation (Promega) in the presence of [^{35}S]methionine (Amersham Biosciences, Inc.) using full-length cDNAs. 2 μg of purified GST, GST-SMN, GST-Gemin2 (13), GST-Gemin2/SMN (36), and GST-Gemin5 bound to 25 μl of glutathione-Sepharose beads were incubated with 10^6 cpm of the *in vitro* translated proteins in 0.8 ml of RSB-200 buffer containing 0.1% Nonidet P-40, 1× complete protease inhibitor mixture (Roche Molecular Biochemicals), and 10 μg/ml RNase A (U. S. Biochemical Corp.). After a 1-h incubation at 4 °C, the beads were pelleted and washed five times with RSB-200. Subsequently, the bound fractions were eluted by boiling in SDS-PAGE sample buffer and resolved on 11.5 or 15% SDS-PAGE gels. The gels were fixed and processed for fluorography by treatment with Amplify solution (Amersham Biosciences, Inc.). For direct binding, purified GST or GST-Gemin5 proteins (2 μg) bound to 25 μl of glutathione-Sepharose beads (Amersham Biosciences, Inc.) were incubated with 1 μg of purified T7/His-tagged SMN or T7/His-tagged SmB (14) in 0.5 ml of RSB-200 buffer containing 0.05% Nonidet P-40 and 1× complete protease inhibitor mixture (Roche Molecular Biochemicals). After incubation for 1 h at 4 °C, the resin was washed 6 times with 0.5 ml of binding buffer. The bound fraction was eluted by boiling in SDS-PAGE sample buffer and analyzed by SDS-PAGE and Western blot using anti-T7 tag antibody (Novagen; diluted 1:10,000 in PBS containing 3% bovine serum albumin).

RESULTS

Gemin5, a Novel Component of the SMN Complex That Contains Multiple WD Repeats and a Coiled-coil Oligomerization Motif—We have previously identified proteins that associate with SMN *in vivo* by coimmunoprecipitations followed by mass spectrometry of the immunopurified proteins (13, 14, 16, 27). This approach revealed that the multimeric SMN complex is comprised of Gemin2 (13), Gemin3 (14), and Gemin4 (16) and contains Sm proteins (13). Using this strategy, we here describe the identification of one of the last major constituents of the SMN complex that has not yet been characterized, a protein of ~170 kDa that consistently coimmunoprecipitates with anti-SMN and anti-Gemin3 antibodies from total HeLa cell lysates (Fig. 1). The 170-kDa protein band was digested with trypsin, and more than 20 of the recovered peptides were sequenced using nanoelectrospray mass spectrometry (32, 33). During sequencing, real-time data base searches were performed. Three partial cDNA clones with the GenBank™ accession numbers AK022748, AL117665, and BC008776 were retrieved using the peptide sequence tag algorithm. Upon alignment, these overlapping cDNAs are predicted to encode a putative protein of 1,508 amino acid residues. Clone AK022748 codes for the amino portion of the protein and comprises 741 amino acids. The predicted first methionine is preceded by an upstream in-frame stop codon. Clones AL117665 and BC008776 code for 1,224 amino acid and 921 amino acid segments of the protein, respectively, and include its carboxyl terminus. 51 peptide masses matched to the assembled protein sequence covering 375 of the 1,508 amino acid residues (25% sequence coverage at a mass accuracy below 15 ppm). Using sequence-specific primers, we isolated the full-length cDNA of this novel protein from a HeLa cDNA library. The open reading frame predicts a protein with a calculated molecular mass of 168.5 kDa and an estimated pI of 6.14 (amino acid sequence given in Fig. 2*B*). Because this protein is a component of the SMN complex and colocalized with SMN in nuclear foci (see below), we name it Gemin5.

Further analysis of the Gemin5 sequence using various protein motif search engines revealed the presence of up to 13 WD repeats in its amino-terminal half and a coiled-coil near its carboxyl terminus (Rep-V1.1 search tool (37); MultiCoil software (38); schematically depicted in Fig. 2*A*). WD repeats form stable propeller-shaped platforms that serve as protein-protein interaction domains (39). Given the fact that Gemin5 contains 13 of these repeats (Rep-V1.1 search tool (37)), the protein may

[2] L. Abel and G. Dreyfuss, unpublished information.

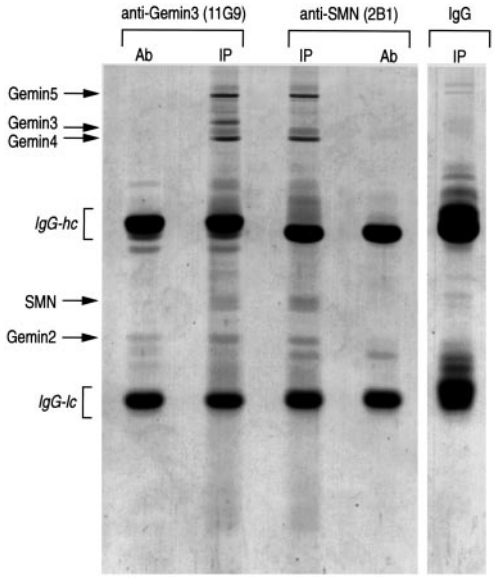

FIG. 1. **Identification of Gemin5 as a novel component of the SMN complex.** Immunoprecipitations (*IP*) with anti-Gemin3 (11G9) and anti-SMN (2B1) monoclonal mouse antibodies were performed using total HeLa cell extract. The immunoprecipitated proteins were analyzed by SDS-PAGE and visualized by Coomassie Blue staining. Antibodies 11G9 and 2B1 coimmunoprecipitated previously characterized proteins corresponding to SMN, Gemins2, -3, -4, and a novel 170-kDa protein (*Gemin5*). The 170-kDa protein was excised and identified by nanoelectrospray mass spectrometry as described under "Materials and Methods." Background from the immunoglobulin G heavy (*IgG-hc*) and light chains (*IgG-lc*) of the antibodies used for coimmunoprecipitation is marked on the *left* and shown in the immunoprecipitations with immunoglobulin G (IgG) as well as in the *first* and *fourth lane* for the antibodies 11G9 and 2B1, respectively. *Ab*, antibody.

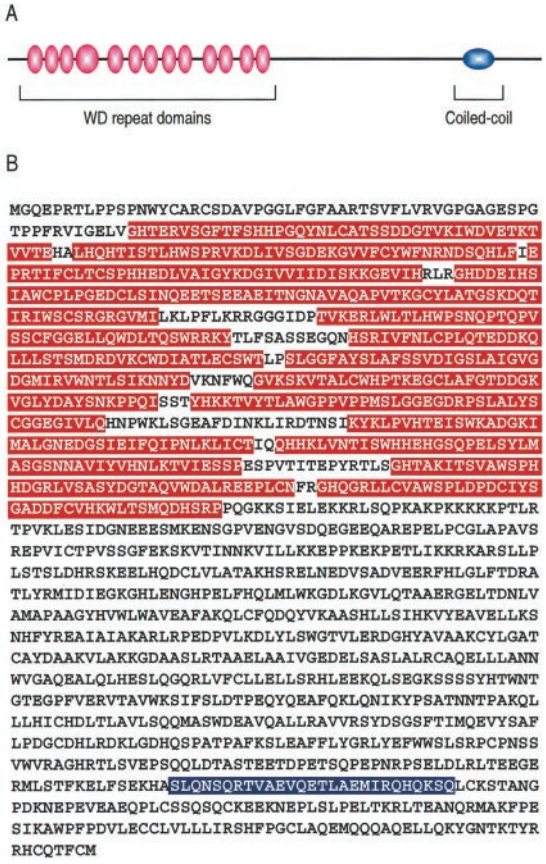

FIG. 2. **Gemin5 encodes a protein containing 13 WD repeats and a coiled-coil motif.** *A*, schematic representation of the modular structure of Gemin5. WD repeat domains are marked by *red ovals*, and the coiled-coil motif is marked by a *blue oval*. *B*, amino acid sequence of human Gemin5. Amino acids that constitute WD repeats are typed in *white* with a *red background*, and the ones that form the coiled-coil motif are in *white* with a *blue background*. The Gemin5 GenBank accession number is AY063750.

either shape one large propeller structure or possibly two or more smaller ones (39). The coiled-coil formed by amino acids 1362–1392 is predicted to permit Gemin5 to oligomerize as a trimer (maximum trimeric residue probability of 0.67, MultiCoil software (38)).

Gemin5 Is in a Complex with SMN, Gemin2, Gemin3, and Gemin4—To further confirm that Gemin5 is part of the SMN complex *in vivo*, we constructed a V5 epitope tag of the Gemin5 open reading frame, transfected it into HEK293T cells, and performed coimmunoprecipitation experiments. Total cell lysates were prepared from V5-Gemin5-transfected cells, and immunoprecipitations were carried out using the anti-V5 antibody. The immunoprecipitated proteins were resolved on SDS-PAGE, immunoblotted, and probed with several antibodies specific for known components of the SMN complex. As shown in Fig. 3*A*, the anti-V5 antibody specifically immunoprecipitated SMN, Gemin2, Gemin3, and Gemin4 (*second lane*). These proteins were neither detectable in similar immunoprecipitates from untransfected HEK293T cells using the anti-V5 antibody (Fig. 3*A*, *third lane*) nor in immunoprecipitates from transfected cells using the non-immune antibody SP2/0 (Fig. 3*A*, *fourth lane*). Furthermore, the abundant heterogeneous nuclear RNP A1 protein was not present in the anti-V5 immunoprecipitate, providing additional evidence for the specificity of the immunoprecipitations. In a reciprocal approach, total cell lysates from untransfected HEK293T cells were immunoprecipitated with the anti-SMN antibody 2B1, and the immunoprecipitate was probed with anti-Gemin5 antiserum. This antiserum specifically recognized a single protein of the expected size of Gemin5 on an immunoblot of total HeLa cell lysate (Fig. 3*B*). As shown in Fig. 3*C*, Western blotting of the 2B1 immunoprecipitate using the anti-Gemin5 antiserum provided further evidence that Gemin5 associates with SMN *in vivo* (*second lane*). The absence of a detectable signal for Gemin5 in the SP2/0 immunoprecipitate further confirmed the specificity of the immunoprecipitations (Fig. 3*C*, *third lane*). Taken together, these results demonstrate that Gemin5 is a component of the SMN complex *in vivo*.

Colocalization of Gemin5 and SMN—To determine the subcellular localization of Gemin5, we performed indirect laser confocal immunofluoresence microscopy using mouse polyclonal anti-Gemin5 antiserum on HeLa cells. As depicted in Fig. 4*A*, the anti-Gemin5 antiserum displayed a strong general staining throughout the cytoplasm. In the nucleus, there was a weaker but significant staining of the nucleoplasm and a prominent labeling of nuclear foci (see below; Ref. 17). This localization pattern is similar to that of SMN, Gemin2, and Gemin3 and partially overlaps with that of Gemin4, which is also detectable in the nucleoli (13, 14, 16, 17). Most strikingly, Gemin5 and SMN colocalize in gems, as demonstrated by double-labeling immunofluorescence using a polyclonal anti-SMN antibody and anti-Gemin5 antiserum (Figs. 4, *A*, *B*, and merged image *C*). To perform these double-labeling experiments, we used a rabbit polyclonal anti-SMN antibody because none of our anti-SMN monoclonal antibodies were compatible with the mouse polyclonal anti-Gemin5 antiserum. This rabbit antibody and the various mouse monoclonal anti-SMN antibodies we have show a very similar staining pattern to that seen for Gemin5 (13, 14, 16, 17). We conclude that Gemin5 and the components of the SMN complex colocalize *in vivo*. Gemin5 is thus the fifth component of gems identified so far.

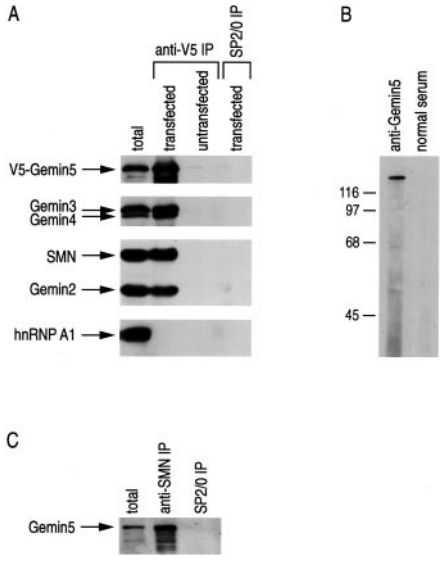

FIG. 3. **Association of Gemin5 with the SMN complex *in vivo*.** *A*, total cell lysates from HEK293T cells transiently transfected with V5-Gemin5 or untransfected were immunoprecipitated (*IP*) with anti-V5 monoclonal antibody or control non-immune antibody (SP2/0). Immunoprecipitates were resolved on SDS-PAGE and analyzed by Western blotting with antibodies specific for the indicated proteins. 2.5% of the total cell lysate is shown on the left. *B*, Western blotting using anti-Gemin5 antiserum on total HeLa cell lysates. The position of the molecular mass markers is indicated on the left in kDa. *C*, total cell lysates from HEK293T cells were immunoprecipitated with the anti-SMN monoclonal antibody 2B1 or control non-immune antibody (SP2/0). Immunoprecipitates were resolved on SDS-PAGE and analyzed by Western blotting with mouse polyclonal anti-Gemin5 antiserum. 2.5% of the total cell lysate is shown on the *left*.

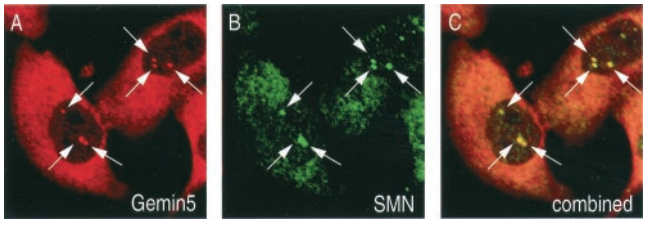

FIG. 4. **The Gemin5 protein colocalizes with SMN in gems.** *A–C*, laser confocal images of indirect double-labeling immunofluorescence on HeLa cells using anti-Gemin5 antiserum (*A*, *red*), and anti-SMN/exon7 antibody (*B*, *green*). Note that the anti-Gemin5 antiserum displays strong general cytoplasmic and weaker nucleoplasmic staining and an intense labeling of gems. *C*, superimposed laser confocal images of *A* and *B*. Colocalization of *red* (anti-Gemin5) and *green* (anti-SMN/exon7) results in a *yellow signal*. Gems are marked by *arrows* in all *panels*.

Gemin5 Binds SMN in Vitro—In vitro binding assays using V5-Gemin5 and GST-bound forms of SMN and Gemin2 were performed to assess whether these proteins interact with each other. Purified GST alone, GST-SMN, GST-Gemin2, and GST-Gemin2/SMN fusions were immobilized on glutathione-Sepharose and incubated with [^{35}S]methionine-labeled Gemin5 generated by *in vitro* transcription and translation in rabbit reticulocyte lysate. After extensive washing, bound proteins were eluted by boiling in SDS-PAGE sample buffer, resolved on SDS-PAGE, and visualized by fluorography. As shown in Fig. 5*A*, V5-Gemin5 specifically and efficiently bound to GST-Gemin2/SMN but not to purified GST alone. A distinct though lower signal for direct binding was also detected for V5-Gemin5 and GST-SMN, whereas no binding was evident between V5-Gemin5 and GST-Gemin2 (Fig. 5*A*). These findings were confirmed by binding between *in vitro* produced [^{35}S]methionine-labeled SMN and GST-Gemin5, whereas no interaction was apparent between *in vitro* produced [^{35}S]methionine-labeled

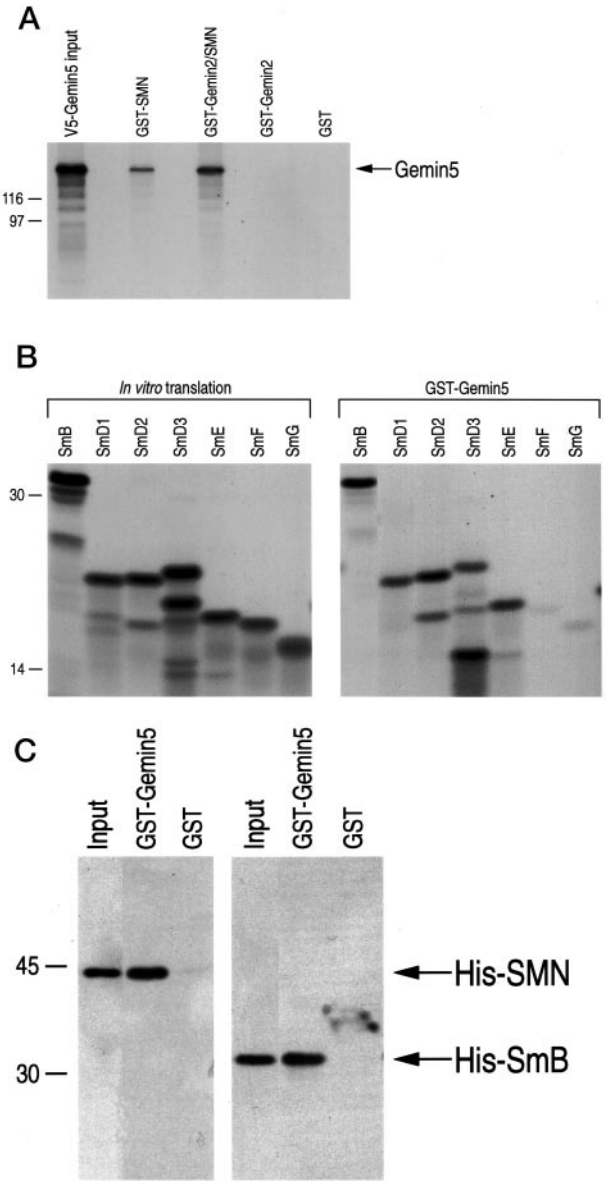

FIG. 5. **Gemin5 binds SMN and several Sm proteins *in vitro*.** *A*, *in vitro* translated [^{35}S]methionine-labeled V5-Gemin5 was incubated with purified GST-SMN, GST-Gemin2/SMN, GST-Gemin2, and GST alone as described under "Materials and Methods." The bound protein fraction was analyzed by SDS-PAGE and fluorography. 10% of the *in vitro* translated V5-Gemin5 input is shown, and the position of molecular mass markers is indicated on the left in kDa. *B*, *in vitro* translated [^{35}S]methionine-labeled myc-Sm proteins B, D1, D2, D3, E, F, and G were incubated with purified GST-Gemin5 or GST alone (data for GST alone is not shown). The bound protein fraction was analyzed by SDS-PAGE and fluorography. 8.5% of the *in vitro* translated Sm protein input is shown, and the position of molecular mass markers is indicated on the left in kDa. *C*, recombinant His-SMN or His-SmB proteins were incubated with purified GST-Gemin5 or GST alone. *Input lanes* show 50% of His-SMN and His-SmB. Bound proteins were analyzed by SDS-PAGE and Western blot. The position of molecular mass markers is indicated on the *left* in kDa.

Gemin2 and GST-Gemin5 (data not shown). These data suggest that Gemin5 can bind to SMN and that this interaction may be more efficient for the SMN/Gemin2 heteromer.

Several of the Sm snRNP core proteins interact with SMN, Gemin3, and Gemin4. These interactions are likely to be important for the function of the SMN complex in snRNP assembly. To investigate whether Gemin5 also interacts with Sm proteins, GST-Gemin5 was used in binding assays with the *in vitro* produced [^{35}S]methionine-labeled Sm proteins B, D1, D2,

D3, E, F, and G. As shown in Fig. 5*B*, the Sm proteins B, D1, D2, D3, and E bound to GST-Gemin5, whereas no binding was detectable to GST alone (data not shown). In contrast, the interaction of Gemin5 with SmF or SmG was very weak (Fig. 5*B*). We therefore conclude that, like several other components of the SMN complex, Gemin5 interacts with a subset of the Sm proteins.

To determine whether the binding of Gemin5 to SMN and SmB is direct, full-length SMN and SmB produced as recombinant T7/His-tagged proteins (14) were incubated with GST or GST-Gemin5. After several rounds of washing, bound proteins were solubilized by boiling in SDS-PAGE sample buffer, resolved by SDS-PAGE, immunoblotted, and probed with a mouse monoclonal antibody specific for the T7 tag of the recombinant proteins. As shown in Fig. 5*C*, His-SMN and His-SmB bound specifically to GST-Gemin5 but not to GST alone. We conclude that both SMN and SmB bind directly to Gemin5.

DISCUSSION

Employing a biochemical immunoaffinity strategy to identify additional components of the SMN complex, we have identified, cloned, and characterized a novel protein termed Gemin5. Gemin5 associates with SMN and Gemins2, -3, and -4 *in vivo* and binds to SMN by direct protein-protein interaction *in vitro*. Furthermore, Gemin5 interacts with several of the snRNP core proteins including SmB, SmD1, SmD2, SmD3, and SmE, suggesting that it plays a role in the activities of the SMN complex in snRNP assembly. Finally, the subcellular distribution of Gemin5 is similar to that of SMN; both Gemin5 and SMN display a general localization in the cytoplasm, are found throughout the nucleoplasm, and are highly enriched and clearly colocalized in nuclear gems.

Sequence analysis of Gemin5 predicts two structural motifs, a series of up to 13 WD repeats in its amino-terminal half and an α-helical coiled-coil near its carboxyl terminus. With regard to the predicted WD repeats, it should be noted that depending on the protein motif search engine used, there is a slight variability in the number of WD repeats identified. For example, the SMART tool (simple modular architecture research tool (40)) predicts 12 WD repeats, the Protein Sequence Analysis tool (BMERC-www.bu.edu/wdrepeat) identifies 11 repeats, and the Rep-V1.1 search tool (37) predicts 13 repeats. The positions of the WD domains predicted by these software tools greatly overlap. However, the putative WD repeats spanning amino acid residues 193–264 and 280–321 are not recognized by all search engines.

Interestingly, both WD domains and coiled-coil motifs are protein-protein interaction domains. WD repeats have been identified in more than 150 functionally diverse proteins that are involved in processes such as signal transduction, pre-mRNA processing, transcriptional regulation, vesicular traffic, and cytoskeleton assembly (for reviews, see Refs. 39 and 41). Each repeat comprises a conserved core of ~40–60 amino acids that is bracketed by the dipeptides glycine-histidine (GH) and tryptophan-aspartic acid (WD) and a variable domain of 7–11 amino acids (41). However, neither the GH nor the WD dipeptides are entirely conserved but are often replaced by alternative amino acids that allow the same tertiary structures to form (39). The distribution and frequency of amino acid residues found in more than 750 non-redundant WD repeats are listed at BMERC-www.bu.edu/wdrepeat.

The crystal structure of the Gβ subunit of G proteins, a prototypic WD protein with seven repeats, has revealed that it forms an α-propeller structure of seven blades (for review, see Ref. 39). This propeller-like structure is thought to be a common feature of all WD repeat proteins and is predicted to offer a surface for protein-protein interactions. The number of "blades" within the "propeller," however, varies according to the number of WD repeats per protein, and proteins that encompass more the eight repeats may form one large or several smaller propellers (39). Accordingly, determination of the crystal structure of Gemin5 will be required to reveal whether its 13 WD repeat domains assemble into several smaller or one large propeller-like structure. Gemin5 is one of the largest WD repeat proteins described to date.

Coiled-coil domains are a major oligomerization motif of proteins (for review, see Ref. 42). They are composed of two or more α-helices that wrap around one another to form a superhelical structure and are stabilized by a characteristic "knobs-into-holes" packaging of apolar side chains into a hydrophobic core (42, 43). Coiled-coils are a common and versatile folding motif found in proteins with cytoskeleton-like structural function, motor proteins, membrane fusion proteins, and transcription factors (Ref. 44 and references therein and Ref. 42). In the case of Gemin5, the amino acids of its coiled-coil most likely enable it to form a trimeric oligomer (MultiCoil software (38)).

Because both WD repeat domains and coiled-coil motifs are found in a wide range of functionally diverse proteins, their presence does not provide direct clues about the biological activity of a given protein (39, 42). Correspondingly, the combination of WD repeats and coiled-coils has been observed in a number of divergent proteins that include the actin-binding protein coronin (45), the mitochondria-associated protein NET2 of *Saccharomyces cerevisiae* (46), and the photomorphogenic repressor COP1 of *Arabidopsis* (47). However, the presence of 13 WD repeats in the amino-terminal half of Gemin5 clearly endow it with numerous surfaces for protein interactions, and it is therefore probable that Gemin5 acts as part of a heteromeric protein complex or serves as a structural platform for protein assembly. Moreover, the coiled-coil motif in the carboxyl portion of Gemin5 further increases the potential complexity of the structures that Gemin5 may form.

Taken together, our data demonstrate that Gemin5 is a novel interactor of SMN. Gemin5 is the fifth integral component of the SMN complex characterized so far. Based on immunoprecipitations with antibodies specific for SMN and other components of the SMN complex (13, 14, 16, 27), it appears most likely that, with the identification of Gemin5, the inventory of intrinsic components of the SMN complex is close to being complete.

Acknowledgments—We are grateful to members of our laboratory, in particular Drs. Westley Friesen, Livio Pellizzoni, and Bernard Charroux for stimulating discussions and critical comments on this manuscript.

REFERENCES

1. Pearn, J. (1980) *Lancet* **1,** 919–922
2. Melki, J. (1997) *Curr. Opin. Neurol.* **10,** 381–385
3. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A., and Melki, J. (1995) *Cell* **80,** 155–165
4. Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., Wood, D., Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L., and Gilliam, T. C. (1990) *Nature* **344,** 540–541
5. Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Le Paslier, D., Cohen, D., Weissenbach, J., and Munnich, A. (1994) *Science* **264,** 1474–1477
6. Lorson, C. L., Hahnen, E., Androphy, E. J., and Wirth, B. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6307–6311
7. Schrank, B., Gotz, R., Gunnersen, J. M., Ure, J. M., Toyka, K. V., Smith, A. G., and Sendtner, M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9920–9925
8. Miguel-Aliaga, I., Culetto, E., Walker, D. S., Baylis, H. A., Sattelle, D. B., and Davies, K. E. (1999) *Hum. Mol. Genet.* **8,** 2133–2143
9. Hannus, S., Buhler, D., Romano, M., Seraphin, B., and Fischer, U. (2000) *Hum. Mol. Genet.* **9,** 663–674
10. Owen, N., Doe, C. L., Mellor, J., and Davies, K. E. (2000) *Hum. Mol. Genet.* **9,** 675–684
11. Paushkin, S., Charroux, B., Abel, L., Perkinson, R. A., Pellizzoni, L., and Dreyfuss, G. (2000) *J. Biol. Chem.* **275,** 23841–23846

12. Wang, J., and Dreyfuss, G. (2001) *J. Biol. Chem.* **276,** 9599–9605
13. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997) *Cell* **90,** 1013–1021
14. Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999) *J. Cell Biol.* **147,** 1181–1194
15. Campbell, L., Hunter, K. M., Mohaghegh, P., Tinsley, J. M., Brasch, M. A., and Davies, K. E. (2000) *Hum. Mol. Genet.* **9,** 1093–1100
16. Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000) *J. Cell Biol.* **148,** 1177–1186
17. Liu, Q., and Dreyfuss, G. (1996) *EMBO J.* **15,** 3555–3565
18. Matera, A. G., and Frey, M. R. (1998) *Am. J. Hum. Genet.* **63,** 317–321
19. Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998) *Cell* **95,** 615–624
20. Carvalho, T., Almeida, F., Calapez, A., Lafarga, M., Berciano, M. T., and Carmo-Fonseca, M. (1999) *J. Cell Biol.* **147,** 715–728
21. Young, P. J., Le, T. T., thi Man, N., Burghes, A. H., and Morris, G. E. (2000) *Exp. Cell Res.* **256,** 365–374
22. Young, P. J., Le, T. T., Dunckley, M., Nguyen, T. M., Burghes, A. H., and Morris, G. E. (2001) *Exp. Cell Res.* **265,** 252–261
23. Tucker, K. E., Berciano, M. T., Jacobs, E. Y., LePage, D. F., Shpargel, K. B., Rossire, J. J., Chan, E. K., Lafarga, M., Conlon, R. A., and Matera, A. G. (2001) *J. Cell Biol.* **154,** 293–307
24. Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11167–11172
25. Friesen, W. J., and Dreyfuss, G. (2000) *J. Biol. Chem.* **275,** 26370–26375
26. Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001) *Curr. Biol.* **11,** 1079–1088
27. Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001) *J. Cell Biol.* **152,** 75–85
28. Gangwani, L., Mikrut, M., Theroux, S., Sharma, M., and Davis, R. J. (2001) *Nat. Cell Biol.* **3,** 376–383
29. Fischer, U., Liu, Q., and Dreyfuss, G. (1997) *Cell* **90,** 1023–1029
30. Buhler, D., Raker, V., Luhrmann, R., and Fischer, U. (1999) *Hum. Mol. Genet.* **8,** 2351–2357
31. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000) *Hum. Mol. Genet.* **9,** 1977–1986
32. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) *Anal. Chem.* **68,** 850–858
33. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) *Nature* **379,** 466–469
34. Pinol-Roma, S., Choi, Y. D., Matunis, M. J., and Dreyfuss, G. (1988) *Genes Dev.* **2,** 215–227
35. Mann, M., and Wilm, M. (1994) *Anal. Chem.* **66,** 4390–4399
36. Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001) *Mol. Cell* **7,** 1111–1117
37. Andrade, M. A., Ponting, C. P., Gibson, T. J., and Bork, P. (2000) *J. Mol. Biol.* **298,** 521–537
38. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci.* **6,** 1179–1189
39. Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999) *Trends Biochem. Sci.* **24,** 181–185
40. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) *Nucleic Acids Res.* **28,** 231–234
41. Neer, E. J., Schmidt, C. J., Nambudripad, R., and Smith, T. F. (1994) *Nature* **371,** 297–300
42. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) *Trends Cell Biol.* **11,** 82–88
43. Lupas, A. (1997) *Curr. Opin. Struct. Biol.* **7,** 388–393
44. Newman, J. R., Wolf, E., and Kim, P. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 13203–13208
45. Asano, S., Mishima, M., and Nishida, E. (2001) *Genes Cells* **6,** 225–235
46. Cerveny, K. L., McCaffery, J. M., and Jensen, R. E. (2001) *Mol. Biol. Cell* **12,** 309–321
47. Torii, K. U., McNellis, T. W., and Deng, X. W. (1998) *EMBO J.* **17,** 5577–5587