The Reaction Mechanism of Cellular U snRNP Assembly

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Bayerischen Julius-Maximilians-Universität Würzburg

vorgelegt von

Ashwin Chari

Aus Bangalore (Indien)

Würzburg 2009

Eingereicht am:

Mitglieder der Promotionskommission:
Vorsitzender: Prof. Dr. M. Müller
1. Gutachter: Prof. Dr. U. Fischer
2. Gutachter: Prof. Dr. U. Scheer

Tag des Promotionskolloquiums:
Doktorurkunde ausgehändigt am:

Erklärung

Erklärung gemäss §4 Absatz 3 der Promotionsordnung der Fakultät für Biologie der Bayerischen Julius-Maximilians-Universität Würzburg vom 15. März 1999

1. Hiermit erkläre ich ehrenwörtlich, dass ich die vorliegende Dissertation selbstständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

2. Ich erkläre, dass die vorliegende Dissertation weder in gleicher noch in ähnlicher Form bereits in einem Prüfungsverfahren vorgelegen hat.

3. Ich erkläre, dass ich ausser den mit dem Zulassungsantrag urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht habe.

Würzburg, 2009

Ashwin Chari

Table of Contents

1. Summary

Macromolecular complexes, also termed molecular machines, facilitate a large spectrum of biological reactions and tasks crucial to the survival of cells. These complexes are composed of either protein only, or proteins bound to nucleic acids (DNA or RNA). Prominent examples for each class are the proteosome, the nucleosome and the ribosome. How such units are assembled within the context of a living cell is a central question in molecular biology. Earlier studies had indicated that even very large complexes such as ribosomes could be reconstituted from purified constituents *in vitro*. The structural information required for the formation of macromolecular complexes, hence, lies within the subunits itself and, thus, allow for self-assembly. However, increasing evidence suggests that *in vivo* many macromolecular complexes do not form spontaneously but require assisting factors ("assembly chaperones") for their maturation.

In this thesis the assembly of RNA-protein (RNP) complexes has been studied by a combination of biochemical and structural approaches. A resourceful model system to study this process is the biogenesis pathway of the uridine-rich small nuclear ribonucleoproteins (U snRNPs) of the spliceosome. This molecular machine catalyzes pre-mRNA splicing, i.e. the removal of non-coding introns and the joining of coding exons to functional mRNA. The composition and architecture of U snRNPs is well defined, also, the nucleo-cytoplasmic transport events enabling the formation of these particles in vivo have been analyzed in some detail. Furthermore, recent studies suggest that the formation of U snRNPs *in vivo* is mediated by an elaborate assembly machinery consisting of protein arginine methyltransferase (PRMT5)- and survival motor neuron (SMN)-complexes. The elucidation of the reaction mechanism of cellular U snRNP assembly would serve as a paradigm for our understanding of how RNA-protein complexes are formed in the cellular environment.

The following key findings were obtained as part of this study:

1) Efforts were made to establish a full inventory of the subunits of the SMN-complex. This was achieved by the biochemical definition and characterization of an atypical component of this complex, the unrip protein. This protein is associated with the SMN-complex exclusively in the cytoplasm and influences its subcellular localization.

2) With a full inventory of the components in hand, the architecture of the SMN-complex was defined on the basis of an interaction map of all subunits. This study elucidated that the proteins SMN, Gemin7 and Gemin8 form a backbone, onto which the remaining subunits adhere in a modular manner.

3) The two studies mentioned above formed the basis to elucidate the reaction mechanism of cellular U snRNP assembly. Initially, an early phase in the SMN-assisted formation of U snRNPs was analyzed. Two subunits of the U7 snRNP (LSm10 and 11) were found to interact with the PRMT5-complex, without being methylated. This report suggests that the stimulatory role of the PRMT5-complex is independent of its methylation activity.

4) Key reaction intermediates in U snRNP assembly were found and characterized by a combination of biochemistry and structural studies. Initially, a precursor to U snRNPs with a sedimentation coefficient of 6S is formed by the pICln subunit of the PRMT5-complex and Sm proteins. This intermediate was shown to constitute a kinetic trap in the U snRNP assembly reaction. Progression towards the assembled U snRNP depends on the activity of the SMN-complex, which acts as a catalyst. The formation of U snRNPs is shown to be structurally similar to the way clamps are deposited onto DNA to tether poorly processive polymerases.

5) The human SMN-complex is composed of several subunits. However, it is unknown whether all subunits of this entity are essential for U snRNP assembly. A combination of bioinformatics and biochemistry was applied to tackle this question. By mining databases containing whole-genome assemblies, the SMN-Gemin2 heterodimer is recognized as the most ancestral form of the SMN-complex. Biochemical purification of the Drosophila melanogaster SMN-complex reveals that this complex is composed of the same two subunits. Furthermore, evidence is provided that the SMN-Gemin2 heterodimer is necessary and sufficient to promote faithful U snRNP assembly.

Future studies will adress further details in the reaction mechanism of cellular U snRNP assembly. The results obtained in this thesis suggest that the SMN and Gemin2 subunits are sufficient to promote U snRNP formation. What then is the function of the remaining subunits of the SMN-complex? The reconstitution schemes established in this thesis will be

instrumental to address this question. Furthermore, additional mechanistic insights into the U snRNP assembly reaction will require the elucidation of structures of the assembly machinery trapped at various states. The prerequisite for these structural studies, the capability to generate homogenous complexes in sufficient amounts, has been accomplished in this thesis.

2. Zusammenfassung

Makromolekulare Komplexe, auch molekulare Maschinen genannt, ermöglichen eine grosse Vielfalt biologischer Reaktionen und Aufgaben, die für das Überleben von Organismen kritisch sind. Diese Komplexe bestehen entweder nur aus Protein, oder setzen sich aus Protein und Nukleinsäure (DNA oder RNA) zusammen. Prominente Beispiele für diese Klassen molekularer Maschinen sind das Proteosom, das Nukleosom oder das Ribosom. Wie sich solche Einheiten innerhalb einer Zelle zusammenlagern ist eine grundlegende Frage der Molekularbiologie. Frühere Studien hatten angeduetet, dass es möglich ist sogar sehr grosse Komplexe wie das Ribosom *in vitro* aus gereinigten Bestandteilen zu einem aktiven Partikel zu rekonstruieren. Die Strukturinformation, die für die Bildung von makromolekularen Komplexen erforderlich ist, liegt also in den Untereinheiten selbst. Im Gegensatz dazu mehren sich heute die Hinweise dafür, dass sich viele makromolekulare Komplexe nicht spontan zusammenlagern, sondern die Aktivität assistierender Faktoren („Assembly Chaperone“) für ihre Reifung benötigen.

In dieser Arbeit wurde der Zusammenbau von RNA-Protein (RNP) Partikeln durch eine Kombination aus Biochemie und Strukturbiologie untersucht. Ein ergiebiges System, um diesen Prozess zu studieren, ist die Biogenese der RNPs (U snRNPs) des Spleissosoms. Aufgabe dieser molekularen Maschine ist das Herausschneiden nicht-kodierender Introns und das Zusammenfügen kodiereneder Exons um so funktionelle mRNA zu bilden. Die Zusammensetzung und Architektur von U snRNPs sind gut definiert. Auch ist der Kern-Zytoplasma Transport, der für die Reifung dieser Partikel notwendig sind, detailliert beschrieben worden. Außerdem weisen neueste Studien darauf hin, dass die Bildung von U snRNPs *in vivo* durch eine komplexe Maschinerie, die aus den Protein-Arginin-Methyltransferase 5 (PRMT5)- und Survival-Motor-Neuron (SMN)- Komplexen besteht, vermittelt wird. Die Entschlüsselung des Reaktionsmechanismus des zellulärem U snRNP Zusammenbaus würde als Musterbeispiel für unser Verständnis dienen, wie RNPs in einer Zelle gebildet werden.

Folgende Erkenntnisse wurden in dieser Arbeit gewonnen:

1) Es wurde zunächst versucht eine komplette Bestandsliste der Untereinheiten des SMN-Komplexes zu erstellen. Dies wurde durch die biochemische Definition und

Charakterisierung einer atypischen Komponente dieses Komplexes, des Unrip Proteins, erreicht. Dieses Protein bindet ausschliesslich im Zytoplasma an den SMN-Komplex und beeinflusst dessen subzelluläre Lokalisation.

2) Die komplette Inventarisierung des SMN-Komplexes ermöglichte die Untersuchung der Wechselwirkung aller Untereinheiten und somit die Untersuchung seiner Architektur. Diese Studie zeigte, dass die Proteine SMN, Gemin7 und Gemin8 das Rückgrat des SMN-Komplexes bilden auf dem die restlichen Untereinheiten modular angeordnet werden.

3) Die zwei oben erwähnten Studien bildeten die Grundlage, den Reaktionsmechanismus zellulärer U snRNP Zusammenlagerung zu entschlüsseln. Zunächst wurde eine frühe Phase im SMN-vermittelten U snRNP Zusammenbau analysiert. Es konnte gezeigt werden, dass zwei Untereinheiten des U7 snRNP (LSm10 und 11) mit dem PRMT5-Komplex wechselwirken, ohne methyliert zu werden. Dies deutet darauf hin, dass die unterstützende Rolle des PRMT5-Komplexes von seiner Methylierungsaktivität unabhängig ist.

4) Schlüsselintermediate im Zusammenschluss von U snRNPs wurden identifiziert und durch eine Kombination von Biochemie und Strukturbiologie charakterisiert. In einer ersten Stufe bildet sich ein Vorgänger von U snRNPs mit einem Sedimentationskoeffizienten von 6S aus. Dieses Intermediat, bestehend aus pICln (einer Untereinheit des PRMT5-Komplexes) und Sm Proteinen, stellt eine kinetische Falle in der U snRNP Zusammenlagerung dar. Das Voranschreiten zum maturen U snRNP hängt von der Aktivität des SMN-Komplexes ab, der als Katalysator wirkt. Weiterhin konnte gezeigt werden, dass die Ausbildung von U snRNPs strukturell ähnlich zu der Reaktion verläuft, die Polymerasen mit geringer Prozessivität an der DNA verankert und die als „clamp-loading“ bezeichnet wird.

5) Der menschliche SMN-Komplex setzt sich aus mehreren Untereinheiten zusammen. Es ist jedoch unbekannt, ob alle Teile des Komplexes für die Zusammenlagerung von U snRNPs notwendig sind. Diese Frage wurde durch eine Kombination aus Bioinformatik und Biochemie adressiert. Durch Datenbanksuchen in komplett sequenzierten Genomen wurde festgestellt, dass die evolutionär ursprüngliche Form des SMN-Komplexes aus

den zwei Proteinen SMN und Gemin2 besteht. Die biochemische Reinigung des Komplexes der Taufliege Drosophila melanogaster offenbarte, dass er auch in diesem Organismus aus denselben zwei Untereinheiten zusammengebaut ist. Außerdem wurde der Beweis erbracht, dass das SMN-Gemin2 heterodimer notwendig und hinreichend ist, um U snRNPs akkurat zusammenzulagern.

Zukünftige Studien werden weitere detaillierte Ansichten des Reaktionsmechanismus in der zellulären Zusammenlagerung von U snRNPs liefern. Die Ergebnisse, die in der vorliegenden Arbeit erhalten wurden, deuten darauf hin, dass die Untereinheiten SMN und Gemin2 des SMN-Komplexes für den Zusammenbau von U snRNPs hinreichend sind. Was also ist die Funktion der weiteren Untereinheiten des SMN-Komplexes? Die Rekonstitutionsschemata, die in dieser Arbeit etabliert wurden, werden essentiell für die Beantwortung dieser Frage sein. Darüberhinaus werden weitere mechanistische Einsichten in die Zusammenlagerung von U snRNPs von der Ermittlung von Strukturen der Assembly-Maschinerie in verschiedenen Zuständen abhängen. Die Voraussetzung für diese strukturbiologische Untersuchungen, die Möglichkeit ausreichende Mengen homogener Komplexe herzustellen, ist ebenfalls in dieser Arbeit geschaffen worden.

3. Introduction

Daily life is greatly simplified by a multitude of gadgets, which help us perform our tasks. In order to manufacture these machines, for instance cars, initially manufactured components and an industrial design by an engineer are required. This blueprint essentially specifies where each individual part of the car should be located in three dimensions (Figure 1A, left column). It is then in factories that assembly lines add all the components in a sequential manner, as specified in the blueprint, to build the car. Biological devices, also termed macromolecular complexes, conduct the biochemical reactions, which are critical for the survival of organisms [1,2]. A major focus of scientific research today deals with the question: How is this machinery built in organisms? It turns out that similar principles apply as outlined above for the assembly of cars.

Biological machines are formed by the interaction of proteins amongst themselves or by their association with DNA or RNA. Prominent examples include the proteosome, the nucleosome and the ribosome. The formation of these structures also requires a blueprint (Figure 1A, right column). As indicated by early studies, it is possible to reconstitute nucleosomes, or even the 2.6 MDa ribosome from purified constituents into an active assembly *in vitro* [3-6]. These findings thus indicate that the structural information (blueprint) for the formation of macromolecular complexes is encoded within the subunits itself. Furthermore, these observations have also given rise to the notion that the machines of life are built spontaneously without the requirement for accessory factors. Despite this general tendency of self-assembly of macromolecular complexes *in vitro*, it is becoming increasingly clear that assisting factors are involved *in vivo*. They consist in an unrelated group of proteins termed molecular chaperones, and act as a cellular counterpart to industrial assembly lines [7]. They associate with the individual subunits of a macromolecular complex and promote the formation of the final structure, as illustrated here for nucleosomes (Figure 1A, right column).

3.1 Principles Governing Macromolecular Complex Assembly *In Vivo*

From a biophysical view the assembly of macromolecular complexes depends on the diffusion-driven, random collisions between the subunits [8]. The stability of the resulting assembly is determined by the ratio of the individual dissociation and association rate constants. However, within cells variations to this simplified view appear prevalent. First,

many subunits of an assembly often complete their individual folding reaction only upon association with their interaction partner. Second, the interaction interfaces between two or more subunits are mainly hydrophobic in nature, and exposed on the surface of the folded subunit. These properties make isolated proteins of a macromolecular complex particularly prone to irreversible aggregation within cells. Third, an additional pitfall towards the formation of macromolecular entities arises from the fact that cellular milieus do not behave as ideal fluids. Rather, the interior of cells is chararacterized by a high total concentration of many different macromolecules. As a consequence, no single species accumulates in high amounts, but in total proteins and other macromolecules are so abundant that they occupy 20-30% of the interior of cells. Therefore, this volume is unavailable to other molecules giving rise to an excluded volume effect that is commonly referred to as macromolecular crowding (Figure 1B) [9,10]. This property of cells lowers the diffusion rate of individual subunits within a cell. Thus, the likelihood of random encounters of the building blocks of a macromolecular entity is strongly reduced. Simultaneously, specific, as well as non-specific association reactions, are strongly favored under crowding conditions. Therefore, the propensity that productive interactions between cognate molecules occur within cells is substantially lowered, giving rise to misassembly. Fourth, partially assembled complexes potentially are capable of interaction with substrates, thereby adversely affecting the activity of the fully assembled complex.

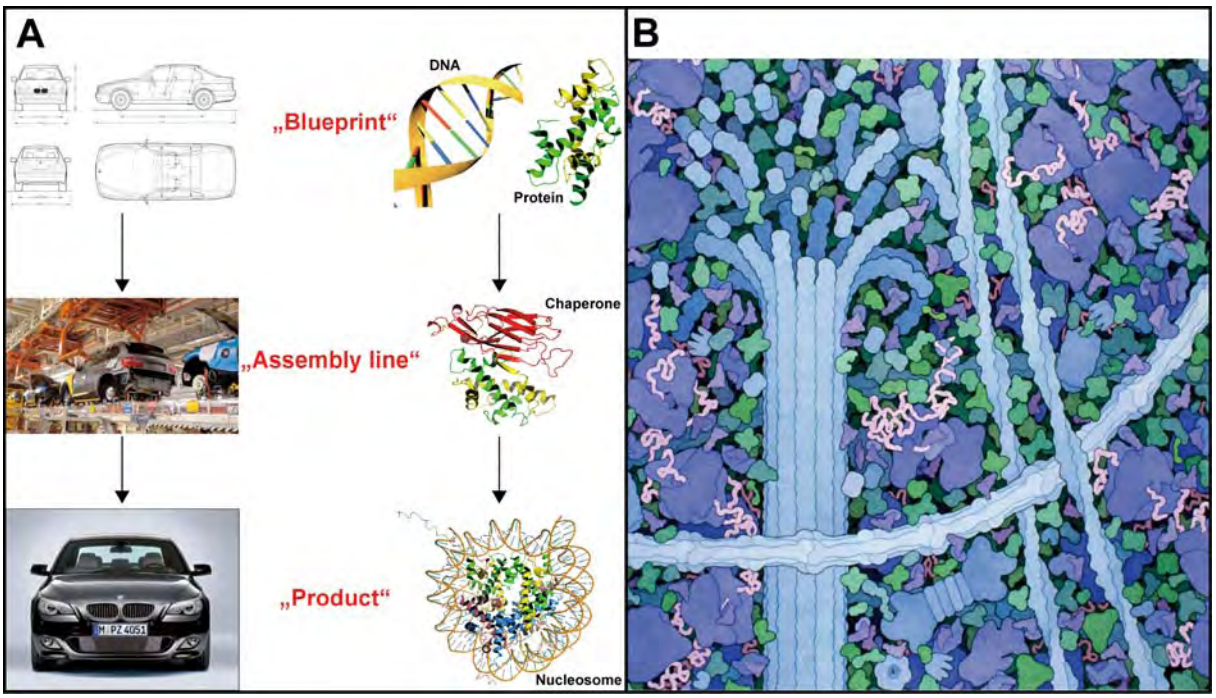


Figure 1: A) The formation of industrial and biological machines. The left column illustrates the industrial assembly of cars. Initially, a blueprint designed by an engineer specifies where the individual parts of a car should be located. An assembly line then adds all components together as specified in the blueprint, which gives rise to the product, a car. Similar principles apply to the formation of biological machines (right column). Also these structures require a blueprint, which albeit is encoded in the DNA and protein components (PDB ID: 2HIO). A chaperone then acts as a counterpart for the industrial assembly line within cells (PDB ID: 2HUE). It associates with the protein components and orchestrates the formation of the final product; in this case a nucleosome core particle (PDB ID: 1AOI). **B) The cytoplasm of cells is a crowded environment.** The illustration depicts an artist´s rendition of macromolecular crowding conditions within a cell. A small portion of eukaryotic cytoplasm is shown, including three types of filaments that make up the cytoskeleton: a microtubule (the largest), an intermediate filament (the knobby one) and two actin filaments (the smallest ones). The large blue molecules are ribosomes, busy in their task of synthesizing proteins. The large protein at bottom center is a proteosome. Reproduced, with permission, from D.S. Goodsell.

Cells have evolved an arsenal of strategies to encounter these obstacles to faithful macromolecular assembly. The most prominent are listed in the following: 1) The synthesis of the individual constituents of a macromolecular entity is temporally controlled. This strategy ensures that all components of a macromolecular complex are simultaneously present when assembly is initiated. 2) The transport of individual subunits from their site of biosynthesis to a spatially distinct subcellular compartment is a prerequisite for assembly. In the case of RNA- and DNA- protein complexes transport across the nuclear envelope is necessary, as the translation of proteins happens in the cytoplasm, while nucleic acids are made in the nucleus. This is often followed by yet another transport step after assembly to the subcellular place of function. The latter strategy prevents partially assembled complexes from adversely affecting the activity of the fully assembled complex. 3) Additionally, cells often resort to the activity of assembly chaperones. These proteins enable the formation of intermediate higher-order structures. This ensures a spatial pre-organization of individual subunits in positions, which they also occupy in the assembled macromolecular complex. Simultaneously, they prevent the premature occurrence of steps, which are intended to happen at a later stage in assembly, ensuring temporal control. The result of this strategy is in many cases the formation of kinetically trapped intermediates, which require the activity of additional factors for assembly to progress. 4) Trans-acting factors, which act as scaffolds, to bring together all the individual subunits or higher-order structures, increase their local concentration, resolve kinetic traps, and actively promote their association to the final product.

Prominent examples for the utilization of these strategies *in vivo* include the assembly of Rubisco, proteosomes and nucleosomes [11-13]. In contrast, how RNA-protein complexes

(RNPs) are formed in the crowded cellular environment is largely unknown today. Therefore, the major goal of this thesis was the elucidation of the mechanisms underlying cellular RNP assembly. An ideal model system to study these processes is the biogenesis pathway of the uridine-rich small nuclear ribonucleoproteins (U snRNPs) of the spliceosome. The composition and architecture of these particles is well defined and also the nucleo-cytoplasmic transport events leading to the formation of U snRNPs have been analyzed in some detail. Furthermore, as outlined below, key steps in U snRNP formation *in vivo* are factor-mediated and active processes. Therefore, the elucidation of the reaction mechanism of cellular U snRNP assembly was expected to serve as a paradigm for our understanding of how RNA-protein complexes are formed in the cellular environment.

3.2 Pre-mRNA Splicing

A cruicial step in the eukaryotic gene expression cycle consists in the excision of long, intervening segments (introns) and the joining of protein-coding parts (exons) from precursor mRNA (pre-mRNA) molecules (Figure 2A) [14]. This gives rise to mature templates (mRNA) for protein expression. The sequence elements that direct removal of introns from pre-mRNA are located at the 5´ or 3´- end of each intron and consist in the 5´splice site (5´ss), the branch point sequence (BPS), and the 3´splice site (3´ss; Figure 2B, upper panel). Chemically, the splicing reaction consists of two S_N2-type transesterification reactions, in which first the 2´-OH of an invariant adenosine in the BPS attacks the phosphodiester bond at the 5´ss (Figure 2B, lower panel). This leads to the formation of a free 5´exon and a lariat-shaped intron-3´exon. Subsequent attack of the phospodiester bond at the 3´ss by the 3´-OH of the liberated 5´exon, leads to the joining of the two exons and the release of the intron lariat. Splice site sequences are very short and poorly conserved in eukaryotic genes. Therefore, they are incapable of forming stable 3-dimensional structures that would position these elements in spatial proximity to each other as observed in group I and II introns, which are capable of self-splicing [15]. Thus, splicing in eukaryotes is dependent on a large number of *trans*-acting factors that assemble on introns and form the spliceosome.

Major building blocks of the spliceosome are the uridylic-rich small nuclear ribonucleoprotein particles (U snRNPs) U1, U2, U4/U6 and U5 [14]. U4/U6 and U5 snRNPs are found associated with each other in cells and form the U4/U5/U6 tri- snRNP [16]. Additional non-snRNP proteins complement these RNA-protein complexes. The spliceosome

forms anew on each intron, starting with the binding of the U1 snRNP to the 5´ss (Figure 2B, lower panel). In a next step the U2 snRNP recognizes the BPS giving rise to the spliceosomal A complex. It is only then that the pre-formed U4/U5/U6 tri- snRNP joins, this addition leads to the formation of the B complex. Although, all U snRNPs are present in the B complex the spliceosome is catalytically inactive at this stage. Catalytic activation requires extensive rearrangement and the formation of a RNA network composed of U snRNAs and the pre-mRNA. Finally, U1 and U4 snRNPs dissociate from the spliceosome to form the B* complex, in which U6 base pairs with the 5´ss and with additional sequences in U2 snRNA. Only at this stage can the first step of splicing occur, which yields the C complex. The C complex then performs the second step of splicing, the matured mRNA is released and the U snRNPs are recycled to participate in additional rounds of splicing. Some metazoan species contain a second so- called minor spliceosome, in which the U snRNPs U1, U2, and U4/U6 are functionally replaced by U11, U12, and U4atac/U6atac [17]. This spliceosome is dedicated to the removal of a separate class of introns, which are delineated by AU at the 5´ss and AC at the 3´ss, respectively. The reaction mechanism of splicing of this class of introns is however identical to that described for the major spliceosome above.

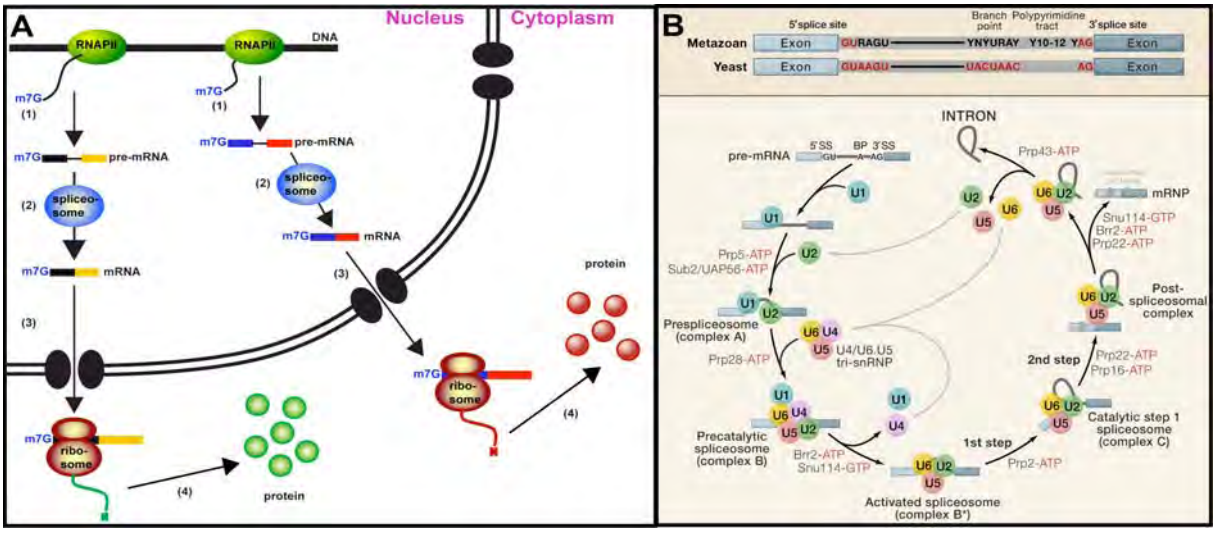


Figure 2: A) Eukaryotic gene expression. Gene expression in eukaryotes starts with the transcription of DNA-encoded genes by RNA polymerase II (RNAPII, step 1). This gives rise to pre-mRNA, which contains non-coding segments (introns) and protein-coding sequences (exons). Pre-mRNA splicing catalyzed by the spliceosome removes introns and leads to the formation of mRNA (step 2). After export of the mRNA to the cytoplasm (step 3), it serves as a template for the ribosome, which synthesizes proteins (step 4). **B) Pre-mRNA splicing.** In the upper panel yeast and metazoan sequence elements are depicted, which direct the removal of introns (grey) and allow the ligation of exons (blue). The lower panel shows the spliceosomal assembly and

disassembly cycle and the chemical steps in splicing. Also shown are the proteins responsible for the catalytic rearrangement steps during the spliceosomal cycle. Reproduced, with permission, from Elsevier Press [14].

3.3 Architecture of Spliceosomal U snRNPs

U snRNPs of the major and minor spliceosome share some common architectural principles. First, they are composed of one (U1, U2, U5) or two (U4/U6) small eponymous RNA(s), which is rich in uridylic acid. U snRNAs commonly contain an unusual trimethylguanosine (TMG) cap structure at their 5´-end and a conserved single-stranded sequence element flanked by two stem-loop structures termed the Sm site (as exemplified for the U1 snRNP in Figure 3A). U snRNPs are additionally characterized by a set of common factors termed Sm proteins, which bind to the Sm site to form the characteristic Sm core domain. Sm proteins do not exist as individual polypeptides, instead they are organized into the heterooligomers D1D2, D3B, and EFG (Figure 3A) [18]. Additionally, each individual U snRNP particle contains a set of specific proteins, which allows it to fulfill its designated role in the spliceosomal cycle (see previous chapter and Figure 2B).

Due to its relative biochemical simplicity the U1 snRNP is to date best characterized in terms of its architecture. However, biochemical and electron microscopic analyses indicate that principles derived from the U1 snRNP are also applicable to the other spliceosomal U snRNPs (Figure 3A) [19,20]. A major breakthrough in the understanding of the architecture of spliceosomal U snRNPs was achieved by the elucidation of the X-ray crystal structures of the two Sm protein heterodimers D1D2 and D3B [21]. These structures revealed that Sm proteins share a common fold, which is defined by an N-terminal α- helix, followed by a strongly bent 5-stranded β- sheet barrel (Figure 3B). Also, the structures showed that the individual Sm proteins associate with each other by hydrophobic interactions of the β4- strand of one Sm protein with the β5- strand of another Sm protein. Based upon these findings the authors put forth a model, in which the 7 Sm proteins form a ring in the Sm core domain [21]. The Sm site of U snRNAs was placed as a single strand in the central cavity of the doughnut-shaped Sm core domain based on size constraints and surface electrostatic charge of the inner cavity. This so-called heptameric ring model has since been corroborated by site-specific, biochemical crosslinking studies [22].

Single particle cryoelectron microscopy (Cryo EM; Figure 3C, left particle) and X-ray crystallography (Figure 3C, right particle) of the U1 snRNP provided additional insight into the architecture of U snRNPs [23,24]. Both structures have similar dimensions. The overall appearance of the U1 snRNP is similar to the face of Mickey Mouse; it consists of a ring-shaped main body and two large protuberances on the upper side of the main body. The heptameric ring model for the Sm core domain is unambiguously corroborated by both structures. The bases of the Sm site are visible as a cartwheel-like density in the central hole of the toroidal Sm core domain in the X-ray crystal structure [24]. Additional details include the finding that each Sm protein decodes one base of the Sm site and that a portion of the 70 kD protein wraps once around the Sm core domain to contact the U1C protein. The latter was suggested previously by biochemical studies [25].

While the structures elucidated to date no doubt have extended our understanding of the architecture of U snRNP particles in unprecedented ways, additional structures at even higher resolution are required to enhance our view. Additionally, the structures of other spliceosomal U snRNPs and distinct states of the spliceosome as recently elucidated should yield mechanistic insights into the function of this dynamic machinery [26-31]. However, the biochemical complexity of these particles and the difficulty to prepare large homogenous amounts of these latter particles makes Cryo EM, along with structural mapping of the subunits the method of choice [32].

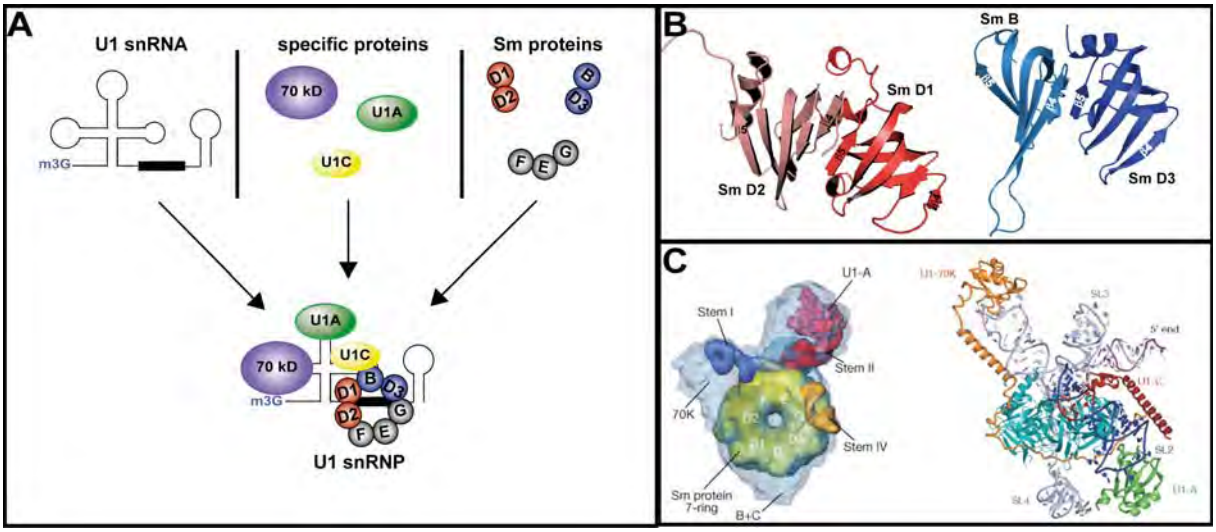


Figure 3: A) The basic architecture of U snRNP particles. The basic architecture of U snRNP particles is presented, exemplified by the U1 snRNP. Each U snRNP is composed of an eponymous small nuclear RNA (U1 snRNA). Each particle also contains a set of seven proteins common to all U snRNPs (Sm proteins). Sm proteins

do not exist as individual polypeptides; instead they are pre-organized into heterooligomers composed of D1D2, D3B and EFG. Additionally, each U snRNP particle contains a set of unique proteins termed specific proteins, which helps it perform its designated role in the spliceosomal cycle (70 kD, U1A and U1C). **B) Sm protein structure.** Shown are the X-ray crystal structures of the two Sm protein heterodimers D1D2 (PDB ID: 1B34) and D3B (PDB ID: 1D3B). Each Sm protein adopts a so-called Sm fold, which consists of an N-terminal α-helix followed by a strongly bent 5- stranded β- sheet. The interaction of the β4- strand of one Sm protein with the β5- strand of another by hydrophobic interactions mediates the interaction of Sm proteins amongst each other. The illustrations were prepared using Pymol (De Lano Scientific). **C) Structures of the U1 snRNP.** On the left side the Cryo EM structure of the human U1 snRNP at a resolution of 10 Å is shown [23], the right side depicts the same structure elucidated by X-ray crystallography at a resolution of 5.5 Å [24] Reproduced, with permission, from Nature Publishing Group.

3.4 The Cell Biology of U snRNP Biogenesis

Prior to their involvement in the spliceosomal cycle the U snRNPs of both the major and minor spliceosome undergo a segmented biogenesis pathway. This involves two subcellular transport events and several steps, where the U snRNA is enzymatically modified and proteins are added to complete the U snRNP (Figure 4A). Upon transcription in the nucleus by RNA polymerase II, the U snRNAs U1, U2, U4, U4atac, U5, U11 and U12 co-transcriptionally acquire a monomethyl cap structure (m^7G; Figure 4A, step 1) [33-35]. This cap structure directs the transient export of newly transcribed U snRNA into the cytoplasm (Figure 4A, step 2) [36]. For cytoplasmic export of U snRNA to occur, the cap-binding complex (CBC), which consists of two polypeptides the cap binding proteins 20 kD in size (CBP20) and 80 kD in size (CBP80), is recruited to the 5´end of U snRNA (Figure 4B) [37]. This heterodimeric complex then binds to the phosphorylated adaptor of export protein (PHAX), which recruits the export receptor CRM1 bound to the small GTPase Ran in its GTP bound form (Figure 4B) [38,39]. These associations give rise to the U snRNA export complex. After translocation through the nuclear pore complex to the cytoplasm, the GTP nucleotide bound to Ran undergoes hydrolysis to GDP, which reduces its affinity for CRM1 [38,40]. Upon dephosphorylation of PHAX, the export complex dissociates and the cargo U snRNA is released into the cytoplasm [39].

In the cytoplasm, the newly exported U snRNAs encounter Sm proteins, which are translated and believed to be stored in the cytoplasm [41]. This cytoplasmic storage pool of Sm proteins appears to include a RNA-free precursor with a sedimentation coefficient of 6S [42]. In an ordered pathway, the Sm proteins next assemble with the Sm site of U snRNAs (Figure 4A,

step3) [43]. Biochemical studies using purified Sm proteins and U snRNAs have elucidated some principles regarding this step of U snRNP biogenesis [18]. Sm proteins do not exist as individual polypeptides; instead they are organized into heterooligomeric units composed of D1D2, D3B, and EFG (Figure 3A and 4C). Neither can the individual Sm protein oligomers associate with each other in the absence of RNA, nor is any oligomer alone capable of association with U snRNA. The cooperative joining of D1D2, EFG, and U snRNA gives rise to the Sm subcore, the first intermediate in U snRNP formation. Subsequently, this intermediate is matured into the Sm core domain upon addition of the D3B Sm protein heterodimer (Figure 4C) [18].

In a next step, the m^7G cap moiety of the U snRNA is converted to the trimethylguanosine cap (TMG; Figure 4A, step 4) [44]. The enzyme required for this conversion has recently been identified and named trimethylguanosine synthase 1 (Tgs1) [45]. This protein is thought to have a binding site on the assembled core domain and thus TMG formation can only occur after assembly of the Sm core domain [44,46]. The TMG constitutes one part of the bipartite nuclear localization signal (NLS), which mediates the re- import of the assembled U snRNP into the nucleus (Figure 4A, step 5) [47,48]. The other part of the NLS is located within the Sm core domain itself [49]. Nuclear transport requires receptor proteins, which mediate the crossing of the nuclear membrane [50]. In the case of U snRNPs some receptors are known (Figure 4D). The m_3G cap is recognized by a receptor called snurportin 1 (Spn1), which associates with importin β (Impβ) [51]. Neither the identity of the NLS on the Sm core domain, nor the receptor that recognizes it is known to date. However, also this receptor is thought to bind to Impβ [49]. After translocation of the U snRNP to the nucleus, the high concentrations of Ran (GTP) cause the dissociation of the import complexes and the U snRNP is released (Figure 4D). After the addition of U snRNP specific proteins and additional spliceosomal proteins (Figure 4A, step6), the assembled U snRNPs are active in splicing (Figure 4A, step 8).

The biogenesis of U6 and U6atac differs from the process described above. The genes for these U snRNAs are transcribed by RNA polymerase III and the transcripts do not leave the nucleus [52]. Furthermore, these U snRNAs do not bind directly to Sm proteins common to the other U snRNP species. Instead, a set of closely related proteins termed Like Sm (LSm) 2-8 binds to a polyuridine stretch located at the 3´-end of these RNAs [53].

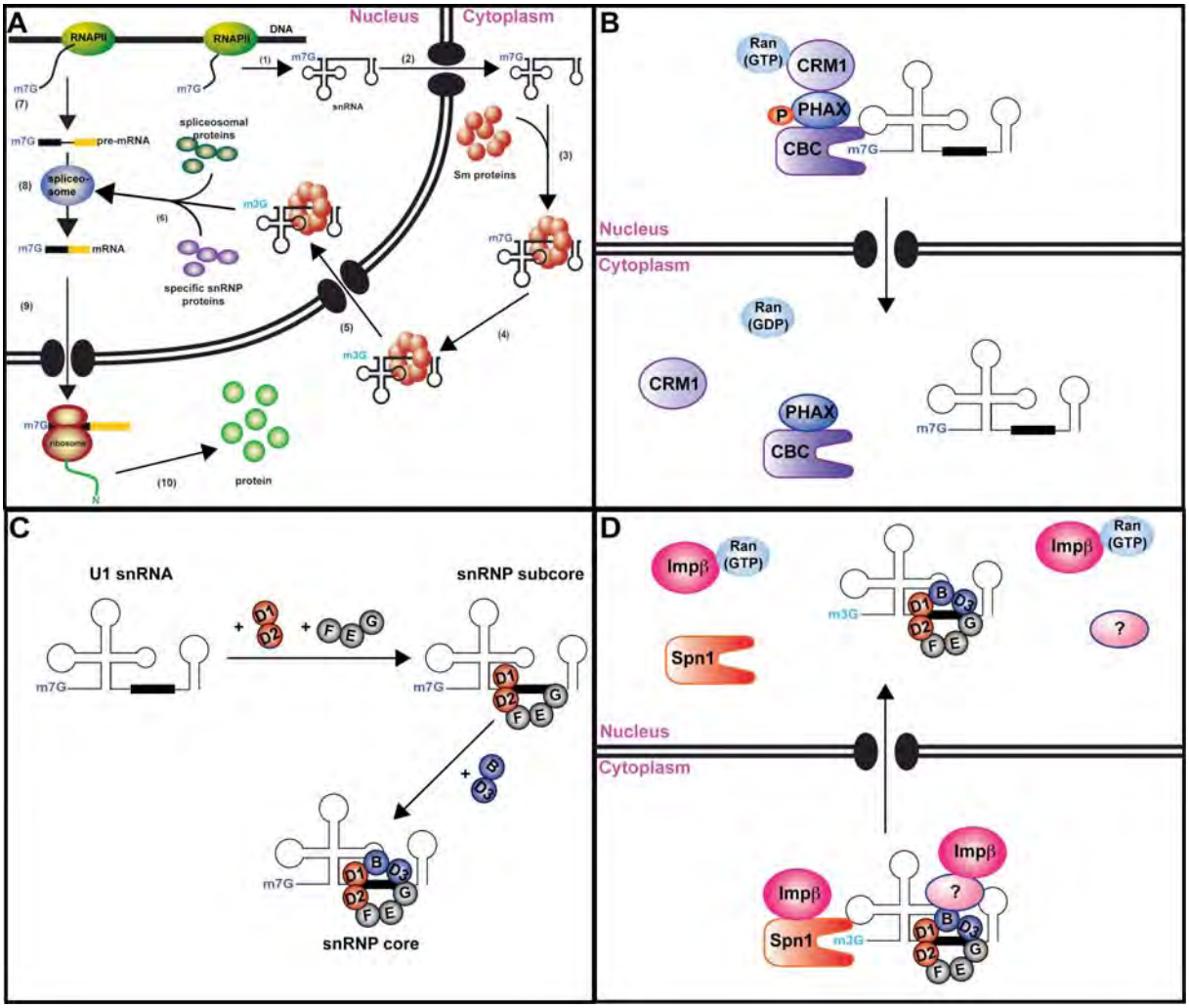


Figure 3: A) U snRNP biogenesis impacts eukaryotic gene expression. RNA polymerase II (RNAPII) transcribes U snRNA genes (1). The newly synthesized U snRNAs are then transiently exported to the cytoplasm (2). In the cytoplasm the seven Sm proteins assemble onto the Sm site to form the Sm core domain (3). Formation of the Sm core domain triggers the conversion of the m^7G cap moiety to the m_3G trimethylguanosine cap structure common to spliceosomal U snRNPs (4). Together m_3G cap and Sm core domain constitute a bipartite nuclear localization signal, which retranslocates the assembled U snRNP into the nucleus (5). Upon the addition of specific proteins and additional spliceosomal proteins (6), the U snRNPs participate in the splicing reaction (8). This step removes intervening segments of RNAPII transcribed genes (7). After splicing mature mRNAs are exported to the cytoplasm (9) and translated into protein (10). **B) U snRNA export from the nucleus.** The heterodimeric cap binding complex (CBC) binds the m^7G cap of U snRNAs. The CBC directly associates with the phophorylated adaptor of export protein (PHAX), which is phosphorylated in the nucleus. PHAX recruits the nuclear export receptor CRM1, which in turn binds to Ran (GTP). Upon export to the cytoplasm the hydrolysis of GTP to GDP in Ran is triggered, which causes its dissociation from CRM1. Upon dephosphorylation of PHAX, the export complex disassembles and the U snRNA is released into the cytoplasm. **C) Formation of the Sm core domain follows an ordered pathway.** No Sm protein heterooligomer is capable of binding to RNA, nor are they able to associate with each other in the absence of RNA. Sm core domain formation is induced by the initial, cooperative binding of D1D2 and EFG to U snRNA, giving rise to the snRNP

subcore domain. This intermediate is then matured into the snRNP core domain upon the addition of D3B. **D) Nuclear import of the assembled Sm core domain.** The import receptor snurportin 1 (Spn1) recognizes the m_3G cap and binds to importin β (Impβ). Neither the identity of the Sm core domain nuclear localization signal (NLS), nor is the receptor that recognizes it known to date. However, it is believed to bind to Impβ. Upon import of the U snRNP into the nucleus Ran (GTP) binds to Impβ to release the U snRNP into the nucleoplasm.

3.5 U snRNP Assembly *In Vivo* is an Active, Factor-Mediated Process

The biochemical studies of U snRNP assembly described in the previous chapter have shown that *in vitro* theses particles are capable of forming spontaneously [18]. Furthermore, these experiments elucidated that the structural information for the formation of the Sm core domain lies within the Sm protein and U snRNA constituents. However as detailed in chapter 3.1 of this thesis, the properties of cellular milieus introduce several impediments to self-assembly processes. These considerations have raised the question whether *in vivo* the assembly of U snRNPs occurs without the assistance of trans- acting factors.

First clues to the cellular assembly of spliceosomal U snRNPs came from the analysis of the Survival Motor Neuron (SMN) protein. Compound heterozygous mutation or deletions of this gene elicits the neuromuscular disorder Spinal Muscular Atrophy (SMA) in humans [54]. When immunoprecipitated from *Xenopus laevis* oocytes, the SMN protein was found to interact with snRNAs exclusively in the cytoplasm [55]. Along with the finding that the SMN protein interacts with Sm proteins, it was suggested to play a role in the assembly of U snRNPs [56]. Subsequent biochemical purification established that SMN acts in the context of a macromolecular unit, termed the SMN-complex with an approximate molecular weight of 1 Megadalton (MDa) [57,58]. Due to their mutual localization with SMN to specialized domains in the nucleus called Gems, the subunits of the SMN-complex are referred to as Gemins [58,59].

Definitive proof for the role of the SMN-complex in U snRNP assembly, came by the development of an *in vitro* assay for Sm core formation using *Xenopus laevis* egg extracts [60]. Addition of purified Sm proteins to these extracts strongly decreased their capacity to form U snRNPs, demonstrating for the first time that Sm core formation does not follow a strict self- assembly route *in vivo*. Immunodepletion of SMN totally abolished the U snRNP assembly activity of these extracts, although Sm proteins were still present after this treatment. The capacity to form Sm cores could be fully restored by the addition of an

immunopurified SMN-complex to extracts previously depleted of SMN. Furthermore, depletion of the endogenous levels of ATP in these extracts inhibited U snRNP assembly; this could be restored by the addition of exogenous ATP and an energy regenerating system [60]. These findings were expanded by the demonstration that the SMN-complex purified from cells, which abundantly contains Sm proteins, is both necessary and sufficient to assemble U snRNPs [61,62]. Thus, the function of the SMN-complex consists in the loading of Sm proteins onto the Sm site of U snRNAs to form the characteristic Sm core domain.

The histone pre-mRNA processing U7 snRNP contains a slightly divergent core domain, which consists of the Sm proteins D3B, EFG [63,64]. In addition this U snRNP has two additional LSm proteins LSm10 and 11, which replace the Sm proteins D1 and D2, respectively [65-67]. Notably, a specialized SMN-complex containing the LSm proteins 10 and 11 is involved in the assembly of the U7 snRNP [65]. Thus, the SMN-complex can be considered the key player in the cellular assembly of RNPs containing Sm and LSm proteins.

After each round of U snRNP assembly the SMN-complex has to be replenished with Sm proteins. To achieve this, it is greatly aided by yet another macromolecular complex, whose name- giving constituent is the protein arginine methyltransferase 5 (PRMT5) protein [68,69]. Additional subunits of this macromolecular entity are the WD- repeat protein WD45 (also termed MEP50) and the pICln protein. The latter binds newly translated Sm proteins prior to their interaction with the SMN-complex through their Sm fold domains, and recruits them to the PRMT5-WD45 heterodimer. In this state PRMT5 posttranslationally modifies designated arginine residues on the Sm proteins D1, D3, and B to symmetrical dimethylarginine. This modification is believed to enhance the interaction of these Sm proteins to the SMN-complex [70,71]. Immunopurification of large amounts of the PRMT5-complex from extracts revealed that it contained substoichiometric amounts of the SMN-complex, suggesting that the two complexes physically interact [61]. This complex was also found to be more efficient in Sm core domain formation, indicating that one role of the PRMT5 complex is to stimulate the assembly reaction. This finding suggests that Sm proteins are directly transferred from the PRMT5-complex onto the SMN-complex. The resulting higher order unit, the SMN-/ PRMT5 complex, is the entity that promotes U snRNP assembly *in vivo* (Figure 5) [57].

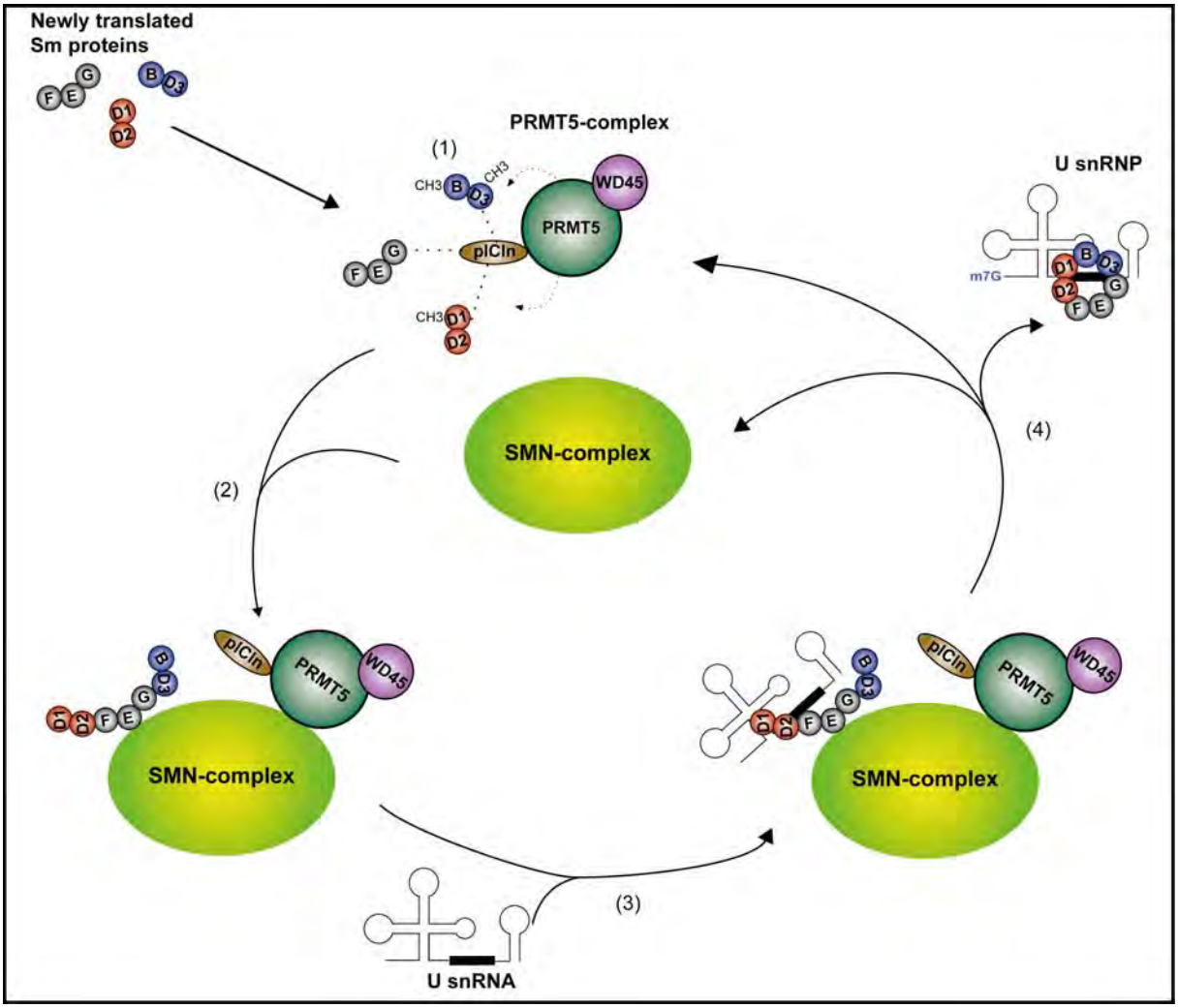


Figure 5: The assembly of U snRNPs is mediated by trans- acting factors *in vivo*. Newly translated Sm proteins, organized as the heterooligomeric units D1D2, D3B, and EFG are initially recruited to the PRMT5-complex via a direct association with the pICln subunit. At the PRMT5-complex the Sm proteins D1, D3, and B are methylated on designated arginine residues (step 1). PRMT5- and SMN-complexes join to form the SMN-/ PRMT5-complex and the Sm proteins are transferred onto the SMN-complex (step 2). Subsequently, U snRNA is recruited to the SMN-complex with bound Sm proteins (step 3). The SMN-complex then promotes the assembly of the Sm core domain. After this reaction the assembled U snRNP is released and PRMT5- and SMN-complexes dissociate to engage in a new round of assembly (step 4).

3.6 References

1. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology**. *Nature* 1999, **402**:C47-52.
2. Alberts B: **The cell as a collection of protein machines: preparing the next generation of molecular biologists**. *Cell* 1998, **92**:291-294.
3. Kornberg RD: **Structure of chromatin**. *Annu Rev Biochem* 1977, **46**:931-954.
4. Nierhaus KH, Dohme F: **Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli**. *Proc Natl Acad Sci U S A* 1974, **71**:4713-4717.
5. Nomura M, Traub P, Bechmann H: **Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins**. *Nature* 1968, **219**:793-799.
6. Kushner DJ: **Self-assembly of biological structures**. *Bacteriol Rev* 1969, **33**:302-345.
7. Frydman J: **Folding of newly translated proteins in vivo: the role of molecular chaperones**. *Annu Rev Biochem* 2001, **70**:603-647.
8. Berg OG, von Hippel PH: **Diffusion-controlled macromolecular interactions**. *Annu Rev Biophys Biophys Chem* 1985, **14**:131-160.
9. Ellis RJ: **Macromolecular crowding: obvious but underappreciated**. *Trends Biochem Sci* 2001, **26**:597-604.
10. Zimmerman SB, Minton AP: **Macromolecular crowding: biochemical, biophysical, and physiological consequences**. *Annu Rev Biophys Biomol Struct* 1993, **22**:27-65.
11. Le Tallec B, Barrault MB, Courbeyrette R, Guerois R, Marsolier-Kergoat MC, Peyroche A: **20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals**. *Mol Cell* 2007, **27**:660-674.
12. Laskey RA, Honda BM, Mills AD, Finch JT: **Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA**. *Nature* 1978, **275**:416-420.
13. Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M: **Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco**. *Cell* 2007, **129**:1189-1200.
14. Wahl MC, Will CL, Luhrmann R: **The spliceosome: design principles of a dynamic RNP machine**. *Cell* 2009, **136**:701-718.
15. Cech TR: **Ribozymes, the first 20 years**. *Biochem Soc Trans* 2002, **30**:1162-1166.
16. Behrens SE, Luhrmann R: **Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells**. *Genes Dev.* 1991, **5**:1439-1452.

17. Patel AA, Steitz JA: **Splicing double: insights from the second spliceosome**. *Nat Rev Mol Cell Biol* 2003, **4**:960-970.
18. Raker VA, Plessel G, Luhrmann R: **The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro**. *EMBO J.* 1996, **15**:2256-2269.
19. Kastner B, Bach M, Luhrmann R: **Electron microscopy of snRNPs U2, U4/6 and U5: evidence for a common structure-determining principle in the major UsnRNP family**. *Mol.Biol.Rep.* 1990, **14**:171.
20. Branlant C, Krol A, Ebel JP, Lazar E, Haendler B, Jacob M: **U2 RNA shares a structural domain with U1, U4, and U5 RNAs**. *EMBO J* 1982, **1**:1259-1265.
21. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, Nagai K: **Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs**. *Cell* 1999, **96**:375-387.
22. Urlaub H, Raker VA, Kostka S, Lührmann R: **Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure**. *EMBO J* 2001, **20**:187-196.
23. Stark H, Dube P, Lührmann R, Kastner B: **Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle**. *Nature* 2001, **409**:539-542.
24. Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K: **Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution**. *Nature* 2009, **458**:475-480.
25. Nelissen RL, Will CL, van Venrooij WJ, Luhrmann R: **The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins**. *EMBO J* 1994, **13**:4113-4125.
26. Boehringer D, Makarov EM, Sander B, Makarova OV, Kastner B, Luhrmann R, Stark H: **Three-dimensional structure of a pre-catalytic human spliceosomal complex B**. *Nat Struct Mol Biol* 2004, **11**:463-468.
27. Golas MM, Sander B, Will CL, Luhrmann R, Stark H: **Molecular architecture of the multiprotein splicing factor SF3b**. *Science* 2003, **300**:980-984.
28. Golas MM, Sander B, Will CL, Luhrmann R, Stark H: **Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy**. *Mol Cell* 2005, **17**:869-883.
29. Jurica MS, Sousa D, Moore MJ, Grigorieff N: **Three-dimensional structure of C complex spliceosomes by electron microscopy**. *Nat Struct Mol Biol* 2004, **11**:265-269.

30. Behzadnia N, Golas MM, Hartmuth K, Sander B, Kastner B, Deckert J, Dube P, Will CL, Urlaub H, Stark H, et al.: **Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes**. *EMBO J* 2007, **26**:1737-1748.
31. Sander B, Golas MM, Makarov EM, Brahms H, Kastner B, Luhrmann R, Stark H: **Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy**. *Mol Cell* 2006, **24**:267-278.
32. Luhrmann R, Stark H: **Structural mapping of spliceosomes by electron microscopy**. *Curr Opin Struct Biol* 2009, **19**:96-102.
33. Mattaj IW, Lienhard S, Jiricny J, De Robertis EM: **An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes**. *Nature* 1985, **316**:163-167.
34. Mattaj IW, Zeller R: **Xenopus laevis U2 snRNA genes: tandemly repeated transcription units sharing 5' and 3' flanking homology with other RNA polymerase II transcribed genes**. *EMBO J.* 1983, **2**:1883-1891.
35. Zeller R, Carri MT, Mattaj IW, De Robertis EM: **Xenopus laevis U1 snRNA genes: characterisation of transcriptionally active genes reveals major and minor repeated gene families**. *EMBO J.* 1984, **3**:1075-1081.
36. Hamm J, Mattaj IW: **Monomethylated cap structures facilitate RNA export from the nucleus**. *Cell* 1990, **63**:109-118.
37. Izaurralde E, Lewis J, Gamberi C, Jarmolowski A, McGuigan C, Mattaj IW: **A cap-binding protein complex mediating U snRNA export**. *Nature* 1995, **376**:709-712.
38. Fornerod M, Ohno M, Yoshida M, Mattaj IW: **CRM1 is an export receptor for leucine-rich nuclear export signals**. *Cell* 1997, **90**:1051-1060.
39. Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW: **PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation**. *Cell* 2000, **101**:187-198.
40. Izaurralde E, Kutay U, von Kobbe C, Mattaj IW, Gorlich D: **The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus**. *EMBO J.* 1997, **16**:6535-6547.
41. Zeller R, Nyffenegger T, De Robertis EM: **Nucleocytoplasmic distribution of snRNPs and stockpiled snRNA-binding proteins during oogenesis and early development in Xenopus laevis**. *Cell* 1983, **32**:425-434.

42. Fisher DE, Conner GE, Reeves WH, Wisniewolski R, Blobel G: **Small nuclear ribonucleoprotein particle assembly in vivo: demonstration of a 6S RNA-free core precursor and posttranslational modification**. *Cell* 1985, **42**:751-758.

43. Hamm J, Kazmaier M, Mattaj IW: **In vitro assembly of U1 snRNPs**. *EMBO J.* 1987, **6**:3479-3485.

44. Mattaj IW: **Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding**. *Cell* 1986, **46**:905-911.

45. Mouaikel J, Narayanan U, Verheggen C, Matera AG, Bertrand E, Tazi J, Bordonne R: **Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron**. *EMBO Rep* 2003, **4**:616-622.

46. Plessel G, Fischer U, Lührmann R: **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.* 1994, **14**:4160-4172.

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

48. Hamm J, Darzynkiewicz E, Tahara SM, Mattaj IW: **The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal**. *Cell* 1990, **62**:569-577.

49. Fischer U, Sumpter V, Sekine M, Satoh T, Luhrmann R: **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.* 1993, **12**:573-583.

50. Gorlich D, Kutay U: **Transport between the cell nucleus and the cytoplasm**. *Annu Rev Cell Dev Biol* 1999, **15**:607-660.

51. Huber J, Cronshagen U, Kadokura M, Marshallsay C, Wada T, Sekine M, Luhrmann R: **Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure**. *EMBO J* 1998, **17**:4114-4126.

52. Fury MG, Zieve GW: **U6 snRNA maturation and stability**. *Exp.Cell Res.* 1996, **228**:160-163.

53. Achsel T, Brahms H, Kastner B, Bachi A, Wilm M, Luhrmann R: **A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'- end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro**. *EMBO J.* 1999, **18**:5789-5802.

54. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, et al.: **Identification and characterization of a spinal muscular atrophy- determining gene**. *Cell* 1995, **80**:155-165.

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

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

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

58. Paushkin S, Gubitz AK, Massenet S, Dreyfuss G: **The SMN complex, an assemblyosome of ribonucleoproteins**. *Curr Opin Cell Biol* 2002, **14**:305-312.

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

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

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

62. Pellizzoni L, Yong J, Dreyfuss G: **Essential role for the SMN complex in the specificity of snRNP assembly**. *Science* 2002, **298**:1775-1779.

63. Mowry KL, Steitz JA: **Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone premessenger RNA's**. *Science* 1987, **238**:1682-1687.

64. Muller B, Schumperli D: **The U7 snRNP and the hairpin binding protein: Key players in histone mRNA metabolism**. *Semin Cell Dev Biol* 1997, **8**:567-576.

65. Pillai RS, Grimmler M, Meister G, Will CL, Luhrmann R, Fischer U, Schumperli D: **Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing**. *Genes Dev* 2003, **17**:2321-2333.

66. Pillai RS, Will CL, Luhrmann R, Schumperli D, Muller B: **Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein**. *EMBO J* 2001, **20**:5470-5479.

67. Schumperli D, Pillai RS: **The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein**. *Cell Mol Life Sci* 2004, **61**:2560-2570.

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

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

70. Brahms H, Meheus L, de Brabandere V, Fischer U, Luhrmann R: **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* 2001, **7**:1531-1542.

71. Friesen WJ, Dreyfuss G: **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* 2000, **275**:26370-26375.

4. Goals of this Thesis

How cells assemble molecular machines is an intensely studied aspect of scientific research today. The principles for the formation of proteinaceous macromolecular complexes, such as the proteosome, or DNA-protein complexes, like the nucleosome, are well understood today. In striking contrast, it is largely unknown how cells assemble RNA-protein (RNP) complexes in the crowded cellular environment. Therefore, the major goal of this thesis was the elucidation of the mechanisms underlying cellular RNP formation. The uridylic-rich small nuclear RNA-protein particles (U snRNPs) of the spliceosome served as a model system to address this question. Previous studies had indicated that these particles assemble in the cytoplasm in a complex pathway, which involves a large number of trans-acting factors. These assembly factors are organized into two units, termed SMN- and PRMT5-complexes. While the involvement of these two macromolecular complexes had been well established, their mode of action was largely unknown.

To elucidate the reaction mechanism of cellular U snRNP assembly 4 strategies were employed in this thesis:

1) Initially, studies were conducted to establish a full inventory of the subunits of the assembly machinery (Chapter 5.1 in the results section).

2) With this information in hand, the second report addressed the question how the individual subunits are organized to form the SMN-complex (Chapter 5.2 in the results section).

3) The previous two studies formed the basis to investigate mechanistic aspects of the cellular U snRNP reaction. In a first step, an early phase in the SMN-assisted assembly reaction was studied. In particular, the stimulatory role of the PRMT5-complex was emphasized in the context of U7 snRNP formation (first part of Chapter 5.3 in the results section). In a second step the mechanistic basis of the cellular U snRNP assembly reaction was examined. For this, key reaction intermediates were elucidated by a combination of biochemistry and structural studies (second part of Chapter 5.3 in the results section).

4) Lastly, even though the human SMN-complex is composed of several subunits it is unknown whether all subunits of this entity are essential for U snRNP assembly. Chapter 5.4 of this thesis deals with the question, which subunits of the SMN-complex make up the most ancestral form of this entity in evolution and, thus, a minimal unit capable of promoting U snRNP formation.

5. Results

5.1 Taking an Inventory of the Subunits of the Human SMN-Complex

Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization

Grimmler M, Otter S, Peter C, Müller F, Chari A, Fischer U: **Hum Mol Genet 2005**, 14:3099-3111.

Thesis author's contribution:

Conception:	10 %
Experimental contribution:	10 %
Formulation of results:	30 %

Human Molecular Genetics, 2005, Vol. 14, No. 20 **3099–3111**
doi:10.1093/hmg/ddi343
Advance Access published on September 13, 2005

Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization

Matthias Grimmler, Simon Otter, Christoph Peter, Felicitas Müller, Ashwin Chari and Utz Fischer*

Department of Biochemistry, Theodor Boveri Institute, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

Received June 21, 2005; Revised August 3, 2005; Accepted September 7, 2005

Spliceosomal Uridine-rich small ribonucleo protein (U snRNP) assembly is an active process mediated by the macromolecular survival motor neuron (SMN) complex. This complex contains the SMN protein and six additional proteins, named Gemin2–7, according to their localization to nuclear structures termed gems. Here, we provide biochemical evidence for the existence of another, yet atypical, SMN complex component, termed unr-interacting protein (unrip). This abundant factor has been previously shown to form a complex with unr, a protein implicated in cap-independent translation of cellular and viral mRNA. We show that unrip is integrated into a complex with unr or with the SMN complex *in vivo* in a mutually exclusive manner. In the latter case, unrip is recruited to the active SMN complex via a stable interaction with Gemin7. However, unlike SMN and Gemins, unrip localizes predominantly to the cytoplasm and is absent from gems/Cajal bodies. Interestingly, RNAi-induced reduction of unrip protein levels leads to enhanced accumulation of SMN in the nucleus as evident by the increased formation of nuclear gems/Cajal bodies. Our data identify unrip as the first component of the U snRNP assembly machinery that associates with the SMN complex in a compartment-specific way. We speculate that unrip plays a crucial role in the intracellular distribution of the SMN complex.

INTRODUCTION

The spliceosomal Uridine-rich small nuclear ribonucleoprotein particles (U snRNPs) are essential components of the spliceosome, the macromolecular machinery that assembles on introns of pre-mRNAs and catalyses the splicing reaction. Each U snRNP consists of one (U1, U2, U5) or two (U4 and U6) small nuclear RNAs (snRNAs) associated with a set of the seven Sm proteins B/B′, D1, D2, D3, E, F and G, and a large number of particle-specific factors (recently reviewed in 1). U snRNPs form *in vivo* in a highly ordered and regulated pathway that is initiated by the post-transcriptional export of the U snRNA from the nucleus to the cytoplasm. Subsequently, the seven Sm proteins bind to the Sm site, a single-stranded sequence motif found in the snRNAs U1, U2, U4 and U5, leading to the formation of the ring-shaped Sm core domain (2,3). The Sm core domain provides a binding site for a methyltransferase (termed Tgs1p/PIMT) that catalyses the formation of the m2,2,7-trimethylguanosine cap (m_3G-cap) of the U snRNA (4,5). Sm core domain and the m_3G-cap both act as nuclear localization signals that target the assembled particles to the nucleus (6–8).

Although binding of Sm proteins onto U snRNAs can occur spontaneously *in vitro* when isolated Sm proteins are incubated with U snRNA, recent studies have indicated that this process requires ATP and the assistance of a large number of trans-acting factors (9–12). These factors are organized in two functional units, termed survival motor neuron (SMN) and PRMT5 complexes that mediate distinct functions in the assembly pathway. In the cytoplasm, the Sm proteins first bind to the PRMT5 complex, which contains the arginine-methyltransferase PRMT5. This enzyme subsequently modifies arginines in the RG-rich tails of the Sm proteins B/B′, D1 and D3 to symmetric dimethylarginines (10,13,14). Arginine methylation increases the affinity of Sm proteins for SMN, thus allowing their transfer onto the SMN

*To whom correspondence should be addressed. Tel: +49 9318884029; Fax: +49 9318884028; Email: utz.fischer@biozentrum.uni-wuerzburg.de

© *The Author 2005. Published by Oxford University Press. All rights reserved.*
For Permissions, please email: journals.permissions@oxfordjournals.org

complex (15–17). In a final step, Sm proteins are loaded onto the Sm site of the U snRNA, which leads to the formation of the Sm core domain (10–12). The finding that the cap-hypermethylase Tgs1p and U snRNP-transport factors associate with SMN further suggests that the SMN complex is also involved in these latter steps of U snRNP biogenesis (18–21). In fact, biochemical studies have shown that the SMN complex accompanies the mature U snRNPs to the nucleus (20). Post-translational modifications of SMN (and possibly also of other assembly factors) may regulate the different functions of the SMN complex in the biogenesis pathway of U snRNPs (22).

SMN interacts with a large number of proteins *in vivo* (23). However, when the SMN complex was isolated from cultivated HeLa cells under stringent conditions, only six stably associated components were found to be part of the assembly machinery. As these proteins localize not only to the cytoplasm but also to the nuclear gemini of coiled bodies (gems), they have been collectively named 'Gemins' (24). More recent studies showed that in many, but not all cells these nuclear structures are equivalent to the well-known nuclear Cajal bodies (25). Upon purification, we and others have identified an additional factor that is associated with SMN *in vivo*, termed unr-interacting protein (unrip) (10,20). Unrip is a WD-repeat protein that forms a complex with unr (upstream of N-ras), a factor implicated in cap-independent translational regulation of mRNAs (26,27).

Here, we show that unrip is the eighth major component of the assembly-active SMN complex but differs from other Gemins in that it is absent from the nucleus. Interestingly, RNA-interference (RNAi)-induced reduction of this factor leads to an increase in the number of gems/Cajal bodies, raising the hypothesis that it may modulate the subcellular distribution of the active SMN complex. Finally, we provide evidence that the interaction with the SMN complex occurs via binding to Gemin7. This interaction is mutually exclusive with binding of unrip to unr, thus allowing unrip to be part of two independent complexes *in vivo*. Our data raise the hypothesis that unrip interacts with the SMN complex in a reversible and compartment-specific manner, thereby regulating its localization and/or activity in U snRNP assembly.

RESULTS

A Sub-fraction of cellular unrip is specifically associated with the SMN complex

Although previous studies have established the interaction of unrip with unr *in vivo* (26,27), its proposed function within the SMN complex remained unclear. To resolve this issue, monospecific polyclonal antibodies directed against human unrip were generated (Fig. 1A). Using this antibody and the previously reported monoclonal antibody 7B10 directed against the N-terminus of human SMN (28), immunoprecipitation studies were performed with HeLa cytosolic extract. SDS–PAGE analysis of the anti-unrip precipitate revealed a large number of proteins, including two prominent ones with apparent molecular weights of ~ 40 and 100 kDa. Mass spectrometry (data not shown) and immunoblotting identified these proteins as unrip and unr, respectively (Fig. 1B, lanes 3, 4 and Fig. 1C lane 2). As expected, affinity purification with 7B10 recovered the SMN complex, consisting of SMN, Gemin2–7 and the Sm proteins (Fig. 1B, lane 4 and Fig. 1C, lane 3). Furthermore, unrip was readily detected in the SMN complex confirming earlier data (Fig. 1B and 1C) (10,20). More importantly, a comparison of both purifications revealed the entire SMN complex to be present in the unrip-immunoprecipitation (Fig. 1B, SMN complex components are indicated by dots and, in part, are also detected by western blot shown in Fig. 1C). None of these proteins were enriched in the pre-immune serum control purification, illustrating that these interactions are specific (Fig. 1C, lane 1). Thus, highly enriched SMN complex contains unrip and affinity-purified unrip contains the SMN complex.

Next, we asked whether unr, the known major interactor of unrip *in vivo* (26), was likewise a component of the SMN complex. Interestingly, in the 7B10 immunoprecipitate unr could be detected neither by silver staining (see Fig. 1B, for enlargement of this portion of gel) nor by immunoblotting with an antibody raised against human unr (Fig. 1C, lane 3). Thus, unrip, but not unr, interacts with the SMN complex in cytosolic extract.

To further strengthen the view that unrip is at least part of two complexes, HeLa cytosolic extract was fractionated by gel filtration chromatography and fractions were probed on western blots with different antibodies. As shown in Figure 1D, SMN and Gemin3 elute in a high molecular weight range (~660 kDa–1 MDa), which corresponds to the size of the SMN complex. Unrip, in contrast, elutes in a much more diffuse pattern ranging from ~150 kDa (the calculated mass of the unr/unrip heterodimer) to ~1 MDa (the mass of the SMN complex). Together, these data suggest that unrip not only binds to unr but also interacts specifically with the SMN complex.

Association of unrip with cytoplasmic SMN complex

A hallmark of SMN and its associated proteins is their subcellular localization to both the cytoplasm and the nuclear gems/Cajal bodies (24,29). Our finding that unrip is a major component of the cytoplasmic SMN complex led us to investigate whether it likewise localizes in gems/Cajal bodies as observed for the other complex components. This question was addressed by indirect immunofluorescence in cultivated HeLa cells. Endogenous unrip and SMN were detected using monospecific antibodies. In line with earlier findings, the antibody 7B10 detected SMN both in the cytoplasm and in the nuclear foci (Fig. 2A, c) (24,30). A partial co-localization of SMN and unrip in the cytoplasm was observed, consistent with the finding that a fraction of cellular unrip is part of the SMN complex (Fig. 2A, d, red and yellow merge). In striking contrast, unrip in the same cells could be detected neither in gems/Cajal bodies nor in other regions of the nucleus (Fig. 2A, b). Consequently, an overlay of both images revealed no co-localization in this compartment (Fig. 2A, d). To confirm these results by biochemical means, we prepared nuclear and cytosolic extract from HeLa cells and determined the amount of SMN and unrip in these fractions by western blotting. In extracts, which were normalized for the amount of SMN, unrip was detected almost exclusively (to >90%)

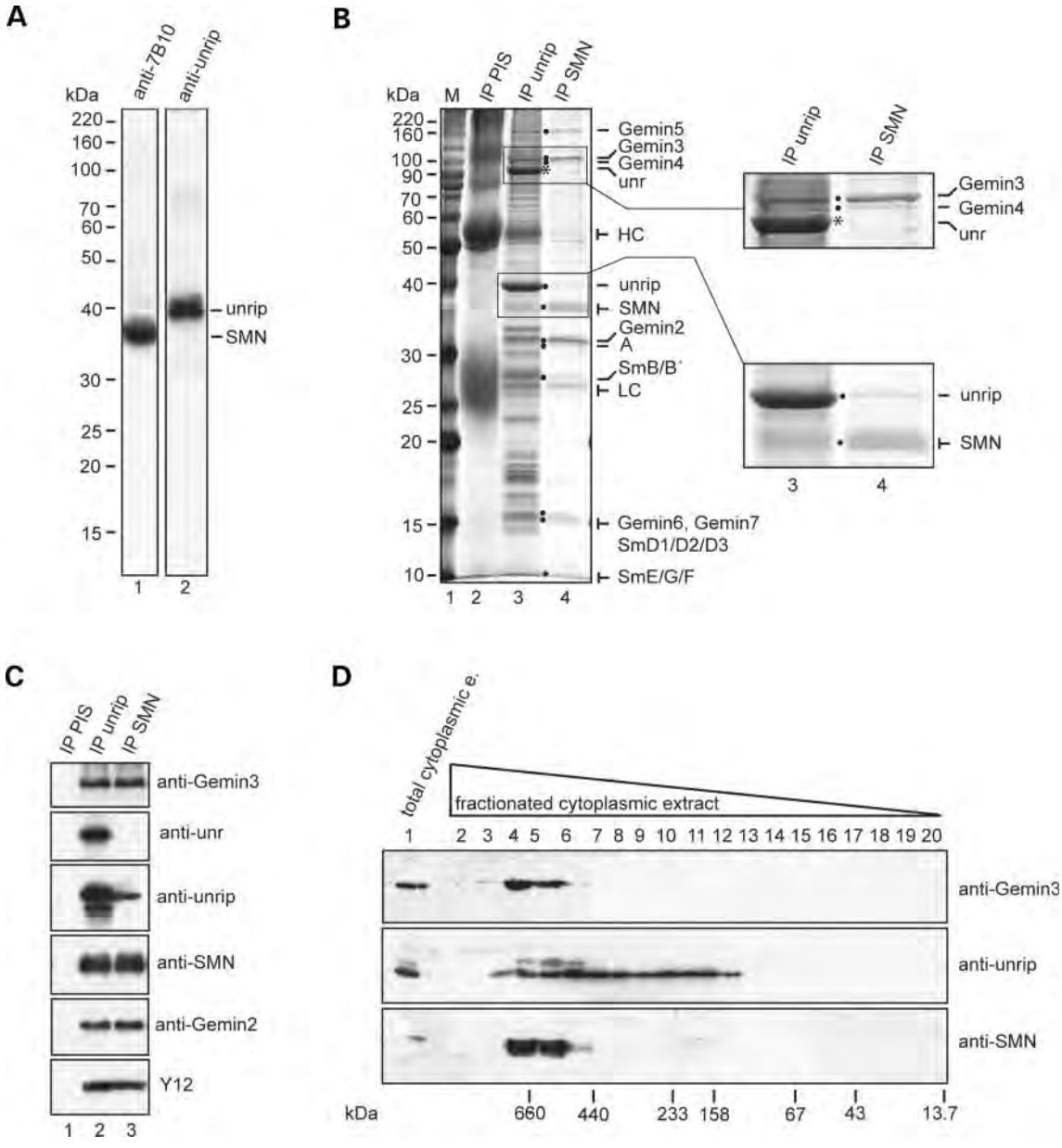


Figure 1. Unrip is part of the SMN complex *in vivo*. (**A**) Detection of unrip and SMN in HeLa cytosolic extract by western blotting using affinity-purified monoclonal anti-SMN antibody 7B10 (lane 1) and rabbit anti-unrip antibodies (lane 2). (**B**) Cytosolic extract from HeLa cells was immunoprecipitated with anti-unrip and anti-SMN antibody. Proteins were eluted from the beads, resolved by SDS–PAGE and analysed by silver staining (lanes 3 and 4). The unrip pre-immune serum was used for a control immunoprecipitation (lane 2). Lane 1 shows a molecular weight marker. Dots indicate known proteins of the SMN complex. An asterisk marks the unr protein, which can be detected only in the unrip immunoprecipitation. The right panel shows enlarged parts of the 40 and 100 kDa range of the gel. (**C**) The immunoprecipitated complexes shown in (A) were resolved by SDS–PAGE and analysed by western blotting with the indicated antibodies. (**D**) Unrip partially co-fractionates with the SMN complex. HeLa cytosolic extract (lane 1) was fractionated on a Superdex 200 pg gel filtration column. Fractions were separated by SDS–PAGE (lanes 2–20) and analysed by western blotting using antibodies directed against SMN, unrip and Gemin3. The indicated molecular masses were determined by an independent chromatography with a gel filtration size standard.

in the cytosolic fraction (Fig. 2B, lane 1). The small amount of unrip in the nuclear fraction most likely originates from cytoplasmic contaminations. SMN, in contrast, was detected in both compartments in similar quantities (compare lanes 1 and 2). Likewise, immunoprecipitation of SMN co-precipitated significant amounts of unrip only from the cytosolic but not from the nuclear extract (data not shown). Thus, unrip associates with the SMN complex only in the cytoplasm, whereas the nuclear complex contains little, if any, unrip.

Unrip-containing SMN complex is active in U snRNP assembly

The association of unrip with cytoplasmic but not with nuclear SMN complex was interesting in the light of earlier findings

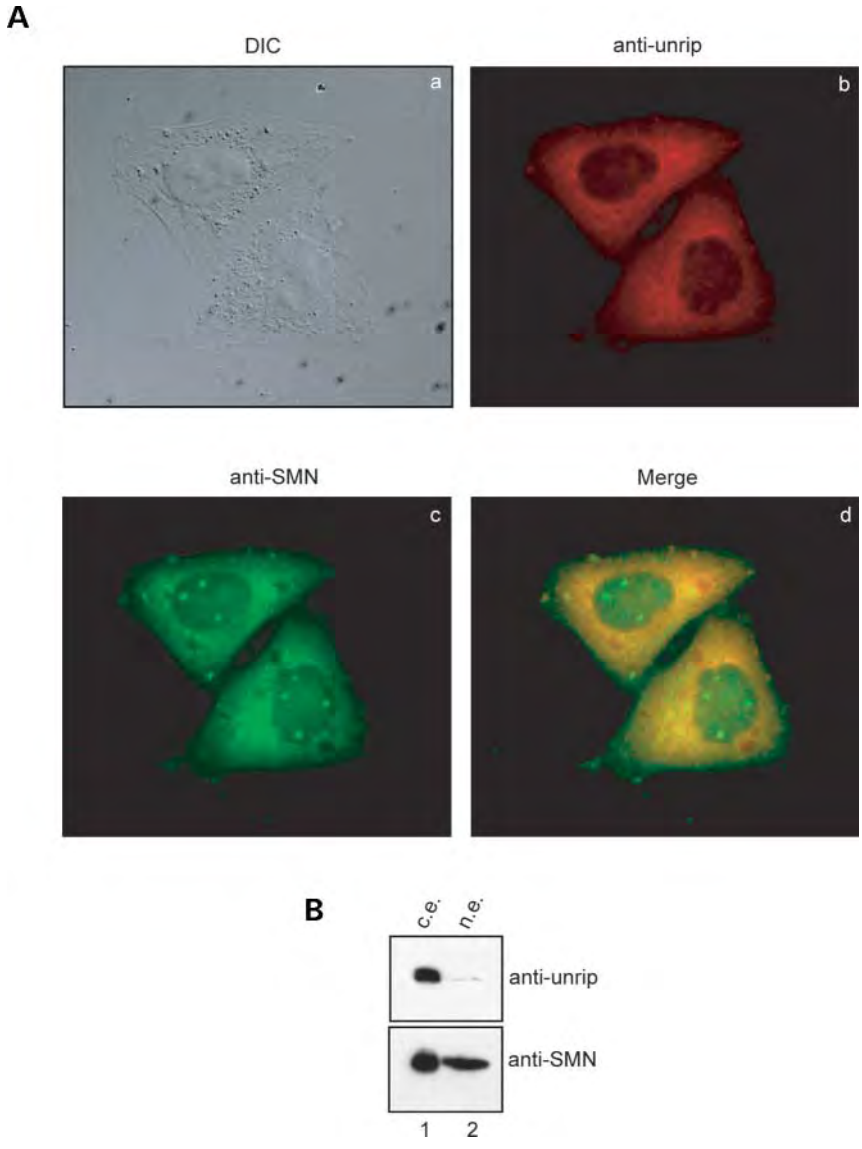


Figure 2. Unrip localizes predominantly to the cytoplasm but is absent from nuclear foci. (**A**) Localization of SMN and unrip in HeLa cells by immunofluorescence. Affinity-purified polyclonal rabbit anti-unrip (b) and monoclonal anti-SMN antibody 7B10 (c) were used as primary antibodies. Rhodamine-labeled anti-rabbit and fluorescein-labeled anti-mouse antibodies were used as secondary antibodies. Co-localization of proteins is shown by overlay of both images (yellow merge, d). (a) shows the analysed HeLa cells in differential interference contrast (DIC) mode. (**B**) HeLa cytosolic (lane 1) and nuclear extract (lane 2) were resolved by SDS–PAGE and analysed by immunoblotting using affinity-purified anti-unrip antibodies (upper panel) and anti-SMN antibody 7B10. The protein amount loaded in each lane was normalized for SMN.

that only the former is active in U snRNP assembly (22). Therefore, we tested whether unrip played a crucial role during the formation of U snRNPs using a previously reported *in vitro* assembly assay. In this assay, *in vitro* transcribed and ^{32}P-labeled U1 snRNA is incubated with cytosolic extract and assembly is subsequently monitored by native gel electrophoresis (10). We first asked whether cytosolic extracts lacking unrip could still promote formation of functional U1 snRNP. Indeed, assembly of the Sm core domain was strongly reduced in extract that had been immunodepleted of ~95% of endogenous unrip (Fig. 3A, lanes 2 and 3). This became evident by the impaired formation of a previously reported 'M-complex', which is super-shifted by the anti-Sm antibody Y12 (Fig. 3A, compare lanes 5 and 6). Formation of a complex marked by an asterisk was not affected. This complex forms SMN independently and consists of U1 snRNA and the U1-specific protein A only (10). Note that a non-specific complex is also retarded to the same position. In contrast, untreated or extract treated with control antibodies showed efficient formation of U1 snRNP (Fig. 3A, lanes 4 and 7). Thus, depletion of unrip from the extract severely impairs U snRNP assembly. We next tested by western blotting whether removal of unrip also had an impact on the abundance of other components of the assembly machinery. Indeed, along with the diminution of unrip also SMN, Gemin2 and 3 were co-depleted from the extract, whereas the level of Sm proteins remained

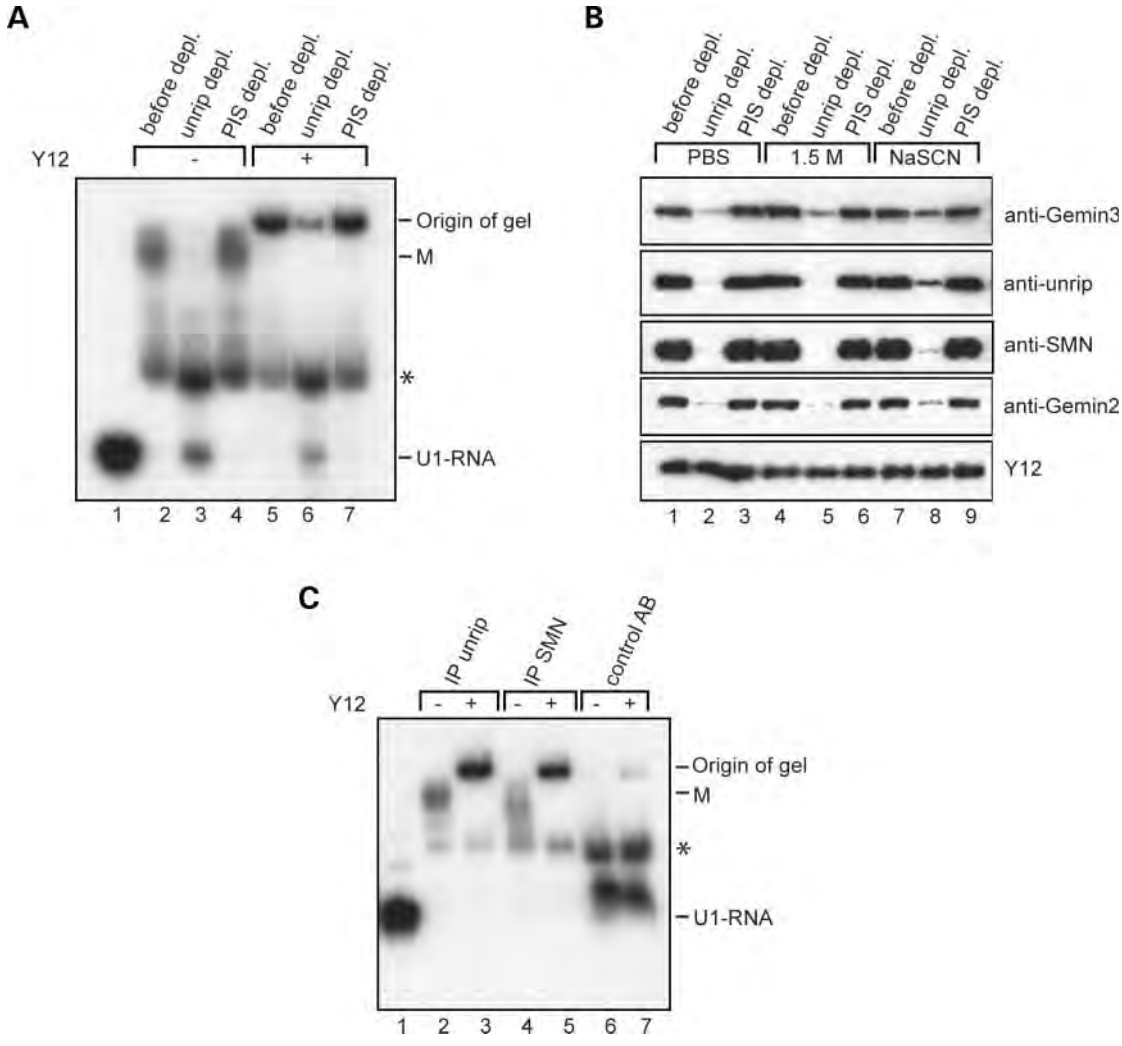


Figure 3. Unrip is associated with the SMN complex, active in U snRNP assembly. (**A**) Cytosolic extracts in PBS [lanes 1–3 in (B)] were incubated with ^{32}P-labeled U1 snRNA and analysed by native gel electrophoresis. Complex 'M' represents the assembled Sm core domain, the band indicated with an asterisk resembles a complex that is formed independent of the SMN complex (lanes 2–4). To verify Sm core assembly, Y12 antibody was added to assembly reactions prior to analysis by native gel electrophoresis. The M-complex is super-shifted to the origin of the gel, illustrating assembly of the Sm core domain (lanes 5–7). Lane 1 shows U1 snRNA in the absence of extract. (**B**) Immunodepletion of unrip under different salt conditions. HeLa cytosolic extracts are shown either before depletion (lanes 1, 4 and 7), after immunodepletion with unrip-antibody (lanes 2, 5 and 8), or after a control-depletion with unrip pre-immuneserum (lanes 3, 6 and 9). Depletions were performed in cytosolic extract containing either 130 mM NaCl (PBS, lanes 1–3), 1.5 M NaCl (1.5 M, lanes 4–6) or 300 mM KCl/500 mM NaSCN (NaSCN, lanes 7–9). Extracts were separated by SDS–PAGE and analysed by western blotting using antibodies directed against unrip, SMN, Gemin2, Gemin3 and the anti-Sm antibody Y12. (**C**) Cytosolic PBS extract from HeLa cells was immunoprecipitated with anti-unrip antiserum, anti-SMN antibody 7B10 or a control antibody. The immunoprecipitates were then incubated with ^{32}P-labeled U1 snRNA and subsequently analysed by native gel electrophoresis (lanes 2, 4 and 6). The specificity of Sm core formation was assessed by addition of the Sm antibody Y12 to the assembly reaction (lanes 3, 5, 7). Lane 1 shows the U1 snRNA in the absence of extract.

largely unchanged (Fig. 3B). All attempts to dissociate unrip from the rest of the SMN complex, which would allow for the specific depletion of merely unrip, failed even at very stringent conditions (1.5 M NaCl or 0.3 M KCl/ 0.5 M NaSCN were tested) (Fig. 3B compare lanes 1–3 with lanes 4–6 and 7–9). These data strongly suggested that unrip is part of the assembly machinery of U snRNPs. To verify this directly, SMN complex isolated either via immunoprecipitation with 7B10 antibody or via anti-unrip antibody was assessed for formation of the U1 snRNP *in vitro*. As shown in Figure 3C, both complexes efficiently generated U1 snRNP cores, as evident by the appearance of complex M (Fig. 3C, lanes 2–5), whereas a control purification with a non-related antibody failed to promote U1 snRNP formation (lanes 6 and 7). In conclusion, these data show that unrip is part of the SMN complex active in U snRNP assembly.

Because depletion of unrip by biochemical means simultaneously led to reduction of other essential assembly factors, the question whether unrip was necessary for assembly activity remained open. To circumvent SMN complex co-depletion, we used RNAi to selectively reduce unrip proteins levels. As shown in Figure 4A, silencing of unrip was nearly complete 72 h after transfection (lane 2), whereas minor amounts of the protein were expressed at later time

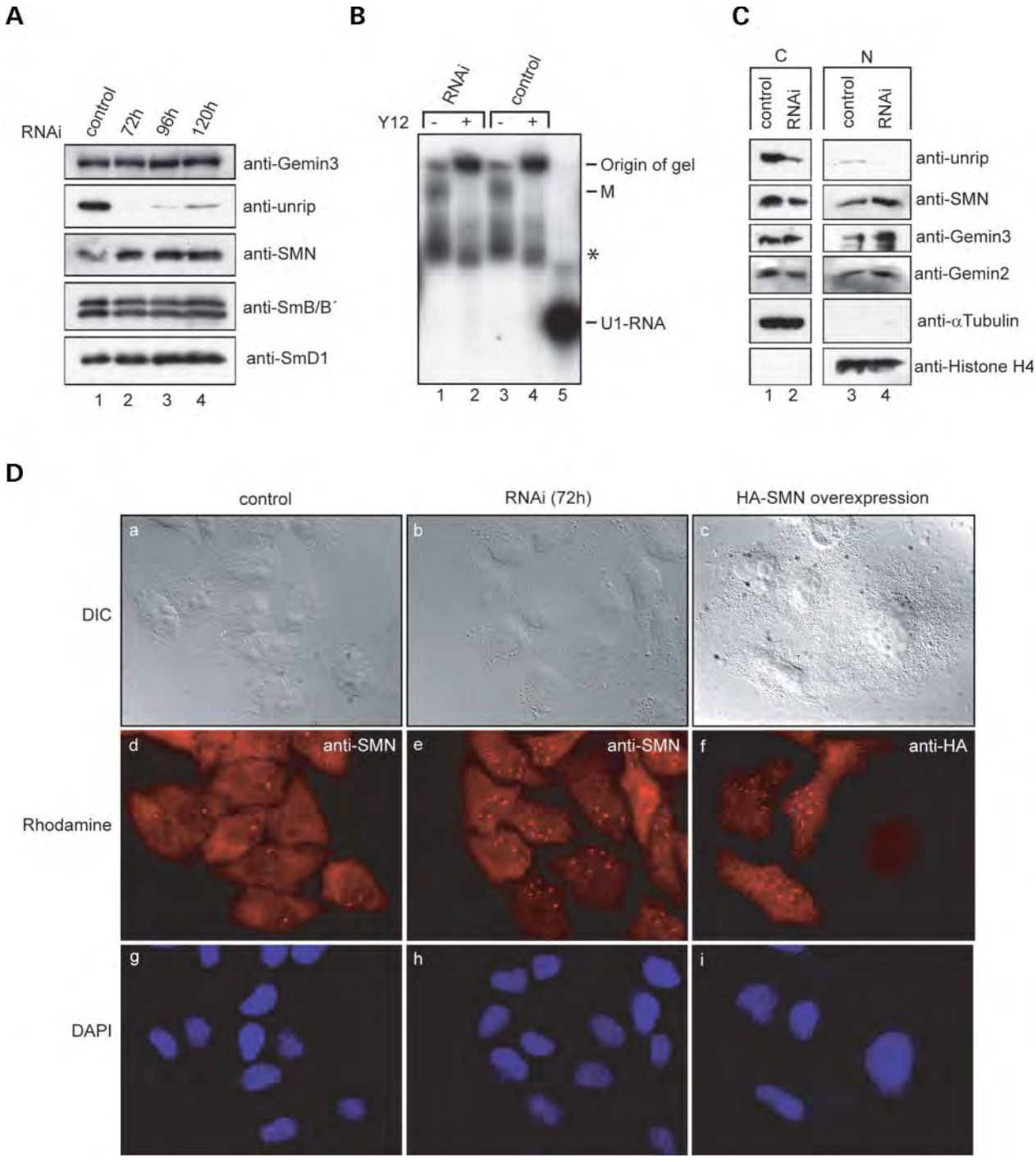


Figure 4. Silencing of unrip leads to an accumulation of the SMN complex in nuclear gems/Cajal bodies. (**A**) Reduction of unrip protein levels in HeLa cells by RNAi. Cells were transfected with siRNA directed against unrip for 72, 96 and 120 h. Whole cell extracts obtained from these cells (lanes 2, 3, and 4, respectively) and control cells (lane 1) were resolved by SDS–PAGE and analysed by immunoblotting with monospecific antibodies directed against SMN, Gemin3, unrip, Sm B/B' and Sm D1, respectively. (**B**) U snRNP assembly activity of extracts derived from unrip-silenced cells. Cytosolic extracts derived from cells treated with unrip siRNA for 72 h (lane 1) or control extract (lane 3) were assessed for their competence in U snRNP assembly by native gel electrophoresis, using ^{32}P-labeled U1 snRNA. Addition of monoclonal antibodies Y12 leads to a super-shift of complex M, verifying Sm core assembly (lanes 2 and 4). Lane 5 shows U1 snRNA in the absence of extract. (**C**) Silencing of unrip leads to an accumulation of SMN complex in the nucleus. HeLa cells were treated with siRNA directed against unrip as described earlier and separated into a cytosolic (C) and a nuclear fraction (N). Successful compartment separation was assessed by immunodetection of αTubulin and Histone H4 as marker proteins for cytoplasm and the nucleus, respectively. SMN, Gemin2 and 3 were analysed by monospecific antibodies. (**D**) Localization of the SMN complex. HeLa cells grown on cover slides were treated with unrip siRNA for 72 h. Overexpression of HA-SMN was achieved by transient transfection. Localization of SMN was analysed by indirect immunofluorescence in control-treated (d) and unrip-silenced cells (e) using monoclonal mouse anti-SMN 7B10 as a primary antibody. Overexpressed HA-SMN was detected using anti-HA (clone 16B12, Babco) as a primary antibody (f). The secondary antibody was a rhodamine-labeled anti-mouse antibody. Nucleus and cytoplasm are indicated by DAPI stain and DIC, respectively.

points (compare lane 2 with lanes 3 and 4). Importantly, silencing of unrip did not affect the expression levels of other components of the SMN complex, such as SMN, Gemin3 or Sm proteins, thus allowing the analysis of unrip reduction directly (Fig. 4B). Cytosolic extracts derived from untransfected (lanes 3 and 4) and unrip-siRNA transfected HeLa cells (lanes 1 and 2) were analysed for U snRNP assembly by native gel electrophoresis. Even though unrip was reduced in RNAi-treated cells by >90%, no significant reduction of the assembly activity was observed when compared to control extract (lanes 3 and 4). This finding suggests that unrip, although being a component of the active SMN complex, is not essential for the assembly reaction *per se*.

RNAi-induced reduction of unrip causes nuclear accumulation of the SMN complex

On the basis of the data described earlier, we next tested whether unrip contributed to the U snRNP biogenesis by influencing the intracellular localization of the SMN complex. Interestingly, immunofluorescence of cells treated with unrip-RNAi revealed a significant increase in SMN and Gemin2-positive nuclear structures (i.e. gems and/or Cajal bodies, Fig. 4D e and Supplementary Material, Section A). Whereas non-transfected control cells contained 1.8 ± 0.12 nuclear foci (Fig. 4D d), their number was increased in unrip-deprived cells more than 3-fold (5.9 ± 0.22 Cajal bodies/cell, see also Statistical Analysis in Supplementary Material Section B). Two lines of evidence suggest that the higher gem/Cajal body number reflected an increased level of SMN complex in the nucleus: Firstly cells that were transfected with a plasmid that allows for overexpression of haemagglutinin (HA)-tagged SMN exhibited a strong increase in nuclear gems/Cajal bodies (13.4 ± 2.53), reminiscent of the effect observed in unrip-deprived cells (Fig. 4D f). Localization of endogenous unrip was not altered in these cells, excluding non-specific effects caused by the overexpression of HA-SMN (Supplementary Material, Section C); furthermore, overexpression of HA-unrip leads to a decrease in SMN-positive nuclear foci (Supplementary Material, Section C). Secondly, when unrip-deprived cells were separated into a nuclear and cytosolic fraction, a significant increase of SMN, Gemin2 and 3 was observed in the nuclear fraction when compared with control cells (Fig. 4C, lanes 3 and 4). Interestingly, this coincided with a decrease of these factors in the cytosol (lanes 1 and 2). Together, our data suggest that unrip influences the intracellular distribution of the SMN complex.

Interaction with Gemin7 incorporates unrip into the SMN complex

As the interaction of unrip with the SMN complex is very stable *in vivo* and even withstands 1.5 M salt (discussed earlier), we anticipated efficient binding to one or several Gemins and/or Sm proteins. This was addressed by *in vitro* protein-binding assays using recombinant glutathione S-transferase (GST)–unrip immobilized to a glutathione–Sepharose matrix and ^{35}S-methionine-labeled SMN complex proteins, i.e. SMN, Gemin2–7 and Sm proteins. After extensive washing, bound proteins were eluted from the resin, resolved by SDS–PAGE and detected by autoradiography. As shown in Figure 5A, only Gemin7 (lane 21) but none of the other core-factors of the SMN complex (i.e. SMN and Gemin2–6) bound to GST–unrip. Binding was specific, as this interaction was not observed with a GST-control protein (Fig. 5A, lane 20). In addition, we detected a weak but reproducible binding of SmB, D1, D2, D3 and E to GST–unrip (Fig. 5A, lane 25–39).

To verify the interaction between unrip and Gemin7, GST–Gemin7 was expressed in *Escherichia coli* either alone or in a complex with His–Gemin6 and incubated with *in vitro* translated and ^{35}S-labeled unrip. An efficient association was observed to both, GST–Gemin7 and the GST–Gemin7/His–Gemin6 complex, but not to GST–protein alone (Fig. 5B). This interaction is direct, rather than mediated by another factor of the reticulocyte lysate, as a complex composed of GST–Gemin7 and His–unrip can be co-purified from bacteria using glutathione–Sepharose (Fig. 5C). To further analyse binding of unrip to Sm proteins, recombinant Sm hetero-oligomers B/D3, D1/D2 and F/E/G were immobilized to CNBr-activated Sepharose and incubated with *in vitro* translated, ^{35}S-labeled unrip. In this assay, we failed to detect significant binding of unrip, leaving open whether these interactions contribute to the incorporation of unrip into the SMN complex (data not shown).

To further characterize the mode of interaction between unrip and Gemin7, *in vitro* binding assays with truncation mutants of both proteins were performed (Fig. 6A). Initially, we wished to determine the sequence within Gemin7 that mediates binding to unrip. As shown in Figure 6B (lane 3 and 4), recombinant GST–Gemin7 and a truncation consisting of the first 56 amino acids (Gemin7$^{1-56}$) bound specifically to full-length unrip, whereas a C-terminal fragment (Gemin 7$^{57-131}$) (lane 5) failed to bind. Therefore, the N-terminus of Gemin7 is necessary and sufficient for binding to unrip. These results were confirmed by *in vitro* binding assays using recombinant unrip as bait and *in vitro* translated Gemin7 as prey. As shown in Figure 6B, a GST–fusion of full-length unrip and a C-terminal fragment encompassing amino acids 195–350 (unrip$^{195-350}$) bound efficiently to Gemin7 and Gemin7$^{1-56}$ (lanes 8, 10, 13 and 15). In contrast, a fragment containing the putative WD-repeats failed to bind in the same assay (lanes 9 and 14). The C-terminal fragment of Gemin7 (Gemin7$^{57-131}$) did not bind any of the unrip truncations (lanes 18-20). Next, we asked whether the very C-terminal tail (amino acids 295–350) of unrip is sufficient for binding to Gemin7. *In vitro* binding assays using unrip$^{295-350}$ showed that this truncation mutant failed to bind to any Gemin7-construct (Fig. 6B, lanes 25, 30 and 35). Thus, neither the putative WD-repeat (unrip$^{1-295}$) nor the C-terminal tail of unrip (unrip$^{295-350}$) alone is sufficient for binding to Gemin7.

We therefore conclude that unrip is recruited to the SMN complex, predominantly if not exclusively by an association with Gemin7. The domains involved in this interaction are the N-terminal 56 amino acids of Gemin7 and the C-terminal part of unrip encompassing amino acids 195–350.

Gemin7 and unr compete for binding to unrip

Our biochemical work clearly indicates that unrip interacts with the SMN complex via Gemin7 and is also associated with unr

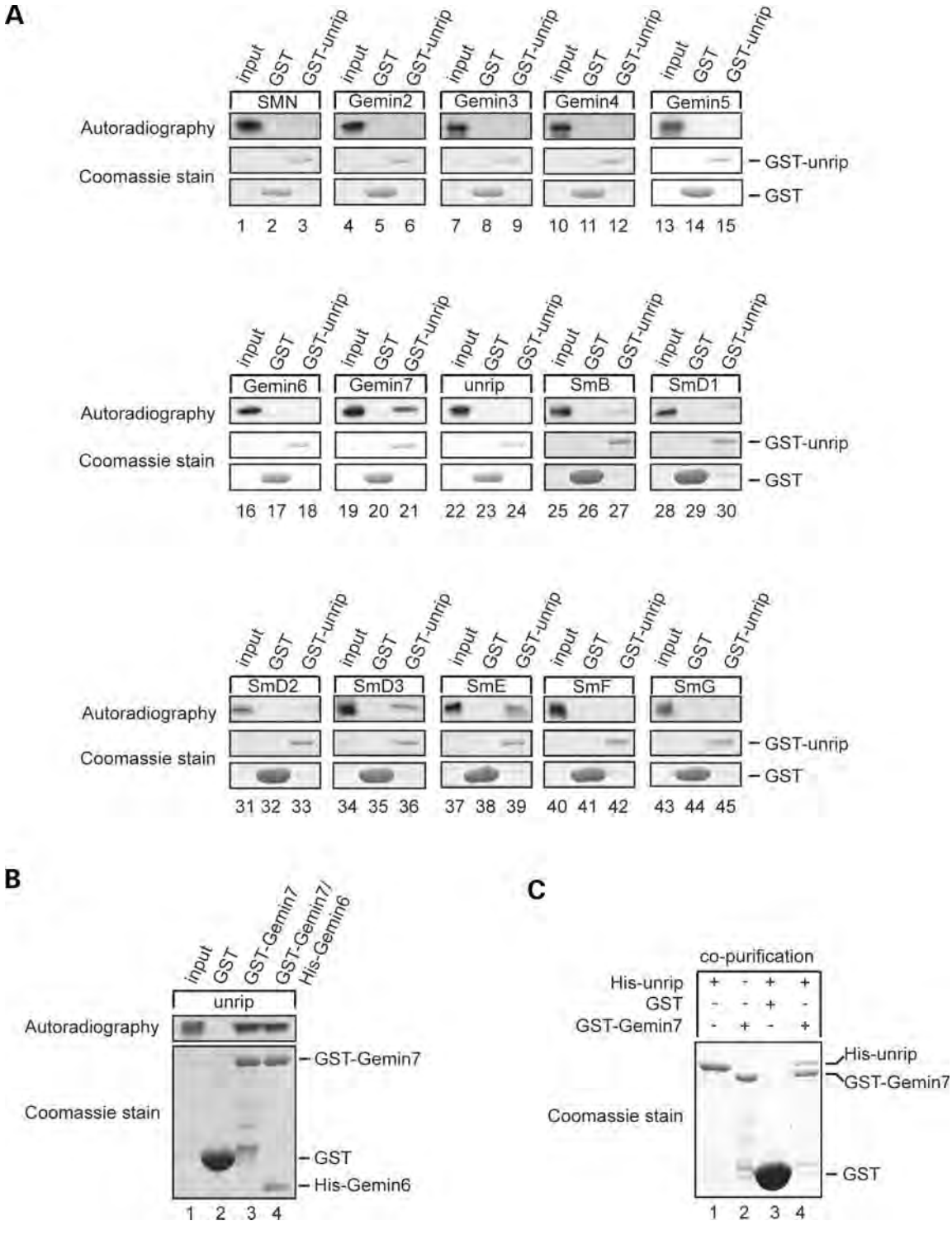


Figure 5. Recruitment of unrip to the SMN complex is mediated by Gemin7. (**A**) Recombinant GST or GST–unrip fusion protein was immobilized on glutathione–Sepharose resin and incubated with the indicated *in vitro* translated ^{35}S-labeled proteins. After extensive washing, the bound proteins were eluted and resolved by SDS–PAGE. The upper panel shows a autoradiography of the binding assay, the middle and lower panels show the Coomassie-stained gel of the corresponding experiment. The input lanes show 10% of the indicated, radiolabeled protein used in the experiment. (**B**) Recombinant GST (lane 2), GST–Gemin7 (lane 3) or a complex composed of GST–Gemin7 and His–Gemin6 (lane 4) was immobilized on glutathione-Sepharose resin and incubated with *in vitro* translated ^{35}S-labeled unrip (10% of the labeled unrip used in this experiment is shown in lane 1). After washing, bound proteins were separated by SDS–PAGE and visualized either by Coomassie staining (upper panel) or by autoradiography (lower panel). (**C**) Direct binding of Gemin7 to unrip. GST–Gemin7 was expressed in *E. coli* either alone (lane 2) or in combination with His-tagged unrip (lane 4) and purified on a glutathione–Sepharose resin. As a control, His–unrip was co-expressed and purified with GST (lane 3). Lane 1 shows His–unrip purified on a nickel resin. Proteins were resolved by SDS–PAGE and visualized by Coomassie staining.

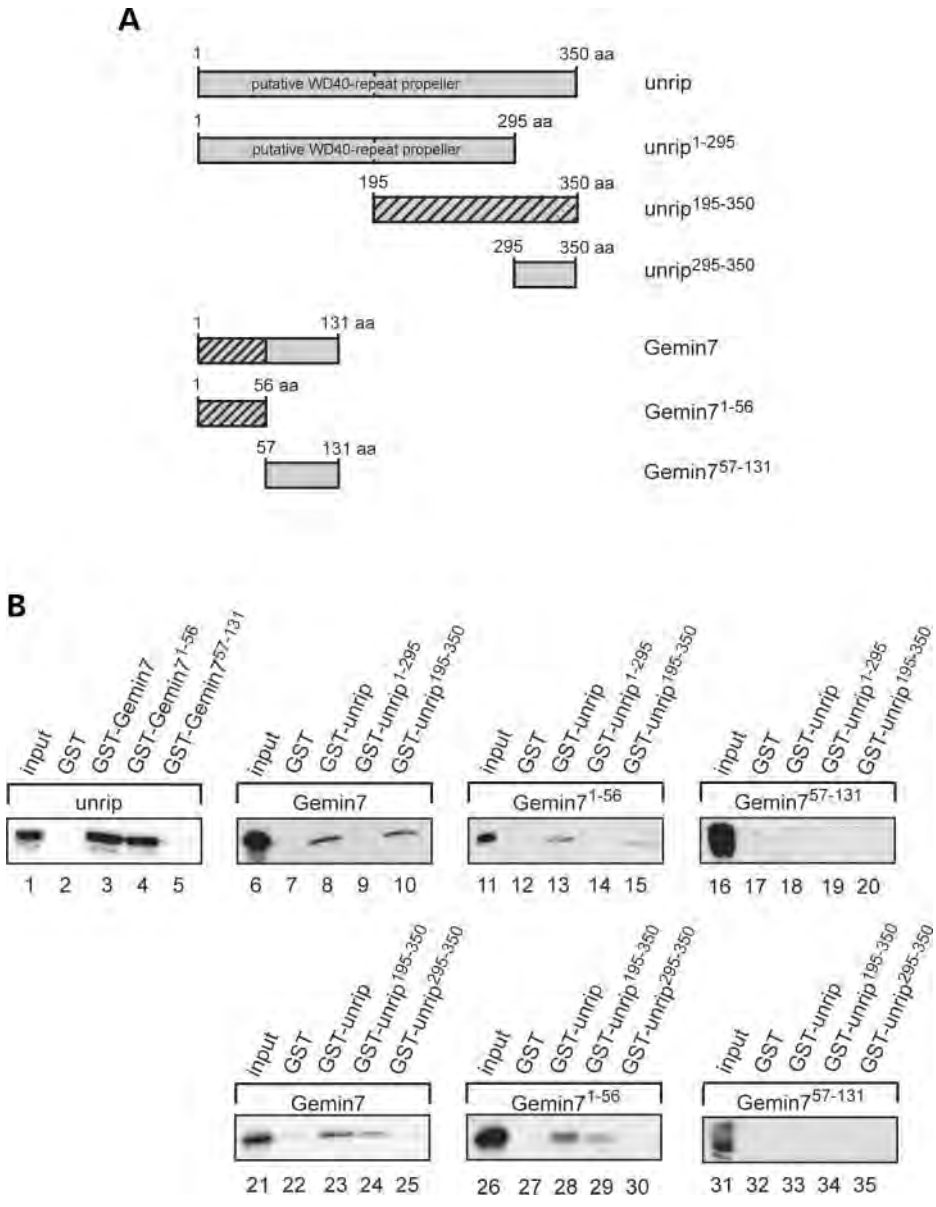


Figure 6. Identification of domains involved in the Gemin7–unrip interaction. (**A**) Schematic presentation of the truncations of Gemin7 and unrip used for the *in vitro* binding assays in (B). Hatched areas mark the putative interacting regions between the two proteins. (**B**) Recombinant GST (lanes 2, 7, 12, 17, 22, 27 and 32), GST–fusions of Gemin7 (lanes 3–5) and GST–fusions of unrip (lanes 8–10, 13–15, 18–20, 23–25, 28–30 and 33–35) were immobilized on glutathione–Sepharose resin and incubated with the indicated *in vitro* translated ^{35}S-labeled proteins. After extensive washing, the bound proteins were eluted, resolved by SDS–PAGE and visualized by autoradiography. The input lanes (1, 6, 11, 16, 21, 26 and 31) show 10% of the indicated radiolabeled protein used in each experiment.

(26). Therefore, the absence of unr in purified SMN complexes raised the question whether the interaction of unrip with unr prevents its association with the SMN complex and *vice versa*. This was assessed by *in vitro* binding assays using immunopurified unrip. For this, unrip immunoprecipitated from HeLa cytosolic extract was incubated with either ^{35}S-labeled Gemin7 (Fig. 7A, upper panel) or unr (lower panel). As a control, unrip pre-immune serum was used. After extensive washing, bound proteins were eluted, resolved by SDS–PAGE and detected by autoradiography. As shown in Figure 7A, both Gemin7 and unr, i.e. the *in vivo* binding partners of unrip, bound efficiently to the unrip-immunoprecipitate but not to the control (compare lanes 2 and 3). We then asked whether the interaction of unr with unrip can be competed by an excess of a purified recombinant Gemin6/7 complex (the heterodimer was used rather than Gemin7 alone as the latter tends to aggregate and may hence associate non-specifically with unrip in this assay). Indeed, the interaction of unr to unrip could be efficiently prevented upon addition of rising amounts of the Gemin6/7 heterodimer to the binding assay (Fig. 7B, lanes 2–6). Bovine serum albumin (BSA) as a control in similar amounts had no effect (lane 7).

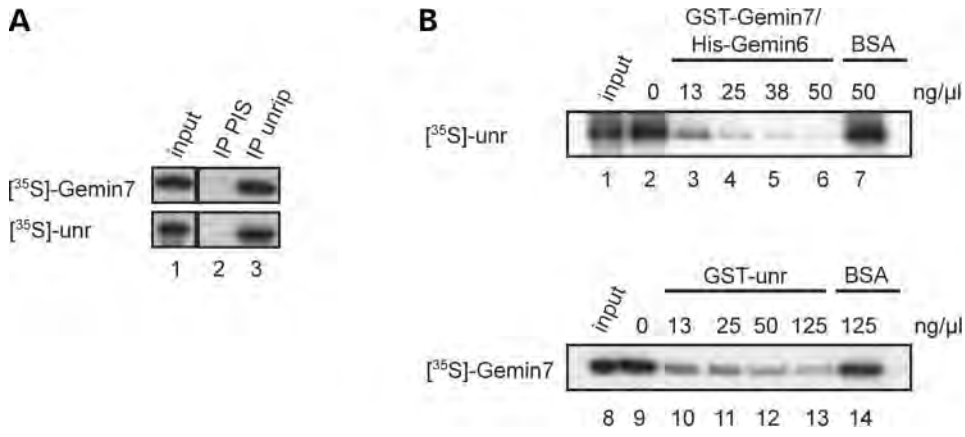


Figure 7. Interaction of unrip with Gemin7 impairs binding of unrip to unr. (**A**) Unrip was affinity-purified from cytosolic HeLa extract. Subsequently, *in vitro* translated ^{35}S-labeled Gemin7 (upper panel) or unr (lower panel) was added, and binding of labeled proteins was analysed by autoradiography of the dried SDS–PAGE gel. As a control, pre-immune serum was used (lane 2). Lane 1 shows 10% of radiolabeled input Gemin7 and unr. (**B**) Immunoprecipitated unrip (~1 µg) from HeLa cytosolic extract was incubated with radiolabeled unr (0.3 fmol) and increasing concentrations (0, 13, 25, 38 and 50 ng/µl) of recombinant, purified Gemin7/Gemin6 complex (upper panel, lanes 2–6). The resin subsequently was washed, proteins were eluted from the matrix, resolved by SDS–PAGE and bound unr was detected by autoradiography. As a control, 50 ng/µl BSA was used as competitor protein (lane 7). In the lower panel, immunoprecipitated unrip (~1 µg) was incubated with radiolabeled Gemin7 (2.4 fmol) and increasing concentrations (0, 13, 25, 50 and 125 ng/µl) of recombinant GST-unr (lanes 9–13). After washing of the resin and SDS–PAGE of the bound proteins, the ^{35}S-labeled Gemin7 was detected by autoradiography. As a control, 125 ng/µl BSA was used as competitor protein (lane 14). Ten percent of ^{35}S-labeled input used in the experiment is shown in lane 1 and 8, respectively.

Finally, we analysed whether recombinant unr could compete for the interaction between unrip and ^{35}S-labeled Gemin7. Increasing amounts of recombinant unr gradually displaced Gemin7 from unrip (Fig. 7B, lower panel, lanes 9–13), although the observed competition was less efficient compared with the inverse experiment (Fig. 7B, upper panel, lanes 2–6). Thus, these *in vitro* binding studies indicate that binding of unrip to either unr or Gemin7 is mutually exclusive.

DISCUSSION

Biochemical studies have revealed a large number of factors that are associated with SMN *in vivo* (reviewed in 23). However, among these interactors, only a subset appears to be stably and stoichiometrically bound. These factors, collectively named 'Gemins', are functionally associated with the SMN complex, the entity that promotes the assembly of spliceosomal U snRNPs. Although the principal role of the SMN complex as a whole in the biogenesis of U snRNPs is well established, only limited information is available regarding the contribution of individual components to the assembly reaction. We have provided several lines of evidence here that unrip, a WD-repeat protein and interacting partner of the translation factor unr, is a major component of the SMN complex. Firstly by immunoprecipitation with monospecific anti-unrip antibodies, we recover not only unr, but also the SMN complex. Consistent with our previous data (10), isolated SMN complex contains significant levels of unrip. Secondly, depletion of unrip from cytosolic extract abolishes its competence to promote formation of U snRNP particles. Furthermore, this procedure simultaneously co-depletes other factors of the assembly machinery, such as SMN, Gemin2 and 3. Thirdly, we identify Gemin7, a known component of the SMN complex (31), as the primary and direct binding partner for unrip. In fact, association of unrip with the SMN complex is stable up to 1.5 M NaCl or 0.3 M KCl/ 0.5 M NaSCN, suggesting high affinity binding in the cytosol. These data strongly argue that unrip is recruited to the SMN complex via an interaction with Gemin7. Our findings are in good agreement with a recent report published by Pellizzoni- and co-workers (32) providing evidence for a interaction of unrip with the SMN complex. This study also reports binding of unrip to Sm proteins. Although we are also able to detect weak interactions between recombinant unrip and *in vitro* translated Sm proteins, immobilized recombinant Sm proteins heterodimers failed to interact with *in vitro* translated unrip in the same assay. It is therefore questionable whether the Sm proteins are in fact interaction partners of unrip *in vivo*.

We show that the interaction between Gemin7 and unrip is mediated via an N-terminal sequence element in Gemin7 that overlaps with a motif previously implicated in SMN-binding (33). Given that unrip is a stoichiometric component of the cytosolic SMN complex, one could imagine that unrip and SMN can bind simultaneously to Gemin7. We have attempted to test this possibility, but failed to see a significant binding of SMN to Gemin7 and the Gemin6/7 dimer. Therefore, further studies will be required to determine the mode of interaction between these factors.

The stable association of unrip with the SMN complex prevented selective depletion of this protein from extracts without simultaneously reducing the levels of other assembly factors. To circumvent this problem, we employed RNAi to specifically reduce unrip expression without affecting other components of the SMN complex. Extracts derived from these cells assembled U1 snRNPs as efficiently as the control extract (Fig. 4B). Although these *in vitro* studies suggest that unrip is not essentially required for the assembly reaction *per se*, the consequence of unrip-depletion *in vivo* is currently less clear. To address this question, an experimental setup to

study the biogenesis of U snRNPs *in vivo* is highly demanded. Such a test system has been established in *Xenopus laevis* oocytes, but an equivalent one in somatic cells is currently not available. The development of this tool will be a major technical challenge for future studies on the *in vivo* function of the SMN complex.

A hallmark of the SMN complex is its localization not only in the cytoplasm but also in the nuclear gems/Cajal bodies. Although the localization of the SMN complex in the cytoplasm correlates well with its function in U snRNP assembly, it is currently unclear why this complex can also be detected in the nucleus. It is hypothesized that the SMN complex accompanies the assembled U snRNP to the nucleus where both entities dissociate. Gems, structures which often overlap with Cajal bodies (also termed coiled bodies), may be sites where U snRNPs and SMN complex are separated and the latter is redirected to the cytoplasm (21,34). The predominant localization of unrip to the cytoplasm could indicate that it either dissociates from the SMN complex prior to nuclear import or rapidly returns to the cytoplasm after import along with the SMN complex. Thus, unrip may participate in the activation of the SMN complex after its return from the nucleus. Alternatively and not mutually exclusive with the first scenario, unrip may help to retard the SMN complex in the cytoplasm while the assembly reaction is in progress. Consistent with the latter model is our observation that RNAi-induced reduction of unrip appears not to interfere with the assembly process *per se*. Rather, it leads to an increase in the nuclear pool of the SMN complex, indicating that the presence of unrip shifts the steady-state distribution of the SMN complex to the cytoplasm. Further studies are required to address how unrip influences the subcellular trafficking of the SMN complex.

Our biochemical work, presented here, shows that only a fraction of the endogenous unrip is associated with the SMN complex, whereas the rest forms the previously described unr/unrip dimer. In functional assays, this dimer has been linked to the cap-independent translation of cellular and viral mRNAs containing internal ribosome entry sites (IRES) (26,27). The role of unrip in this scenario currently remains unclear. Our finding that the binding of unrip to unr and Gemin7 is mutually exclusive argues that overlapping regions of unrip are recognized by both proteins. Understanding the role of unrip in the context of U snRNP assembly may thus provide useful insights into its function in cap-independent translation and *vice versa*. In conclusion, our identification of unrip as a major, yet atypical, component of the SMN complex has set the stage for detailed examinations of both SMN-mediated U snRNP assembly and IRES-driven translation.

MATERIALS AND METHODS

DNA constructs

Full-length cDNAs corresponding to the open-reading frames of unrip, unr (mRNA variant 1), Gemin6 and Gemin7 were amplified from a human brain Marathon cDNA library (Clontech). For *in vitro* translation and expression of recombinant proteins, unrip, unr, Gemin6 and Gemin7 were subcloned into pET28a vector (Novagen). For expression of GST–fusion proteins unrip and Gemin7 were subcloned into pGEX6P-1 (Amersham). Truncations of unrip and Gemin7 were generated with specific primers and subcloned into pGEX6P-1 or pET28a vector, respectively. Plasmids encoding the Sm proteins B, D1, D2, D3, E, F and G, SMN and Gemin2–5 have been described previously (10,28,35,36).

Recombinant proteins and *in vitro* protein binding assays

Full-length human Gemin6, Gemin7, unrip and truncated versions thereof were expressed as GST–fusion proteins in *E. coli* Rosetta(DE3) pLysS (Novagen). *E. coli* harboring the respective plasmids were cultured in superbroth medium. Protein expression was induced at mid-logarithmic phase by the addition of 1 mM IPTG. Expression was performed for 5 h at 16°C. Bacteria were harvested by centrifugation, resuspended in lysis buffer [25 mM NaCl, 20 mM Tris/HCl pH 8.0, 1 mM DTT, 0.01% IgePal and protease inhibitors Leupeptin, PepstatinA, Aprotonin (10 μg/ml each) and AEBSF (0.1 mM)] and lysed by sonication. GST–fusion proteins were purified on glutathione–Sepharose resin (Amersham) following the manufacturer's instructions. Recombinant GST–Gemin7/His-tagged-Gemin6 complex was obtained by co-expression of proteins in *E. coli* strain Rosetta(DE3) pLysS following procedures described earlier. The protein complex was purified by two consecutive affinity purification steps on glutathione–Sepharose as the first step and Pro-bond nickel chelating resin (Invitrogen) as the second step. ^{35}S-methionine-labeled proteins were produced using a TNT-T7 Quick Coupled Transcription/Translation System (Promega). *In vitro* translated proteins were incubated with ∼ 2μg purified GST–fusion proteins, immobilized on a glutathione–Sepharose resin (Amersham) and allowed to bind in lysis buffer at 4°C for 1 h. After washing the resin three times with lysis buffer, bound proteins were eluted by boiling in 2× SDS sample buffer, resolved by SDS–PAGE and analysed by Coomassie staining. Labeled proteins were detected by autoradiography of the dried gel.

Immunofluorescence microscopy

HeLa cells were grown on cover slides in DMEM/10% FCS and were washed once in phosphate-buffered saline (PBS), fixed with 3% paraformaldehyde for 7 min and permeabilized in 0.2% Triton X-100/PBS for 5 min on ice. After blocking with 2% BSA in PBS, cells were incubated for 1 h with monoclonal mouse anti-SMN antibody 7B10 or affinity-purified polyclonal rabbit unrip-antibody. As secondary antibody, fluorescein- or rhodamine-conjugated secondary anti-mouse or anti-rabbit antibodies were used and analysed with a 63× oil immersion lens on a Zeiss Axiovert 200M microscope. Localization of overexpressed, HA-tagged SMN was analysed with a mouse monoclonal HA-antibody (clone 16B12, Babco) and a secondary rhodamine anti-mouse antibody. SMN and Gemin2 double-labeling experiments were performed using affinity-purified rabbit anti-SMN and mouse monoclonal anti-Gemin2 antibody (clone4, BD Transduction Laboratories) as primary antibodies. For SMN and unrip co-immunofluorescence, monoclonal SMN 7B10 and affinity-purified rabbit unrip antibody were used. Secondary antibodies were rhodamine-labeled anti-rabbit and fluorescein-labeled anti-mouse antibodies, respectively.

Preparation of HeLa cell extracts and antibody production

To obtain cytosolic extract, active in U snRNP assembly, 20l cultured suspension HeLa cells were harvested by centrifugation for 5 min at 1.500 *g* and washed once in PBS. The cell pellet (~15 g) was resuspended in 2.5 volumes of PBS containing 0.01% IgePal. Cells were lysed in a Dounce homogenizer, S (B. Braun) by 20 strokes. Cytosolic extract was obtained by centrifugation in a swing out rotor at 17.000 *g* for 10 min. Complete clarification of the extract was achieved by filtration through a 0.45 μm low protein binding filter. Cytosolic and nuclear extracts were prepared following a protocol of Dignam *et al.* (37). The separation of nuclear or cytosolic extract of siRNA-treated cells was achieved with a Qproteome Cell Compartment Kit (Qiagen) according to the manufacturer's protocol. Compartment separation was analysed by immunodetection of cytoplasmic marker protein αTubulin (anti-αTubulin antibody, clone B5-1-1, Sigma) and Histone H4 as a nuclear marker protein (anti-Histone H4 antibody, Cell Signaling Technology), respectively.

Antibodies against unrip and unr were raised by injection of recombinant full-length human proteins into rabbits. Antibodies were affinity-purified on columns with the respective covalently linked antigen. Immunoprecipitation and detection of SMN in western blots were performed with the mouse monoclonal antibody 7B10 (28).

Immunoprecipitations and immunodepletions

Immunoprecipitations and immunodepletions were carried out in HeLa cytosolic extract, using antibodies covalently linked to protein A–Sepharose. To attempt dissociation of unrip from the SMN complex, immunodepletion was carried out in extracts either brought to 1.5 M NaCl or to 0.3 M KCl/0.5 M NaSCN. For comparison, untreated extract was diluted with PBS to the final resulting volume.

RNA interference

Unrip levels were reduced by transfection of a mixture of two double-stranded 21 nt long siRNA (sequences: 5′-AAACU GUUACGCAUAUAUGACTT-3′ and 5′-AACUUAUGGAC GAUCUAUUGCTT-3′, purchased from IBA Nucleic Acids Synthesis, Göttingen) with Oligofectamine™ (Invitrogen) following the protocol of the manufacturer. Silencing of unrip was assayed by western blotting of cell extracts, 48, 72, 96 and 120 h after transfection, using a monospecific unrip antiserum. Transient transfection of adherent HeLa cells with a N-terminally HA-tagged SMN (subcloned in pcDNA3.1 vector, Invitrogen) was carried out with Effectene™ (Qiagen) according to the protocol of the manufacturer and analysed 48 h after transfection.

In vitro assembly of U snRNPs

To analyse U snRNP assembly *in vitro*, 3 μl HeLa cytosolic extract (containing ~10 mg/ml protein) or immunoprecipitated SMN complex were incubated at 37°C for 45 min with 25 fmol ^{32}P-U1 RNA, 2 μg tRNA, 1 mM ATP and 1 μl RNasin in a final volume of 20 μl. The reactions were analysed by native gel electrophoresis as described previously (10). Analysis of the assembly activity of unrip-silenced cells was carried out with cytosolic extract prepared with a Qproteome Cell Compartment Kit (Qiagen) according to the manufacturer's protocol. Assembly reaction was performed as described earlier.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We are indebted to R. Lührmann and J. Steitz for providing reagents and E. Dinkl for technical help. This work was supported by grants of the German Research Foundation (DFG, FOR426 and SFB 581) and families of SMA (fsma).

Conflict of Interest statement: Authors have declared no conflict of interest.

REFERENCES

1. Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5–14.
2. Raker, V.A., Plessel, G. and Luhrmann, R. (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.
3. Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell. Biol.*, **13**, 290–301.
4. Plessel, G., Fischer, U. and Luhrmann, R. (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.
5. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. and Bordonne, R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell*, **9**, 891–901.
6. Hamm, J., Darzynkiewicz, E., Tahara, S.M. and Mattaj, I.W. (1990) The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell*, **62**, 569–577.
7. Fischer, U. and Luhrmann, R. (1990) An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science*, **249**, 786–790.
8. Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Luhrmann, R. (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.
9. Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, **90**, 1023–1029.
10. Meister, G., Buhler, D., Pillai, R., Lottspeich, F. and Fischer, U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell. Biol.*, **3**, 945–949.
11. Meister, G. and Fischer, U. (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J.*, **21**, 5853–5863.
12. Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. *J. Biol. Chem.*, **277**, 7540–7545.
13. Friesen, W.J., Massenet, S., Paushkin, S., Wyce, A. and Dreyfuss, G. (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol. Cell*, **7**, 1111–1117.

14. Branscombe, T.L., Frankel, A., Lee, J.H., Cook, J.R., Yang, Z., Pestka, S. and Clarke, S. (2001) PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. *J. Biol. Chem.*, **276**, 32971–32976.
15. Friesen, W.J. and Dreyfuss, G. (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.
16. Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. and Luhrmann, R. (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.
17. Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C. and Fischer, U. (2001) Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.*, **11**, 1990–1994.
18. Narayanan, U., Ospina, J.K., Frey, M.R., Hebert, M.D. and Matera, A.G. (2002) SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum. Mol. Genet.*, **11**, 1785–1795.
19. Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J. and Bordonne, R. (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular athrophy protein, survival of motor neuron. *EMBO Rep.*, **4**, 616–622.
20. Narayanan, U., Achsel, T., Luhrmann, R. and Matera, A.G. (2004) Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell*, **16**, 223–234.
21. Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W. and Dreyfuss, G. (2002) The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol. Cell. Biol.*, **22**, 6533–6541.
22. Grimmler, M., Bauer, L., Nousiainen, M., Korner, R., Meister, G. and Fischer, U. (2005) Phosphorylation regulates the activity of the SMN complex during assembly of spliceosomal U snRNPs. *EMBO Rep.*, **6**, 70–76.
23. Meister, G., Eggert, C. and Fischer, U. (2002) SMN-mediated assembly of RNPs: a complex story. *Trends Cell. Biol.*, **12**, 472–478.
24. Liu, Q. and Dreyfuss, G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.*, **15**, 3555–3565.
25. Matera, A.G. and Frey, M.R. (1998) Coiled bodies and gems: Janus or gemini? *Am. J. Hum. Genet.*, **63**, 317–321.
26. Hunt, S.L., Hsuan, J.J., Totty, N. and Jackson, R.J. (1999) unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes Dev.*, **13**, 437–448.
27. Mitchell, S.A., Spriggs, K.A., Coldwell, M.J., Jackson, R.J. and Willis, A.E. (2003) The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. *Mol. Cell*, **11**, 757–771.
28. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F. and Fischer, U. (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.
29. Gubitz, A.K., Feng, W. and Dreyfuss, G. (2004) The SMN complex. *Exp. Cell Res.*, **296**, 51–56.
30. Malatesta, M., Scassellati, C., Meister, G., Plottner, O., Buhler, D., Sowa, G., Martin, T.E., Keidel, E., Fischer, U. and Fakan, S. (2004) Ultrastructural characterisation of a nuclear domain highly enriched in survival of motor neuron (SMN) protein. *Exp. Cell Res.*, **292**, 312–321.
31. Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) Identification and characterization of Gemin7, a novel component of the SMN complex. *J. Biol. Chem.*, **277**, 31957–31962.
32. Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. and Pellizzoni, L. (2005) Unrip is a component of SMN complexes active in snRNP assembly. *FEBS Lett.*, **579**, 2348–2354.
33. Ma, Y., Dostie, J., Dreyfuss, G. and van Duyne, G.D. (2005) The Gemin6–Gemin7 heterodimer from the survival of motor neurons complex has an sm protein-like structure. *Structure*, **13**, 883–892.
34. Hebert, M.D., Shpargel, K.B., Ospina, J.K., Tucker, K.E. and Matera, A.G. (2002) Coilin methylation regulates nuclear body formation. *Dev. Cell*, **3**, 329–337.
35. Lehmeier, T., Raker, V., Hermann, H. and Luhrmann, R. (1994) cDNA cloning of the Sm proteins D2 and D3 from human small nuclear ribonucleoproteins: evidence for a direct D1–D2 interaction. *Proc. Natl Acad. Sci. USA*, **91**, 12317–12321.
36. Hermann, H., Fabrizio, P., Raker, V.A., Foulaki, K., Hornig, H., Brahms, H. and Luhrmann, R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.*, **14**, 2076–2088.
37. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.

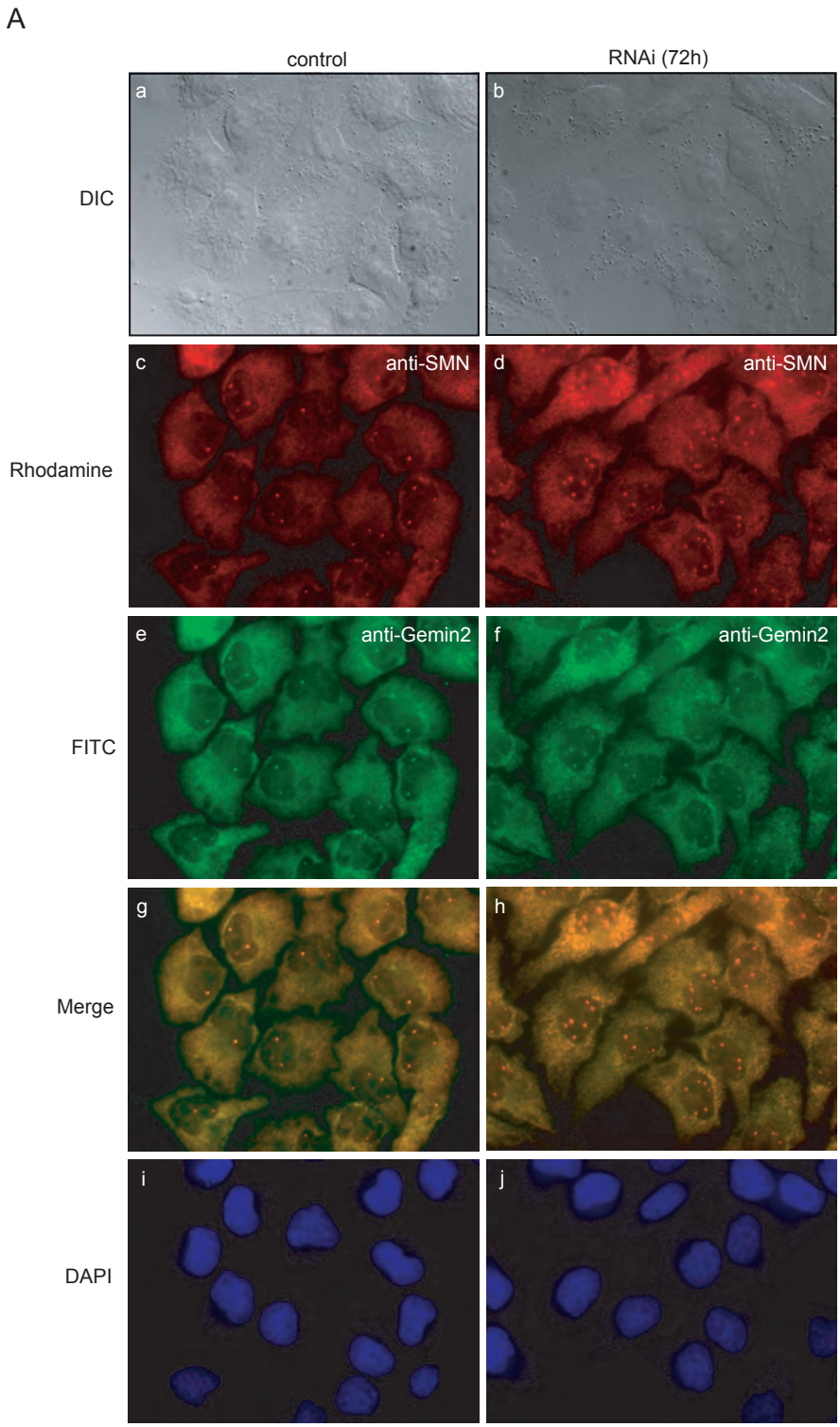
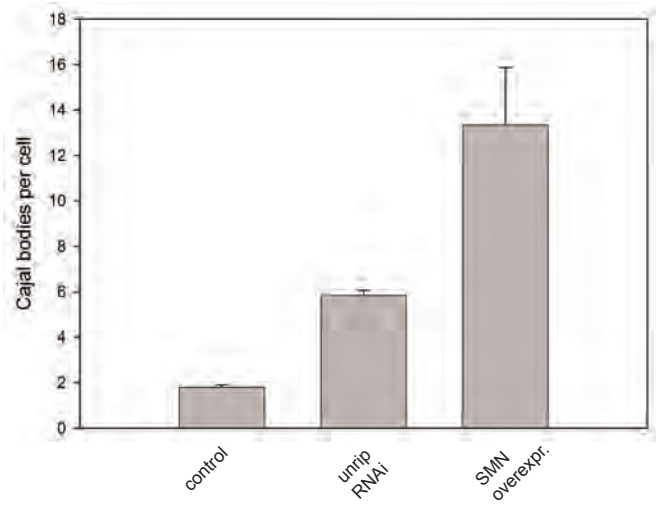


Figure supplement A. Silencing of unrip alters localization of the SMN-complex to nuclear gems/Cajal bodies.
Localization of SMN and Gemin2 in HeLa cells after silencing of unrip (72 h post transfection) was analyzed by indirect immunofluorescence. Affinity purified rabbit polyclonal SMN-antibody (panel d) and monoclonal Gemin2-antibody (clone 4, BD Transduction Laboratories, panel f) were used as primary antibodies. Rhodamine-labeled anti-rabbit and fluorescein-labeled anti-mouse antibody were used as secondary antibodies. Co-localization of proteins is shown by overlay of both images (yellow merge, panel h). As control untransfected HeLa cells were used. Cytoplasm was detected by differential interference contrast (DIC, panels a and b), nucleii were identified by DAPI stain (panels i and j).

B



	control	unrip RNAi	SMN overexpression
	C.b./cell	C.b./cell	C.b./cell
exp. no.1	1,67	5,98	11,00
exp. no.2	1,80	5,59	16,02
exp. no.3	1,90	5,96	12,99
counted cells/ exp.	125		
Cajal bodies per cell	1,79	5,85	13,34
standard deviation	0,12	0,22	2,53

Figure supplement B. Statistic analysis of gem/Cajal body number in unrip-silenced or SMN overexpressing cells.
Unrip was knocked down for 72 h in HeLa cells and gems/Cajal bodies were visualized by immunofluorescence with anti-SMN antibodies. HA-tagged SMN was overexpressed in HeLa cells and gems/Cajal bodies were visualized by immunofluorescence with anti-HA antibodies (see also Figure 4D).
The number of gems/Cajal bodies is presented in a diagram showing three independent transfections for unrip-silenced, SMN overexpressing and control cells. In each experiment the number of Cajal bodies of 125 cells were analyzed. Standard deviations were calculated with Microsoft Excel™ and data were visualized by SigmaPlot™ 9.

C

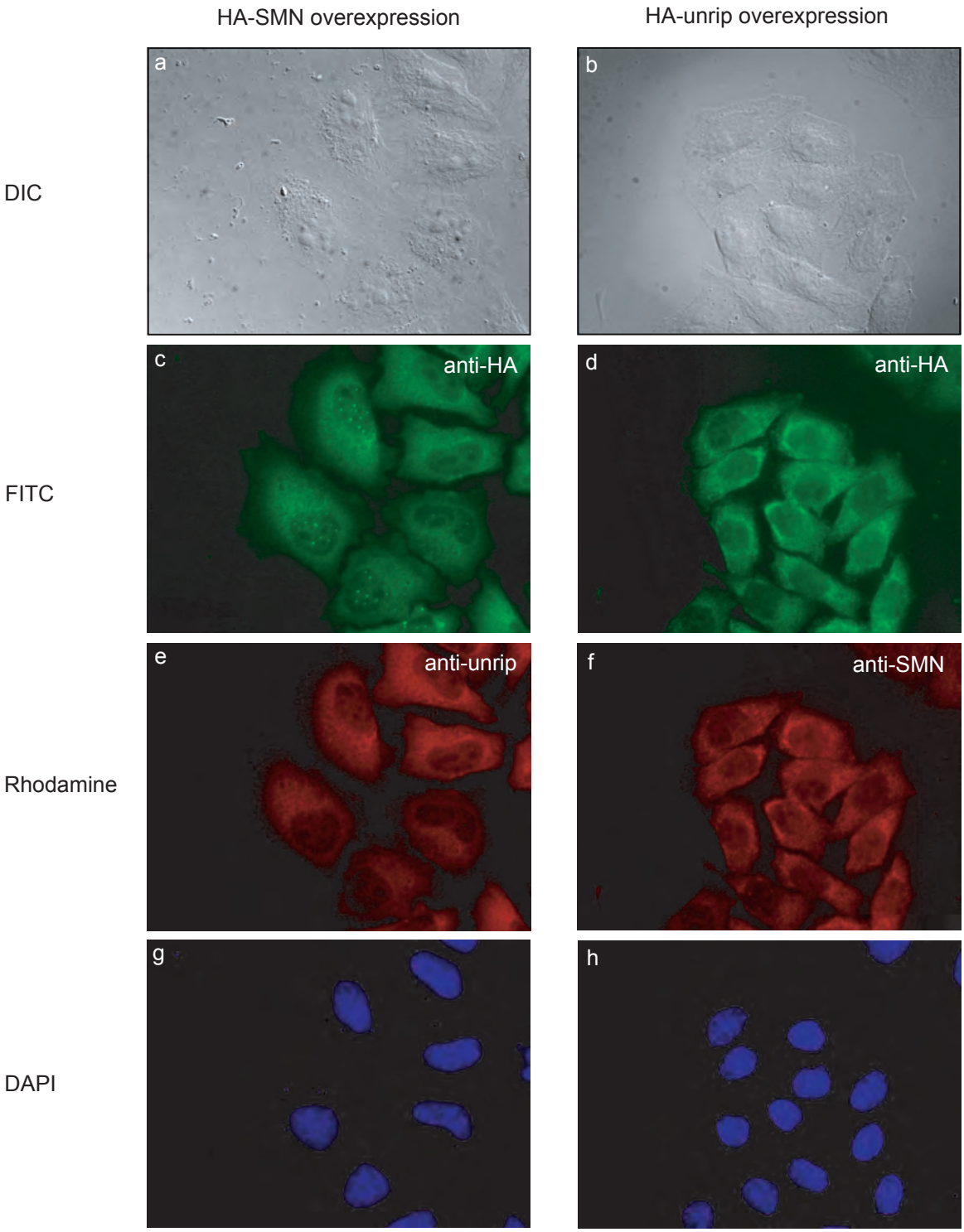


Figure supplement C. Effect of overexpression of unrip or SMN on SMN-complex localization.
HA-tagged SMN (left column) or HA-tagged unrip (right column) were overexpressed in HeLa cells growing on cover slides. The cells were washed once in PBS, fixed with 3% paraformaldehyde for 7 minutes and permeabilized in 0.2% Triton X-100/PBS for 5 minutes on ice. After blocking with 2% BSA in PBS, cells were incubated for 1 h with monoclonal mouse anti-HA antibody (clone 16B12, Babco) to analyze localization of overexpressed HA-tagged protein (panels c and d). As secondary antibody fluorescein-conjugated anti-mouse antibody was used. Subsequently cells were analyzed with a 63X oil immersion lense on a Zeiss Axiovert 200M microscope. Localization of unrip (panel e) or SMN (panel f), respectively, were analyzed with monospecific rabbit polyclonal antisera and rhodamine-conjugated anti-rabbit antibody. Cytoplasm was detected by differential interference contrast (DIC, panels a and b), nucleii were identified by DAPI stain (panels g and h).

5.2 Definition of the Basic Architecture of the Human SMN-Complex

A comprehensive interaction map of the human survival of motor neuron (SMN) complex

Otter S, Grimmler M, Neuenkirchen N, Chari A, Sickmann A, Fischer U**: J Biol Chem 2007,** 282:5825-5833.

Thesis author's contribution:

Conception:	30 %
Experimental contribution:	10 %
Formulation of results:	30 %

Supplemental Material can be found at:
http://www.jbc.org/cgi/content/full/M608528200/DC1

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 8, pp. 5825–5833, February 23, 2007
© 2007 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

A Comprehensive Interaction Map of the Human Survival of Motor Neuron (SMN) Complex*[S]

Received for publication, September 5, 2006, and in revised form, December 1, 2006 Published, JBC Papers in Press, December 18, 2006, DOI 10.1074/jbc.M608528200

Simon Otter[‡1], Matthias Grimmler[‡1], Nils Neuenkirchen[‡], Ashwin Chari[‡], Albert Sickmann[§], and Utz Fischer[‡2]

From the [‡]Theodor-Boveri Institute at the Biocenter, University of Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany and the [§]Rudolf-Virchow-Center, Deutsche Forschungsgemeinschaft-Research Center for Experimental Biomedicine, Versbacher Strasse 9, D-97078 Wuerzburg, Germany

Assembly of the Sm-class of U-rich small nuclear ribonucleoprotein particles (U snRNPs) is a process facilitated by the macromolecular survival of motor neuron (SMN) complex. This entity promotes the binding of a set of factors, termed LSm/Sm proteins, onto snRNA to form the core structure of these particles. Nine factors, including the SMN protein, the product of the spinal muscular atrophy (SMA) disease gene, Gemins 2–8 and unrip have been identified as the major components of the SMN complex. So far, however, only little is known about the architecture of this complex and the contribution of individual components to its function. Here, we present a comprehensive interaction map of all core components of the SMN complex based upon *in vivo* and *in vitro* methods. Our studies reveal a modular composition of the SMN complex with the three proteins SMN, Gemin8, and Gemin7 in its center. Onto this central building block the other components are bound via multiple interactions. Furthermore, by employing a novel assay, we were able to reconstitute the SMN complex from individual components and confirm the interaction map. Interestingly, SMN protein carrying an SMA-causing mutation was severely impaired in formation of the SMN complex. Finally, we show that the peripheral component Gemin5 contributes an essential activity to the SMN complex, most likely the transfer of Sm proteins onto the U snRNA. Collectively, the data presented here provide a basis for the detailed mechanistic and structural analysis of the assembly machinery of U snRNPs.

Several nuclear RNA-protein complexes (RNPs)[3] involved in the processing of mRNAs, such as the snRNPs of the major (U1, U2, U4/6, and U5) and minor (U11, U12, U5, and U4/6atac) spliceosome and the histone-mRNA processing U7 snRNP contain a set of evolutionary conserved proteins of the Sm/LSm class (1, 2). This group of proteins has the propensity to form heptameric rings in the presence of their respective target snRNA. Sm and LSm/Sm rings (also called "cores") can form spontaneously *in vitro* on their target RNAs (3–5). However, assembly *in vivo* occurs in a highly regulated manner and is assisted by *trans*-acting factors. One well characterized entity in this pathway is the SMN complex, whose name-giving component SMN is the product of the spinal muscular atrophy (SMA) disease gene (6, 7). This entity recruits all Sm proteins and promotes their transfer onto the U snRNAs (8–10). Likewise, assembly of the U7 snRNP is facilitated by a specialized SMN complex that is charged with the unique set of Sm and LSm proteins of this particle (11). With a sedimentation coefficient of 25–40 S and an estimated molecular mass exceeding 1 megadalton, the SMN complex represents a macromolecular machine of great complexity. So far, nine major proteins termed SMN, Gemins 2–8, and unrip as well as nine Sm/LSm protein "substrates" (*i.e.* B/B′, D1, D2, D3, E, F, G, LSm10, and LSm11) have been identified as components of this assembly machinery (12, 13).

Our knowledge about the architecture of the core SMN complex (*i.e.* the SMN complex without substrate proteins) is still limited and relies mostly on *in vitro* interaction assays using recombinant proteins. These studies have placed the SMN protein in the center with interactions to Gemins 2, 3, 5, and 7, as well as substrate Sm proteins. In contrast, other factors such as Gemins 2, 4–6, and 8 and unrip appeared to be peripheral and interact only with few other components of the complex (see Refs. 12 and 13 for reviews; Refs. 14–16). The relevance of these interactions in the context of the native SMN complex remained unclear from these studies. To address this issue, we applied a combination of *in vitro* and *in vivo* approaches (*in vitro* binding assays, co-immunoprecipitations and a yeast two-hybrid interaction system) to establish a comprehensive interaction map of the SMN complex. Furthermore, by use of a novel assay we were able to reconstitute the SMN complex in reticulocyte lysate. Interestingly, complex formation is severely compromised by an SMA-causing mutation in the SMN protein. Finally, we have assessed the contribution of the peripheral component Gemin5 to the activity of the SMN complex.

EXPERIMENTAL PROCEDURES

DNA Constructs—Plasmids encoding full-length cDNAs corresponding to the open reading frames of SMN, unrip, and

* This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (Fi FOR426 and SFB581 (to U. F.)) and the Rudolf Virchow Center (DFG-Forschungsprojekt FZT-82 (to A. S.). 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.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–3 and Refs. 1–3.

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 49-931-888-4029; Fax: 49-931-888-4028; E-mail: utz.fischer@biozentrum.uni-wuerzburg.de.

[3] The abbreviations used are: RNP, ribonucleoprotein; U snRNP, uridine-rich small ribonucleoprotein; snRNA, small nuclear ribonucleic acid; SMN, survival of motor neurons; Gemins 2–8, components of gems number 2–8, respectively; GST, glutathione *S*-transferase; HA, hemagglutinin; siRNA, short interfering RNA; RNAi, RNA interference; X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

Gemins 2–7 have been described previously (15). The full-length open reading frame of Gemin8 was purchased from RZPD (Clone IRAUp969F1069D6) and subcloned into the vectors pGEX6P-1 (GE Healthcare), pET28a (Novagen), and pHA (an N-terminal HA-tag containing derivative of pcDNA3.1; Invitrogen).

Recombinant Proteins and in Vitro Protein Binding Assays—Expression and purification of single proteins or protein complexes were performed as described (15). [^{35}S]Methionine-labeled proteins were produced using the TNT-T7 quick coupled transcription/translation system (Promega). *In vitro* co-translations of the entire SMN complex were carried out in one single TNT reaction with [^{35}S]methionine, using a mixture of vectors encoding for the different SMN complex components. In the case of GST binding assays, *in vitro* translated proteins were incubated with ~2 μg of purified GST fusion proteins, immobilized on glutathione-Sepharose (GE Healthcare), and allowed to bind in lysis buffer (50 mM Tris/HCl, pH 7.5, 200 mM NaCl, 0.01% Igepal, 1 mM dithiothreitol, 5 mM EDTA, 5 mM EGTA, 1 μg/ml bovine serum albumin) at 4 °C for 1 h. After washing the resin extensively, bound proteins were eluted by boiling in 2× SDS sample buffer, resolved by SDS-PAGE, and analyzed by Coomassie staining and autoradiography of the dried gel.

Preparation of HeLa Cell Extract, Antibodies, and Immunoprecipitations—HeLa cytosolic extract was prepared as described (17). Antibodies against Gemins 5–8 were raised by injection of full-length human proteins into rabbits. Antibodies were affinity purified on columns with the respective covalently linked antigen. Anti-Gemin4 and anti-SmB/B′ were purchased from Santa Cruz Biotechnology. Immunoprecipitations of the SMN complex were carried out in HeLa cytosolic extract, using a covalently linked 7B10 monoclonal SMN antibody (8). Precipitated SMN complex was washed extensively with buffer (50 mM Tris/HCl, pH 7.5, 1 mM EDTA, 0.01% Igepal) of rising ionic strength (150 mM, 250 mM, 750 mM, 500 mM, 1 M, 1.5 M NaCl). Resulting SMN complexes were eluted by 2× SDS buffer and resolved on 8–20% SDS-PAGE gradient gels. For reconstitution of SMN complexes with SmB/D3, an immobilized complex was washed with 1.5 M NaCl high salt buffer and incubated with recombinant SmB/D3 heterodimer, purified as described (18). After extensive washing, complexes were eluted by 2× SDS buffer and resolved in 8–20% SDS-PAGE.

RNA Interference (RNAi)—Gemin2 levels were reduced by transfection of HeLa cells with two double-stranded 21-nucleotide-long siRNAs (sequences: RNAi-1, 5′-GGAAGCAAAGUGUGAAUAUTT-3′ and RNAi-2, 5′-GCAGCUCAAUGUCCAGAUGTT-3′). siRNAs were purchased from IBA Nucleic Acids Synthesis (Göttingen, Germany) and transfected with Oligofectamine™ (Invitrogen), following the protocol of the manufacturer. Silencing of Gemin2 was assessed 52 h after transfection by Western blotting of total cell extract, using a monospecific Gemin2 antibody. Immunoprecipitations from these extracts were performed as described above, using extracts normalized for SMN by Western blotting with specific antiserum.

Purification of Baculoviral Expressed Gemin5 and in Vitro Assembly of U snRNPs—His-tagged Gemin5 protein was expressed using the MultiBac system (19). Cells were harvested 5 days post-infection by centrifugation and lysed with detergent (20 mM NaHEPES, 200 mM NaCl, 0.25% Triton X-100, pH 7.5). The clarified lysate was subjected to nickel-nitrilotriacetic acid purification, washed, and eluted with imidazole. 60–200 mM imidazole elution fractions were pooled and dialyzed against 1× phosphate-buffered saline, 0.01% Igepal, and 1 mM dithiothreitol before further use.

To analyze the role of Gemin5 in U snRNP assembly, an immunoprecipitated SMN complex in the presence and absence of recombinant SmB/D3 was incubated with 0.75 μg of recombinant Gemin5 in an assembly assay using ^{32}P-labeled U1 snRNA (15). Sm core formation was assessed by native gel electrophoresis after addition of buffer containing 4 M urea, 12.5 mg/ml heparin, and 16% glycerol as described (8).

Glycerol Gradient Centrifugation of the in Vitro Reconstituted SMN Complex—*In vitro* reconstituted, ^{35}S-labeled SMN complexes were diluted with 1× phosphate-buffered saline to a final volume of 250 μl, layered on a 10–30% glycerol gradient, and centrifuged for 16 h in an SW60Ti rotor (4 °C, 24,000 rpm, Beckman Coulter Optima L-80 XP). Gradients were manually harvested in 22 fractions from top to bottom.

Yeast Two-hybrid Interaction Trap—All yeast manipulations and the yeast two-hybrid interaction trap assay were carried out as described (20, 21). In brief, to analyze all possible protein-protein interactions within the SMN complex, the complex components (*i.e.* SMN, Gemins 2–8, and unrip) were subcloned into pEG202 vector (Clontech), which allows constitutive expression of LexA fusion proteins (DNA-binding domain, bait) and into the galactose inducible vector pJG4-5 to express B42-HA-fusion proteins (activation domain, prey). Plasmids were co-transformed in *Saccharomyces cerevisiae* strain EGY48 harboring the reporter plasmid pSH18–34 (Invitrogen). Positive clones were selected by blue staining of colonies on X-gal containing galactose/URA$^-$HIS$^-$TRP$^-$ dropout plates. Yeast clones were grown on glucose-containing dropout plates as a control.

RESULTS

Interactions of Recombinant SMN Complex Components in Vitro—In a first series of experiments we used GST-tagged recombinant proteins to analyze the interactions with *in vitro* translated ^{35}S-labeled components of the SMN complex, produced in reticulocyte lysate (Fig. 1*A*). In our hands, only SMN and Gemins 2 and 6–8 could be produced in *Escherichia coli* as soluble, full-length proteins, whereas Gemins 3–5 and unrip were degraded and/or misfolded. Robust binding of SMN, Gemins 2, 3, and 8 to immobilized GST-SMN was detected (*lane 3*), but no significant association of any translated protein was observed to GST alone (*lane 2*). An identical pattern could be observed when a complex composed of GST-SMN and Gemin2 was used as bait (*lane 4*), suggesting that Gemin2 binds predominantly to SMN. Consistently, GST-Gemin2 bound SMN only but no other protein of the SMN complex in this assay (*lane 5*). Next, we analyzed GST-Gemins 6 and 7, which have convincingly been shown to form a heterodimer (13). In

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

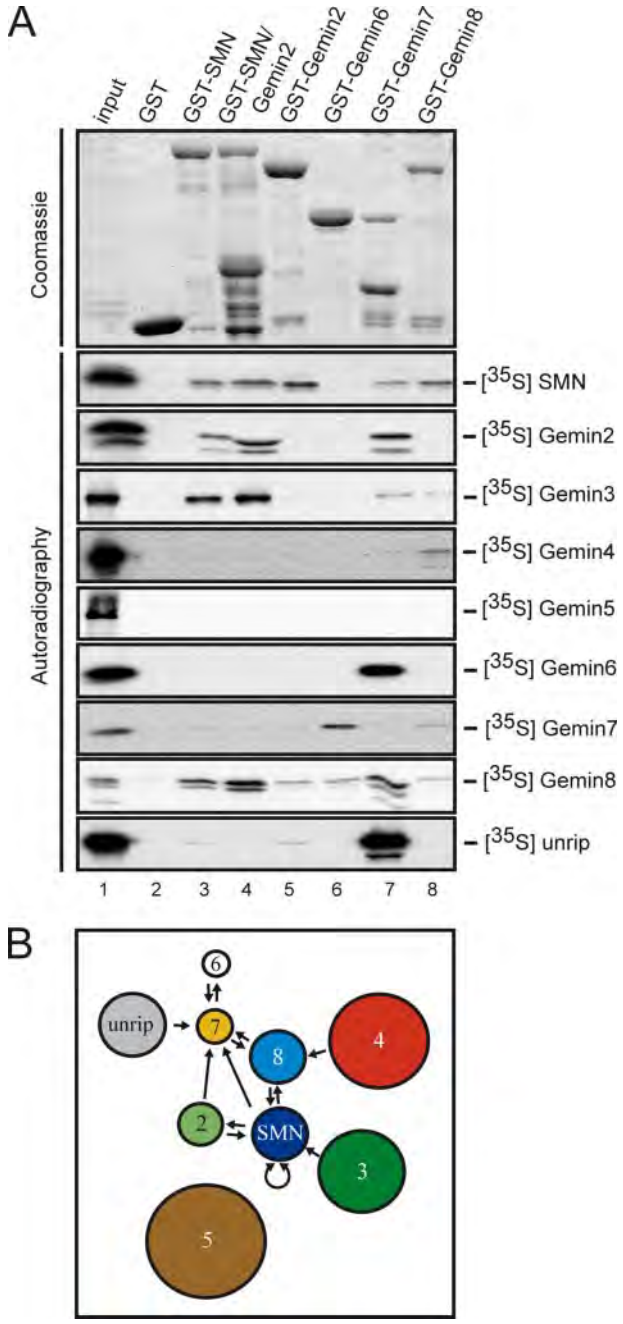


FIGURE 1. **GST pulldown assay with SMN complex components.** *A*, GST-fusion proteins (*i. e.* SMN, Gemin2, GST-SMN/His_6-Gemin2, and Gemins 6–8) were immobilized on glutathione-Sepharose and incubated with *in vitro* translated, ^{35}S-labeled components of the SMN complex. Bound proteins were eluted by boiling in loading buffer, resolved by SDS-PAGE, visualized by Coomassie staining (*upper panel*, *lanes 3–8*), and by autoradiography of the dried gel (*lower panels*, *lanes 3–8*). 10% input of the ^{35}S-labeled proteins is shown in *lane 1*. GST was used as a control (*lane 2*). *B*, schematic of all resulting interactions shown in *A*.

agreement with earlier findings, we observed binding of translated Gemin7 to GST-Gemin6 and vice versa (*lanes 6* and *7*). Whereas Gemin6 appears to bind to no other component, strong binding of Gemin7 to Gemins 2, 6, and 8 and unrip, and weak binding to SMN could be detected (*lane 7*). In the same assay, GST-Gemin8 bound to SMN as well as Gemin4 (*lane 8*). In conclusion, this first experimental strategy suggests that Gemin8 links the proteins Gemins 6 and 7 and unrip with the remainder of the SMN complex (see also supplemental Fig. 1 for additional binding assays that support this view). The data obtained from this set of experiments are depicted in Fig. 1*B*.

A Co-immunoprecipitation Strategy Reveals Interactions of Gemins 3, 4, and 5—Three proteins of the SMN complex, namely Gemins 3, 4, and 5 could not be produced in *E. coli* as full-length proteins in a properly folded manner and in sufficient amounts to perform the experiments described above. Hence, the interactions of these proteins were analyzed by co-immunoprecipitation of co-translated and ^{35}S-labeled components. In a first set of experiments, the interactions of Gemin3 were studied in detail (Fig. 2*A*). For this, Gemin3 was co-translated with single proteins of the SMN complex, tagged with either an HA-epitope (SMN, Gemins 2, 4, and 6–8 and unrip) or a V5 epitope (Gemin5). The co-translations were then precipitated via the respective tag and analyzed by autoradiography (see *lanes 4–11* for inputs). Using this assay, we observed reproducible association of Gemin3 with SMN, Gemins 2, 4, and 5 (*lanes 12–15*). No other component of the SMN complex was co-precipitated in this experiment, and neither antibody precipitated Gemin3 alone (Fig. 2*A*, *lanes 2*, *3*, and *16–19*), implying specific interactions.

Next, binding of Gemin4 with co-translated proteins of the SMN complex was determined by the same strategy (see Fig. 2*B*, for inputs see *lanes 4–11*). Only Gemin3 and to a lesser extent Gemin5 could be co-precipitated with Gemin4 (compare *lanes 14* and *15* with *lanes 12*, *13*, and *16–19*). Finally, V5-tagged Gemin5 was analyzed (Fig. 2*C*). Weak but reproducible binding was observed to Gemin2 and Gemin4 (*lanes 12* and *14*), whereas the other components failed to interact in the same assay (*lanes 11*, *13*, and *15–18*). To further validate the unexpected interaction between Gemin2 and Gemin5, expression levels of Gemin2 were reduced by RNAi in HeLa cells. As shown in Fig. 2*E*, transfection of two different 21-nucleotide-long siRNA duplices complementary to the Gemin2 mRNA but not a control siRNA reduced Gemin2 protein levels. SMN and Gemin7 levels were unaffected, whereas Gemin5 expression was reduced marginally (Fig. 2*E*, *left panel*, compare *lane 1* with *lanes 2* and *3*). Extracts prepared from these cells were then immunoprecipitated with an anti-SMN monoclonal antibody and co-precipitated components detected by Western blotting. Whereas Gemins 2, 5, and 7 could be co-precipitated with SMN from extracts derived from control cells, the association of SMN with Gemin2 and Gemin5 was markedly reduced in the Gemin2 knockdown extracts (by at least 80%), while Gemin7 association was unaltered (Fig. 2*E*, *right panel*, compare *lane 1* with *lanes 2* and *3*). These data illustrate that Gemin5 is tethered to the SMN complex predominantly via an interaction with Gemin2 *in vivo*. This conclusion was further supported by additional co-immunoprecipitation experiments and bacterial co-expression studies with truncation mutants of Gemin5 (supplemental Fig. 2). These confirmed the results obtained so far and further revealed the C-terminal half of Gemin5 (amino acids 721–1508) as the binding site for Gemin2 (supplemental Fig. 2).

Analysis of Protein Contacts within the SMN Complex by a Yeast Two-hybrid System—The experiments described above revealed specific interactions of individual SMN complex com-

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

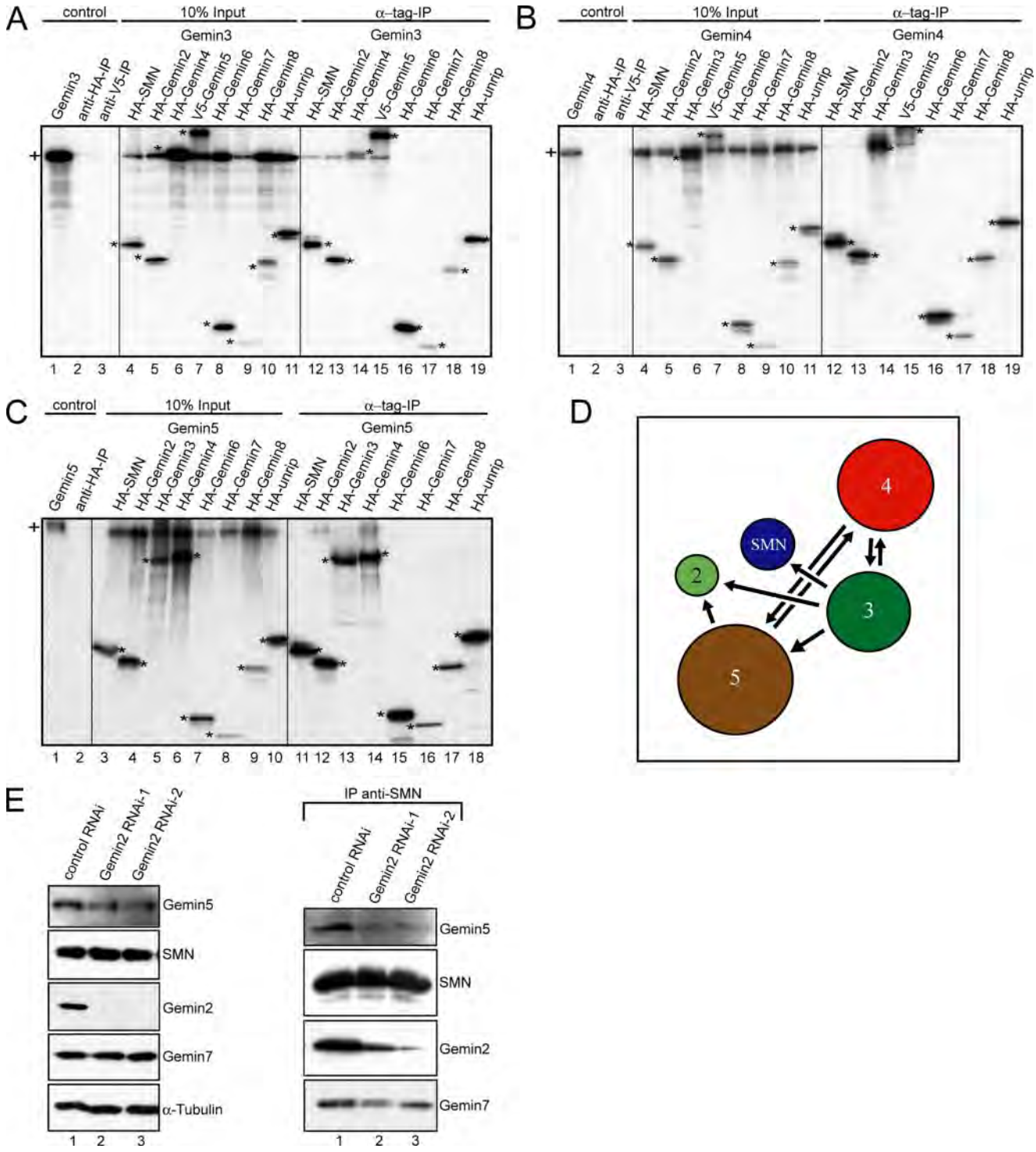


FIGURE 2. **Co-immunoprecipitation assay with Gemins 3–5.** HA-tagged versions of all SMN complex components were co-translated *in vitro* with His_6-tagged Gemin3 (*A*), Gemin4 (*B*), or Gemin5 (*C*) and labeled with [^{35}S]methionine, respectively. Proteins were co-immunoprecipitated with either an α-HA (SMN, Gemins 2–4, Gemins 6–8, and unrip) or an α-V5 antibody (Gemin5) and resolved by SDS-PAGE. Labeled proteins were visualized by autoradiography. Immunoprecipitated bait proteins are marked by *asterisks* and prey proteins by *plus signs*. *A*, co-immunoprecipitations of Gemin3 with SMN complex components (*lanes 12–19*). As a control ^{35}S-labeled His_6-Gemin3 was immunoprecipitated by α-HA (*lane 2*) and α-V5 antibody (*lane 3*). 10% input is shown (*lanes 1* and *4–11*). *B*, co-immunoprecipitations of Gemin4 with SMN complex components (*lanes 12–19*). As a control ^{35}S-labeled His_6-Gemin4 was immunoprecipitated by α-HA (*lane 2*) and α-V5 antibody (*lane 3*). 10% input is shown (*lanes 1* and *4–11*). *C*, co-immunoprecipitations of Gemin5 with SMN complex components (*lanes 11–18*). As a control ^{35}S-labeled V5-Gemin5 was immunoprecipitated by α-HA antibody (*lane 2*). 10% input is shown (*lanes 1* and *3–10*). *D*, schematic of all resulting interactions shown in *A–C*. *E*, *left panel*: Gemin5 is incorporated into the SMN complex predominantly via Gemin2. HeLa cells were transfected with Gemin2 siRNAs (RNAi-1, *lane 2* and RNAi-2, *lane 3*). Silencing efficiency was assessed 52 h post-transfection by Western blotting with the indicated antibodies. An unrelated siRNA was used as control (*lane 1*). *E*, *right panel*: SMN complex was immunoprecipitated from extract (*E*, *left panel*) with 7B10 anti-SMN antibody, resolved on an 8–20% gradient gel, and analyzed by Western blotting with specific antibodies.

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

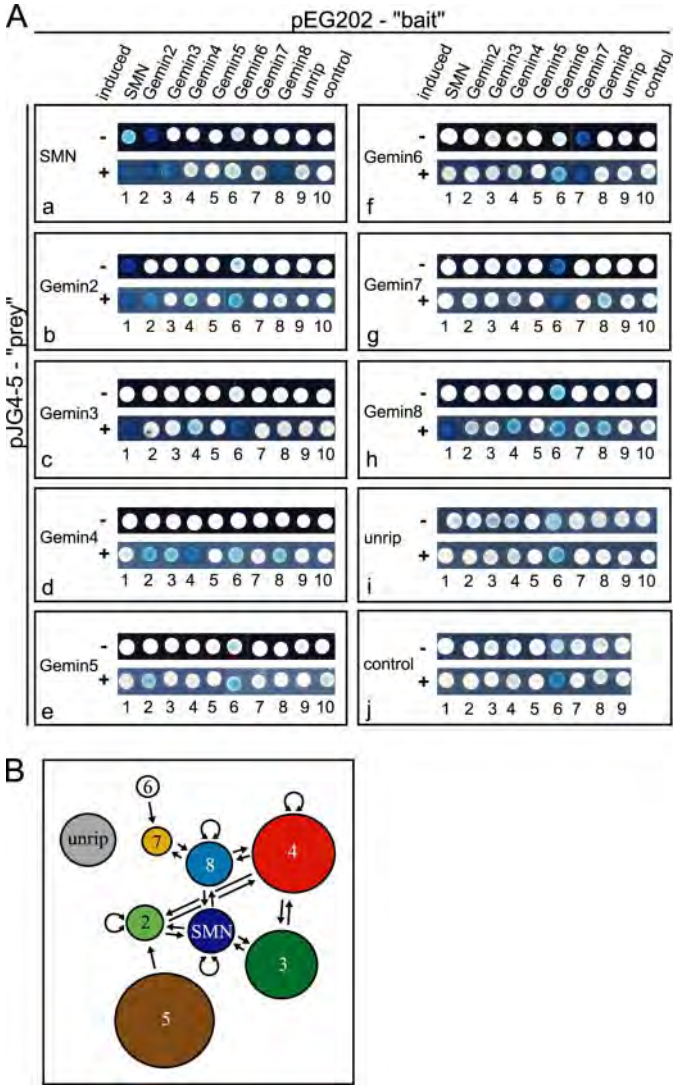


FIGURE 3. **Yeast two-hybrid interaction trap assay with SMN complex components.** *A*, all components of the SMN complex (*i. e.* SMN, Gemins 2–8, unrip) were transformed into *S. cerevisiae* as galactose-inducible prey plasmids pJG4-5 (*panels a–j*). A single tested clone of each transformation was co-transformed with bait plasmids pEG202, encoding all SMN complex components (*lanes 1–9* in each panel), respectively. As a control, yeast clones with complex components subcloned in pJG4-5 were co-transformed with the empty vector pEG202 (*lane 10*). Positive clones were selected by blue staining on galactose/URA$^-$HIS$^-$TRP$^-$ X-gal dropout plates (*panels a–j*, *lower row*). As a control, clones were tested on glucose containing dropout plates (*upper row*). *B*, schematic of all resulting interactions shown in *A*.

ponents to each other *in vitro*. We next set forth to analyze interactions within the SMN complex *in vivo* using a yeast two-hybrid assay (20, 21). For this purpose, a panel of "bait"-vector constructs were created, encoding for fusions of the DNA-binding domain of the transcription factor LexA and either SMN, Gemins 2–8, or unrip (Fig. 3*A*, *lanes 1–9*, *lane 10* shows the LexA control without a fusion partner). These plasmids were co-transformed with "prey"-vectors encoding for fusions of the activation domain B42 and the same proteins as in the bait vectors (*panels a–i*, *panel j* shows the B42 control). Binding of prey to bait fusion proteins was then assessed in a β-galactosidase enzyme assay upon induction of prey fusion expression.

Interactions of SMN expressed from the prey-vector are shown in Fig. 3*A*, *panel a*. In agreement with earlier data and those shown in Figs. 1 and 2, SMN interacts with itself, Gemin2, Gemin3, and Gemin8 (*panel a*, *lanes 1–3* and *8*). Importantly, these interactions can also be observed when SMN is expressed as bait protein (see *panels a–c* and *h*, *lane 1*). The weak β-galactosidase activity observed in yeast co-transformed with SMN and Gemin6 most likely does not reflect a true interaction, as the latter fusion protein is auto-activating (*panels a–j*, *lane 6* and data not shown). In the same system, we also observed binding of Gemin3 with Gemin4 and Gemin7 with Gemins 6 and 8. Furthermore, we could detect novel protein contacts of Gemins 4–8, and Gemin 2–4 as well as Gemin5 (compare *panels a–i*, *lanes 1–9*). Note that certain interactions such as SMN and Gemin2 (*panel a*, *lane 2*; *panel b*, *lane 1*), Gemins 6 and 7 (*panel f*, *lane 7*), and SMN oligomerization (*panel a*, *lane 1*) can be detected prior to induction, implying very strong binding. Taken together, this strategy supports most data obtained *in vitro* (summarized in Fig. 3*B*). However, binding partners of unrip that could readily be detected *in vitro* were not identified in the yeast system possibly due to misfolding in this host.

A Novel Reconstitution System Reveals the Framework of the SMN Complex—The studies described thus far focused on interactions of single, isolated components. We next developed an experimental strategy to address whether the observed interactions also occur in the context of the SMN complex and how they contribute to its overall architecture. To this end, all proteins of the complex, including HA-tagged SMN were simultaneously translated and radiolabeled in rabbit reticulocyte lysate. Following immunoprecipitation by anti-HA antibodies, the proteins were subjected to SDS-PAGE and autoradiography. Strikingly, this led to an efficient co-precipitation of Gemins 2–4 and 6–8, whereas no proteins were precipitated from a mixture lacking tagged SMN (Fig. 4*A*, *lanes 8* and *9*, the inputs are depicted in *lanes 1* and *2*). Only Gemin5 and unrip are incorporated less efficiently into the complex and therefore co-precipitate weakly. Thus, a significant proportion of the SMN complex forms when all core components are co-translated. This observation is further supported by gradient centrifugation studies shown in Fig. 5. Co-translated SMN complex components sediment very similar to the endogenous complex (Fig. 5*A*, *lanes 12–22* and data not shown), indicating the formation of a stable entity. In contrast, when SMN is omitted from the co-translation reaction the sedimentation behavior of all Gemins is in a lower molecular weight range (compare Fig. 5, *A* and *B*, *lanes 12–22*; note that lower migrating subcomplexes or single translated proteins can be observed in both translations, Fig. 5, *A* and *B*, *lanes 2–11*). Next, co-translation experiments were performed in the absence of selected single components. Consistent with the *in vitro* binding assays shown above all other complex components could still be co-immunoprecipitated with HA-SMN when Gemin2 was omitted from the co-translation (Fig. 4*A*, *lane 10*). Gemin2 is therefore a peripheral protein that is recruited to the complex predominantly via SMN. In contrast, co-translation in the absence of Gemin3 led to the formation of a complex that also lacked Gemin4 (*lane 11*), whereas omission of Gemin4 had no effect on the incorporation of the other proteins into the complex (*lane 12*). These data are consistent with the notion that Gemin3 and 4 interact and associate with the SMN complex predominantly via

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

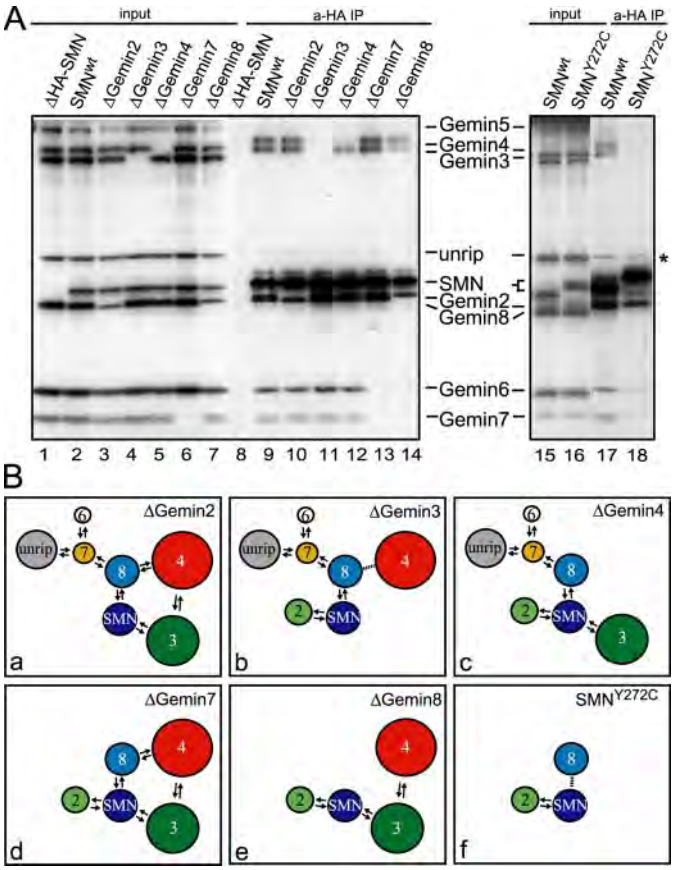


FIGURE 4. ***In vitro* reconstitution of the entire SMN complex in reticulocyte lysate.** *A, left panel*: all SMN complex components were ^{35}S-labeled and co-translated as His_6 fusion proteins together with HA-tagged SMN in one TNT reticulocyte lysate reaction as indicated. The resulting SMN complexes were co-precipitated by an α-HA antibody (*lanes 9–14*). As a control all SMN complex components were co-translated without HA-SMN and immunoprecipitated as above (*lane 8*). 10% input is shown (*lanes 1–7*). *A, right panel*: all SMN complex components were ^{35}S-labeled and co-translated as His_6 fusion proteins together with HA-tagged SMN^{wt} or SMN^{Y272C} and co-immunoprecipitated (*lanes 17* and *18*). 10% input is shown (*lanes 15* and *16*). An unspecific band is indicated by an *asterisk*. *B*, schematic of all precipitated partial SMN complexes. *Panels a–e*, co-precipitations of reconstituted SMN complexes lacking the indicated components. *Panel f*, co-precipitated reconstituted SMN^{Y272C} complex.

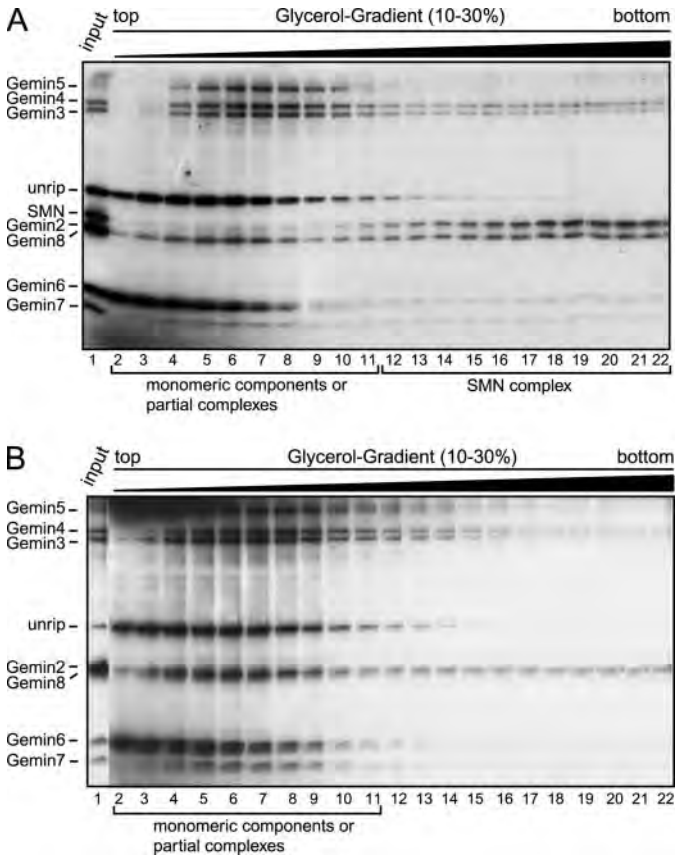


FIGURE 5. **Analysis of the *in vitro* translated SMN complexes by gradient centrifugation.** All SMN complex components were ^{35}S-labeled and co-translated as His_6 fusion proteins together with (*A*) or without (*B*) HA-tagged SMN in the TNT reticulocyte lysate system. The reactions were then layered on 10–30% glycerol gradients. Gradients were harvested manually from top to bottom in 22 fractions. A sample of each fraction was analyzed by 12% SDS-PAGE and autoradiography of the dried gel (*lanes 2–22*). 10% input is shown (*lane 1*).

Gemin3 (22, 23). We note, however, a weak but reproducible binding of Gemin4 to the SMN complex in the absence of Gemin3 (*lane 11* and data not shown). This may indicate that Gemin4 also associates weakly with another protein, presumably Gemin8 as has been observed in the *in vitro* binding experiments and in the yeast interaction assay (Figs. 1 and 3). Co-translation in the absence of Gemin7 strongly affected binding of Gemin6 and unrip (Fig. 4*A*, *lane 13*), and omission of Gemin8 led to the specific loss of Gemins 6 and 7 and unrip association to the complex (*lane 14*). These data in conjunction with those in the preceding sections indicate that Gemin8 constitutes the bridge between the Gemin6/Gemin7/unrip trimer and the rest of the SMN complex. These findings are summarized in Fig. 4*B* (*panels a–e*).

The neuromuscular disorder SMA is caused by reduced levels of or production of mutant SMN protein. It was therefore interesting to analyze whether pathogenic mutations in SMN would interfere with its incorporation into the complex. To address this, we chose the well characterized C-terminal mutation Y272C that has previously been reported to abolish SMN-oligomerization and Sm protein binding (7, 24, 25). Strikingly, when the Y272C mutant was co-translated with all SMN complex components only Gemin2 could efficiently be co-precipitated, whereas binding to all other components was severely impaired (Fig. 4*A*, *lane 18*). Reduced or aberrant formation of the SMN complex may hence be a biochemical defect in patients carrying this and possibly other mutations.

An Essential Role for Gemin5 in the Assembly Reaction—The assembly map established above places Gemin5 in the periphery of the complex with an interaction to Gemin2 (and possibly Gemin4). As Gemin5 appears to be unimportant for the integrity of the remaining SMN complex, we reasoned that it might have a functional role in the U snRNP assembly reaction. To explore this possibility, we devised a strategy to generate SMN complexes lacking Gemin5. This was achieved by treatment of immobilized purified human SMN complex with increasing concentrations of NaCl. Although most components of the complex remain stably associated up to 1.5 M salt, SmB/B′, D3, and Gemin5 dissociate almost quantitatively at 500 mM and 1.5 M NaCl, respectively (see Fig. 6*A*, *left panel* for a silver-stained gel of the purified complex, *right panel* for Western blots of the same samples with various antibodies). Native and Gemin5/

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

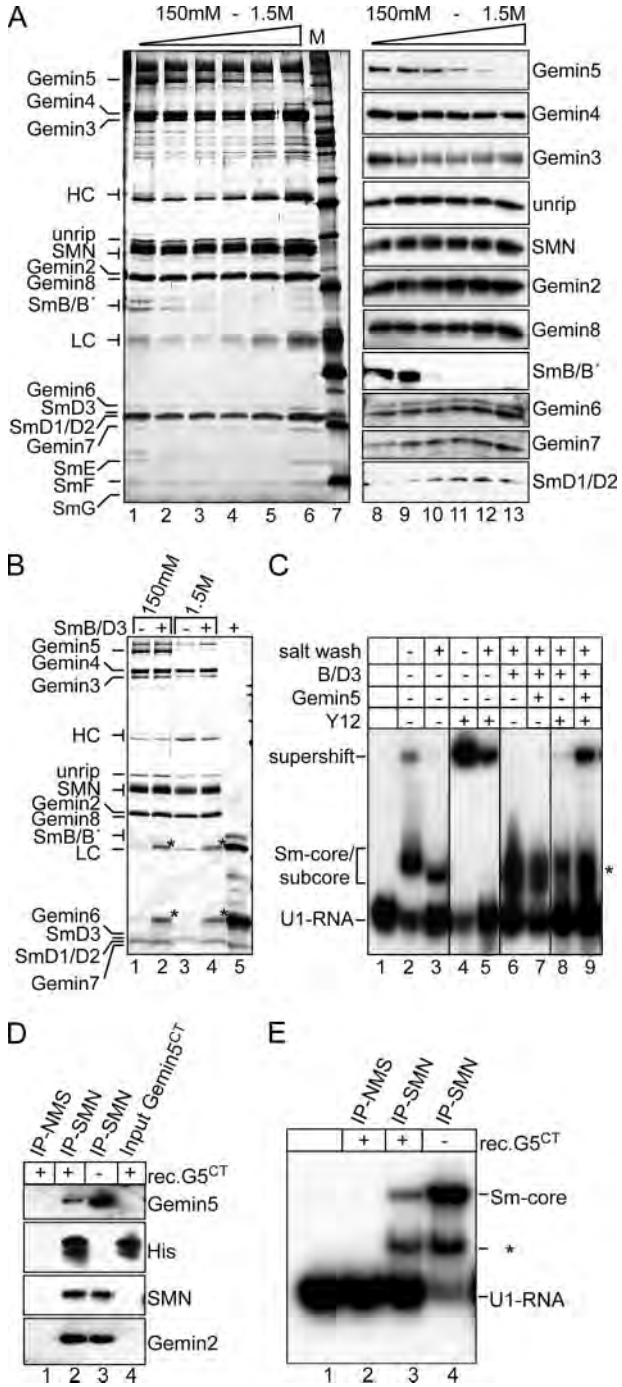


FIGURE 6. **Gemin5 is a peripheral component of the SMN complex but essential for U1 snRNP biogenesis.** *A*, *left panel*: SMN complex was immunoprecipitated from HeLa cytosol. Immobilized complex was washed with buffer containing rising salt concentrations (150 mM, 250 mM, 500 mM, 750 mM, 1 M, and 1.5 M NaCl). Complex was eluted from protein G-Sepharose by SDS-sample buffer, resolved on an 8–20% SDS-polyacrylamide gradient gel, and visualized by silver staining (*lanes 1–6*). *A*, *right panel*: immunoprecipitated complexes from *A* were resolved on an 8–20% gradient gel and analyzed by Western blotting with specific antibodies as indicated. *B*, SMN complex was immunoprecipitated from HeLa cytosol and washed with buffer containing 150 mM NaCl (*lanes 1* and *2*) or 1.5 M NaCl (*lane 3* and *4*). *Lanes 2* and *4* indicate complexes that were reconstituted with purified recombinant B/D3. *Asterisks* indicate bound recombinant SmB/D3. *Lane 5* shows total input of used recombinant B/D3 heterodimer. *C*, Gemin5 is an essential factor in U snRNP biogenesis. Complexes used in *B* were incubated with ^{32}P-labeled U1 snRNA in an assembly assay and analyzed by native gel electrophoresis for Sm core formation (*lanes 2*, *3*, and *6*). *Lane 7* shows the assembly reaction upon addition of 0.75 μg of purified Gemin5. Specificity of Sm core formation was assessed by incubation with Sm-specific antibody Y12 to the assembly reaction (*lanes 4*, *5*, *8*, and *9*). *Lane 1* shows U1 snRNA in the absence of SMN complex. *D*, immunopurified SMN complex was incubated with recombinant His-tagged Gemin5^{CT} (purified as described in supplemental Fig. 2). Mock-treated SMN complex and a normal mouse serum (*NMS*) immunoprecipitation form HeLa cytosolic extract incubated with Gemin5^{CT} served as a control. Western blotting confirmed the incorporation of His-tagged Gemin5^{CT} at the expense of endogenous full-length Gemin5 (compare *lanes 2* and *3*). *Lane 4* shows 10% of histidine-tagged, recombinant Gemin5^{CT}. Indicated proteins were detected with monospecific antibodies after SDS-PAGE and blotting. *E*, complexes used in *D* were incubated with ^{32}P-labeled U1 snRNA in an assembly assay and analyzed by native gel electrophoresis for Sm core formation (*lanes 2–4*). *Lane 1* shows U1 snRNA in the absence of SMN complex.

SmB/D3-depleted SMN complexes were then tested for their ability to promote formation of the Sm core domain on radiolabeled U1 snRNA. Core formation was assessed by native gel electrophoresis. The native SMN complex efficiently promoted the formation of the Sm core domain, as evident by the slower migration of the U1 snRNA in the native gel and further retardation upon addition of the anti-Sm antibody Y12 (Fig. 6*C*, *lanes 2* and *4*). Interestingly, the salt-washed complex lacking Gemin5, B/B′, and D3 also promoted the formation of an RNP that could be super-shifted with antibody Y12 (*lanes 3* and *5*). This Sm-containing RNP most likely corresponds to the subcore particle that has been shown previously to form *in vitro* when isolated Sm proteins D1, D2, E, F, and G (but not B/B′ and D3) are incubated with U1 snRNA (3). Next, we asked whether Sm core domain formation could be restored upon re-addition of the missing Sm proteins B/B′ and D3. Surprisingly, although a recombinant B/D3 heterooligomer was efficiently bound by the salt washed complex (Fig. 6*B*, *lane 4*, indicated by *asterisks*), no Sm core formation could be observed. Note that the retardation of U1 snRNA in the native gel could not be further increased upon addition of Y12 and hence was not caused by Sm proteins (Fig. 6*C*, *lanes 6* and *8*, indicated by an *asterisk*). Thus, an SMN complex containing the complete set of Sm proteins but lacking Gemin5 can neither promote subcore formation nor Sm core assembly. To address the question whether the functionality of this defective SMN complex could be restored upon re-addition of Gemin5, this protein was expressed in insect cells and purified to homogeneity (supplemental figure 3). Strikingly, addition of recombinant Gemin5 to the assembly mixture partially restored the assembly reaction, as evident by the formation of a complex that could be super-shifted with Y12-antibody (Fig. 6*C*, *lanes 7* and *9*). In contrast, upon incorporation of a truncated version consisting of the C-terminal part of the Gemin5 protein, the SMN complex was inactive (Fig. 6 D and E, asterisk indicates unspecific RNA shift). In sum, these data are consistent with the notion that Gemin5 mediates an essential step in the assembly of U snRNPs, most likely in the transfer of Sm proteins onto the U snRNA.

DISCUSSION

The function of the multisubunit SMN complex relies on the ordered interplay of the nine components SMN, Gemins 2–8, and unrip. Previous studies have revealed a great number of interactions among these proteins using various *in vitro* and *in vivo* assays (12–16). However, as most interaction assays are error prone, it is not surprising that conflicting data regarding

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

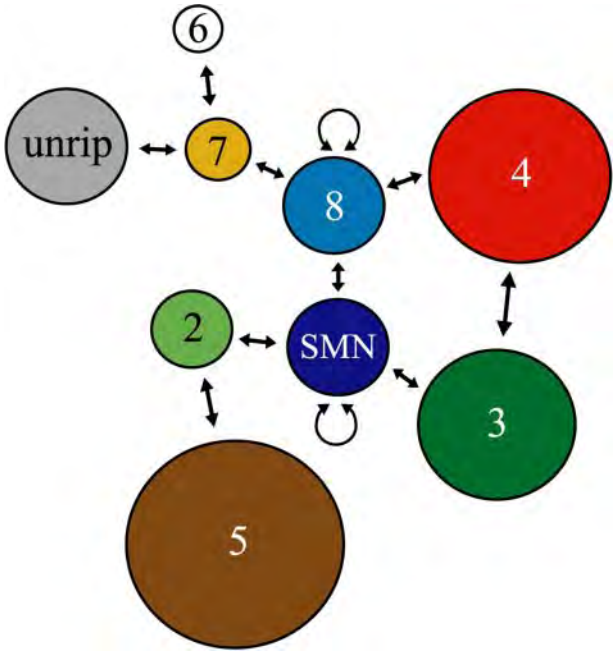


FIGURE 7. **Consensus interaction map of the human SMN complex.** Schematic of all interactions within the SMN complex obtained by the methods described in this work. Only protein-protein interactions that were observed in at least two independent experimental systems were considered.

the interactions within the core SMN complex have been obtained. To circumvent this problem, we have delineated an interaction map that relies on several different and independent interaction assays. Only protein contacts that could be observed in at least two independent experimental systems were considered in the construction of an interaction network. Our studies have revealed a framework of the SMN complex with SMN, Gemin7, and Gemin8 as its backbone. These three proteins provide a binding platform for the other components of the complex via multiple interactions: SMN binds to Gemin2 (26), Gemin3 (27, 28), and Gemin8. Gemin8, in turn, interacts with Gemins 4 and 7. Finally, Gemin7 recruits unrip and Gemin6 via direct interactions (14, 15, 29). Fig. 7 shows a consensus interaction map derived from published data (summarized in Refs. 12 and 13) and from results obtained in this report. Whereas most components of the SMN complex form a very stable structure, the peripheral protein Gemin5 can be dissociated upon treatment with high salt. This finding suggests its incorporation into the complex via weak interactions mediated by Gemin2 and possibly Gemin4. Based on data presented here and in previous reports, we are now confident that we understand the architecture of the SMN complex. Nevertheless, we cannot exclude the possibility that some protein-protein contacts might have escaped our detection or were erroneously excluded based on the stringent criteria described above. In addition, with the availability of the experimental strategies detailed above, we are now in a position to map the binding sites of other peripheral components of the SMN complex. These include substrate Sm/LSm proteins (12, 13), the nuclear import factors importin β and snurportin (30, 31), and coilin, a binding partner of SMN in nuclear coiled bodies (32). These studies will provide additional mechanistic insights into the U snRNP biogenesis cycle.

Most cases of SMA are caused by the reduced production of SMN protein. However, some cases of disease are also known, where patients express mutant versions of the protein. The majority of pathogenic mutations (deletions and missense mutations) cluster in the C-terminal domain responsible for SMN oligomerization and/or interaction with Sm proteins (7, 24, 25). Here, we provide evidence that the pathogenic SMN mutation Y272C is incorporated inefficiently into the SMN complex. This effect might be due to the inability of SMN to form oligomers and/or the disability to bind other proteins such as Gemin8, which contacts amino acids 242–294 of SMN, where the missense mutation Y272C is located (see supplemental Fig. 1*B*, *lanes 3* and *6*). Future studies based on the novel reconstitution assay reported here will be required to resolve the issue whether other pathogenic SMN mutations result in the reduced formation of functional SMN complexes and if this is the underlying cause for the development of the disease SMA in patients.

Conflicting data regarding the function of Gemin5 have recently been reported (33–35). In two reports RNA interference was used to show that Gemin5 is dispensable for the assembly reaction (33, 34). In contrast, evidence for an essential role of Gemin5 in the formation of U snRNPs has been provided in another recent report (35). In this latter study, Gemin5 was shown to bind snRNAs in a sequence specific manner, suggesting its role in the recruitment of these to the SMN complex. Our finding that an SMN complex lacking Gemin5 fails to promote U snRNP assembly is consistent with this notion. However, we do not rule out additional functions of Gemin5 in the assembly reaction such as transfer of Sm proteins onto the U snRNA. The availability of Gemin5 expressed in insect cells will allow us to address these questions.

Acknowledgments—We are grateful to I. Berger, T. J. Richmond, L. Pellizzoni, G. Morris, and F. Graesser for providing reagents and to A. Farwick and E. Dinkl for help and expert technical assistance.

REFERENCES

1. Seraphin, B. (1995) *EMBO J.* **14,** 2089–2098
2. Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., and Luhrmann, R. (1995) *EMBO J.* **14,** 2076–2088
3. Raker, V. A., Plessel, G., and Luhrmann, R. (1996) *EMBO J.* **15,** 2256–2269
4. Raker, V. A., Hartmuth, K., Kastner, B., and Luhrmann, R. (1999) *Mol. Cell. Biol.* **19,** 6554–6565
5. Achsel, T., Stark, H., and Luhrmann, R. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3685–3689
6. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., and Zeviani, M. (1995) *Cell* **80,** 155–165
7. Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997) *Nat. Genet.* **16,** 265–269
8. Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001) *Nat. Cell Biol.* **3,** 945–949
9. Meister, G., and Fischer, U. (2002) *EMBO J.* **21,** 5853–5863
10. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002) *Science* **298,** 1775–1779
11. Pillai, R. S., Grimmler, M., Meister, G., Will, C. L., Luhrmann, R., Fischer, U., and Schumperli, D. (2003) *Genes Dev.* **17,** 2321–2333
12. Meister, G., Eggert, C., and Fischer, U. (2002) *Trends Cell Biol.* **12,** 472–478
13. Gubitz, A. K., Feng, W., and Dreyfuss, G. (2004) *Exp. Cell Res.* **296,** 51–56
14. Carissimi, C., Baccon, J., Straccia, M., Chiarella, P., Maiolica, A., Sawyer,

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

A., Rappsilber, J., and Pellizzoni, L. (2005) *FEBS Lett.* **579,** 2348–2354
15. Grimmler, M., Otter, S., Peter, C., Muller, F., Chari, A., and Fischer, U. (2005) *Hum. Mol. Genet.* **14,** 3099–3111
16. Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J., and Pellizzoni, L. (2006) *J. Biol. Chem.* **281,** 8126–8134
17. Grimmler, M., Bauer, L., Nousiainen, M., Korner, R., Meister, G., and Fischer, U. (2005) *EMBO Rep.* **6,** 70–76
18. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J., and Nagai, K. (1999) *Cell* **96,** 375–387
19. Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004) *Nat. Biotechnol.* **22,** 1583–1587
20. Gyuris, J., Golemis, E., Chertkov, H., and Brent, R. (1993) *Cell* **75,** 791–803
21. Golemis, E. A., Serebriiskii, I., Finley, R. L., Kolonin, M. G., Gyuris, J., and Brent, R. (1999) *Current Protocols in Molecular Biology*, Vol. 4, pp. 20.21.21–20.21.40, Wiley, New York
22. Meister, G., Buhler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000) *Hum. Mol. Genet.* **9,** 1977–1986
23. Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000) *J. Cell Biol.* **148,** 1177–1186
24. Lorson, C. L., Strasswimmer, J., Yao, J. M., Baleja, J. D., Hahnen, E., Wirth, B., Le, T., Burghes, A. H., and Androphy, E. J. (1998) *Nat. Genet.* **19,** 63–66
25. Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11167–11172
26. Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997) *Cell* **90,** 1013–1021
27. Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999) *J. Cell Biol.* **147,** 1181–1194
28. Campbell, L., Hunter, K. M., Mohaghegh, P., Tinsley, J. M., Brasch, M. A., and Davies, K. E. (2000) *Hum. Mol. Genet.* **9,** 1093–1100
29. Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002) *J. Biol. Chem.* **277,** 31957–31962
30. Narayanan, U., Ospina, J. K., Frey, M. R., Hebert, M. D., and Matera, A. G. (2002) *Hum. Mol. Genet.* **11,** 1785–1795
31. Narayanan, U., Achsel, T., Luhrmann, R., and Matera, A. G. (2004) *Mol. Cell* **16,** 223–234
32. Hebert, M. D., Szymczyk, P. W., Shpargel, K. B., and Matera, A. G. (2001) *Genes Dev.* **15,** 2720–2729
33. Shpargel, K. B., and Matera, A. G. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17372–17377
34. Feng, W., Gubitz, A. K., Wan, L., Battle, D. J., Dostie, J., Golembe, T. J., and Dreyfuss, G. (2005) *Hum. Mol. Genet.* **14,** 1605–1611
35. Battle, D. J., Lau, C. K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006) *Mol. Cell* **23,** 273–279

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

Supplemental Fig. 1. Analysis of binding regions of the interacting proteins within the SMN-complex

Deletion mutants of Gemin8, Gemin7, Gemin5 and Gemin4 were generated by PCR using specific primer sets that amplify the desired DNA-fragments. Truncation mutants of SMN have been described (1). The amplified DNA was subsequently cloned into vectors pGEX6P-1 (GE Healthcare), pET28a (Novagen) or pHA (an N-terminal HA-tag containing derivative of pcDNA3.1, Invitrogen). Expression and purification of GST-tagged fusion proteins was performed as described (2). [^{35}S]-methionine labeled proteins were generated in reticulocyte lysate, using the TNT-T7 Quick Coupled Transcription/Translation System (Promega).

(A) GST-pulldown assay of Gemin8 with SMN-complex components. A GST-fusion of Gemin8 (lane 2-10) and GST as a control protein (lane 11-19) were immobilized on Glutathione Sepharose and incubated with *in vitro* translated, [^{35}S]-labeled components of the SMN-complex. After extensive washing bound proteins were eluted by boiling in 2x loading-buffer, resolved by SDS-PAGE, visualized by Coomassie staining (upper panel, lanes 2-19) and by autoradiography of the dried gel (lower left panel, lanes 2-19). 10% input of the [^{35}S]-labeled proteins is shown in the lower right panel, lanes 20-28. **(B)** In a GST-pulldown assay 2µg of recombinant GST-SMN (*e.g.* SMN^{fl}, SMN$^{90-294}$, SMN$^{1-160}$, SMN$^{1-90}$, SMN$^{242-294}$) was immobilized to Glutathione Sepharose and incubated with *in vitro* translated, [^{35}S]-labeled Gemin8. After extensive washing, bound proteins were eluted by boiling in 2x loading-buffer, resolved by SDS-PAGE and visualized by Coomassie staining (upper panel) and autoradiography (lower panel) of the dried gel (lanes 2-6). 10% input of Gemin8 is shown in lane 9. GST immobilized to Glutathione Sepharose and free Glutathione Sepharose were used as controls (lane 7, 8). **(C)** 2µg of recombinant truncated versions of GST-Gemin8 (*e.g.*. Gemin8^{fl}, Gemin8$^{1-127}$, Gemin8$^{128-242}$), immobilized to Glutathione Sepharose were incubated with *in vitro* translated, [^{35}S]-labeled Gemin4 in a GST-pulldown assay. After extensive washing, proteins were eluted by boiling in 2x loading-buffer, resolved by SDS-PAGE and visualized by Coomassie staining (upper panel) and autoradiography (lower panel) of the dried gel (lanes 2-4). 10% input of Gemin4 is shown in lane 5. GST was used as a control (lane 1). **(D)** In a GST-pulldown assay 2µg of immobilized GST-SMN (lanes 2-4) and GST-Gemin7 (lanes 5-7) were incubated with *in vitro* translated, [^{35}S]-labeled truncated versions of Gemin8 (*i.e.* Gemin8^{fl}, Gemin8$^{1-127}$, Gemin8$^{128-242}$), respectively. After extensive washing, the bound proteins were eluted by boiling in 2x loading-buffer, resolved by SDS-PAGE and visualized by Coomassie staining (upper panel) and autoradiography (lower panel) of the dried gel. 10% input is shown in lanes 11-13. GST was used as a control (lane 8-10).

Supplemental Fig. 2. Bacterial co-expressions of SMN-complex components

(A) To analyze single protein-protein interactions of the SMN-complex, indicated protein-pairs were co-expressed as GST- (SMN, Gemin2 and Gemin7) and his$_6$-fusion proteins (SMN, Gemin2, Gemin5$^{721-1508}$, Gemin6, Gemin8 and unrip) in *E. coli*. The proteins were co-purified on Glutathione Sepharose and analyzed by SDS-PAGE. Protein interactions were detected by silver staining (upper panel, lanes 1-7, GST-fusion proteins are indicated by asterisks) and his$_6$-tagged proteins were additionally verified by α-his$_6$ immunoblotting (lower panel, lanes 1-7), using a monoclonal anti-histidine antibody (Clone His-1, Sigma). As a control all used his$_6$-fusion proteins were co-expressed and purified with GST alone (lanes 8-13). **(B)** The interaction of Gemin2 with truncation mutants of Gemin5 (*i.e.* Gemin5^{fl}, Gemin5$^{1-720}$ and Gemin5$^{721-1508}$) was mapped in a co-immunoprecipitation experiment. HA-Gemin2 was co-translated with Gemin5^{fl}, Gemin5$^{1-720}$ and Gemin5$^{721-1508}$. Proteins were immunoprecipitated with α-HA monoclonal antibody (clone HA.11, BAbCO). After extensive washing, proteins were resolved by SDS-PAGE and visualized by autoradiography of the dried gel (lanes 10-12). Gemin5 and truncated versions thereof were immunoprecipitated by α-HA antibody (lanes 4-6) as a control. 10% input is shown (lanes 1-3 and 7-9).

Supplemental Fig. 3. Ni-NTA purification of His-tagged, baculoviral expressed Gemin5

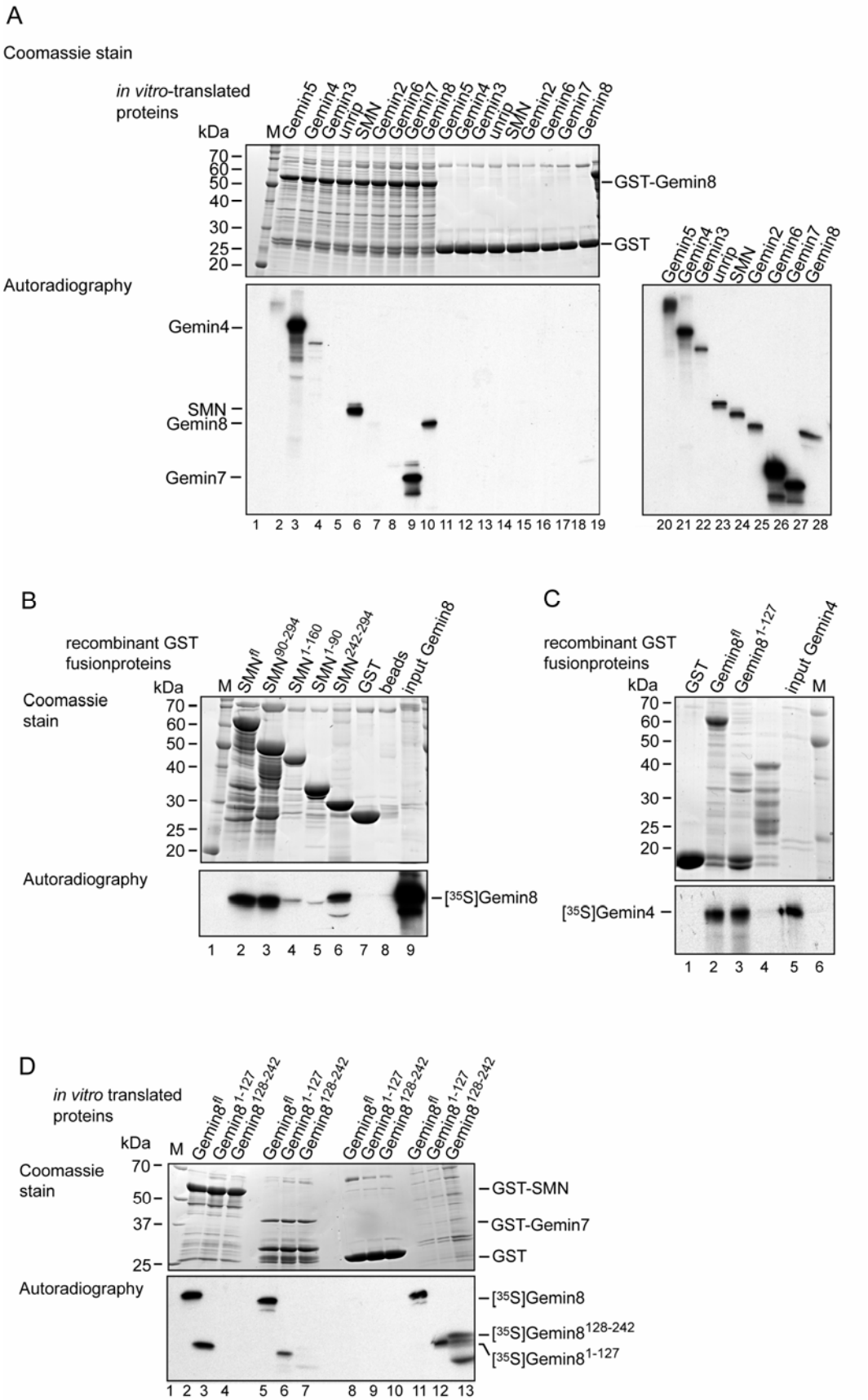
His-tagged Gemin5 protein was generated using the MultiBac insect cell expression system (3). Cells were harvested 5 days post-infection and lysed by detergent (20mM NaHEPES, 200mM NaCl, 0.25% Triton X-100, pH 7.5). Cleared lysate was processed in Ni-NTA purification and eluted fractions were analyzed by 12% SDS-PAGE. Flow-through (ft) is shown in lane 2, elution fractions containing

increasing amounts of imidazole are displayed in lanes 3-13. Fractions holding 60-200mM imidazole were pooled and dialysed (1x PBS, 0.01% Igepal, 1mM DTT) before further use.

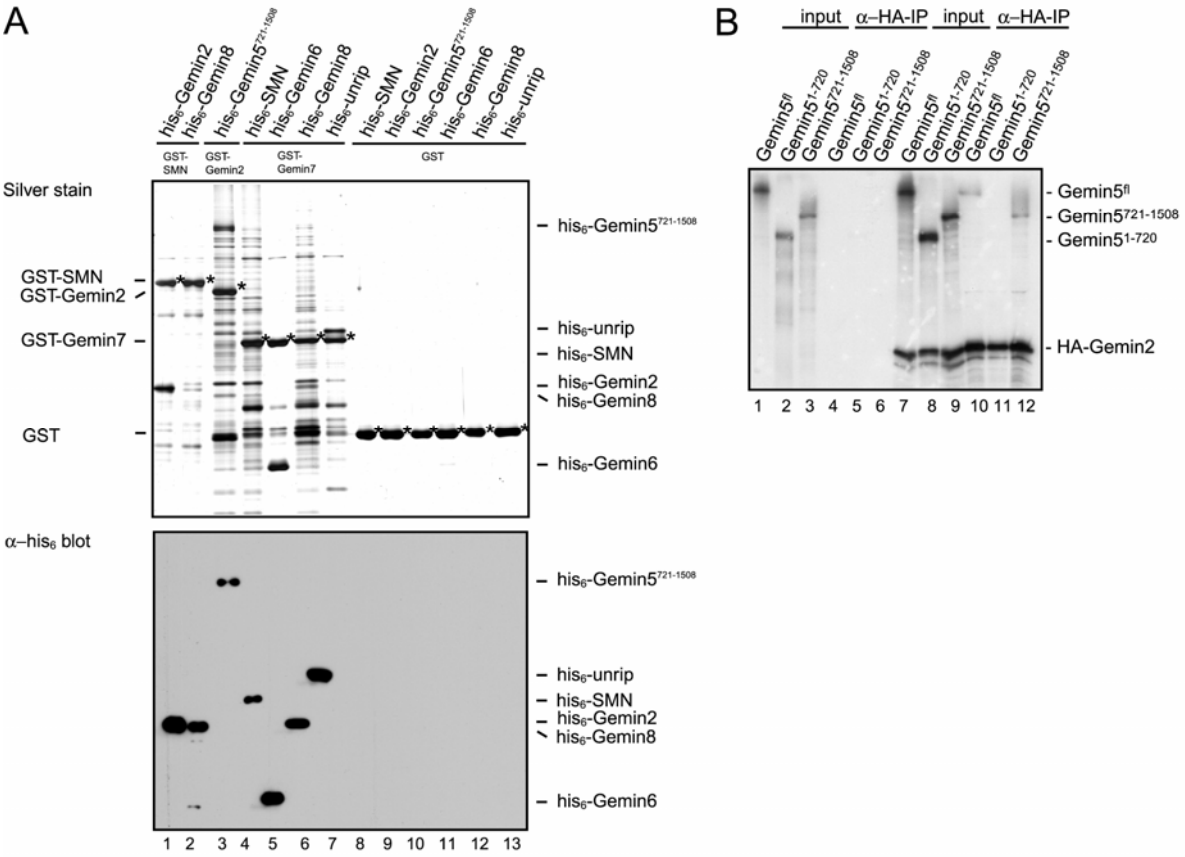
Supplemental References

1. Buhler, D., Raker, V., Luhrmann, R., and Fischer, U. (1999) *Hum Mol Genet* **8**, 2351-2357
2. Grimmler, M., Otter, S., Peter, C., Muller, F., Chari, A., and Fischer, U. (2005) *Hum Mol Genet* **14**, 3099-3111
3. Berger, I., Fitzgerald, D. J., and Richmond, T. J. (2004) *Nat Biotechnol* **22**, 1583-1587

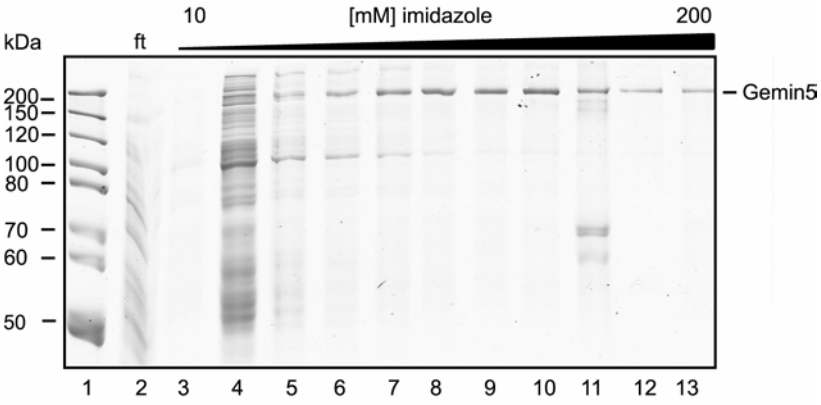
Supplemental Figure 1



Supplemental Figure 2



Supplemental Figure 3



5.3 Mechanistic Aspects of Cellular U snRNP Assembly

Toward an assembly line for U7 snRNPs: interactions of U7-specific Lsm proteins with PRMT5 and SMN complexes

Azzouz TN, Pillai RS, Dapp C, Chari A, Meister G, Kambach C, Fischer U, Schumperli D: **J Biol Chem 2005**, 280:34435-34440.

Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formulation of results:	10 %

An assembly chaperone collaborates with the SMN complex to generate spliceosomal snRNPs

Chari A, Golas MM, Klingenhager M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H, Fischer U: **Cell 2008**, 135:497-509.

Thesis author´s contribution:

Conception:	60 %
Experimental contribution:	80 %
Formulation of results:	80 %

This publication in Cell was also discussed in the press. The press release from Sonja Jülich and Gunnar Bartsch „Zellulärer Maschinenbau“ is appended here as well.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 280, NO. 41, pp. 34435–34440, October 14, 2005
© 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

Toward an Assembly Line for U7 snRNPs

*INTERACTIONS OF U7-SPECIFIC Lsm PROTEINS WITH PRMT5 AND SMN COMPLEXES**

Received for publication, May 9, 2005, and in revised form, August 5, 2005 Published, JBC Papers in Press, August 8, 2005, DOI 10.1074/jbc.M505077200

Teldja N. Azzouz[‡], Ramesh S. Pillai[‡1], Christoph Däpp[‡2], Ashwin Chari[§], Gunter Meister[§3], Christian Kambach[¶], Utz Fischer[§], and Daniel Schümperli[‡4]

From the [‡]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland, [§]Institute of Biochemistry, Biocenter of the University of Würzburg, Am Hubland, 97074 Würzburg, Germany, and [¶]Life Sciences, The Paul Scherrer Institute, 5232 Villigen PSI, Switzerland

The survival of motor neurons (SMN) complex mediates the assembly of small nuclear ribonucleoproteins (snRNPs) involved in splicing and histone RNA processing. A crucial step in this process is the binding of Sm proteins onto the SMN protein. For Sm B/B′, D1, and D3, efficient binding to SMN depends on symmetrical dimethyl arginine (sDMA) modifications of their RG-rich tails. This methylation is achieved by another entity, the PRMT5 complex. Its pICln subunit binds Sm proteins whereas the PRMT5 subunit catalyzes the methylation reaction. Here, we provide evidence that Lsm10 and Lsm11, which replace the Sm proteins D1 and D2 in the histone RNA processing U7 snRNPs, associate with pICln *in vitro* and *in vivo* without receiving sDMA modifications. This implies that the PRMT5 complex is involved in an early stage of U7 snRNP assembly and hence may have a second snRNP assembly function unrelated to sDMA modification. We also show that the binding of Lsm10 and Lsm11 to SMN is independent of any methylation activity. Furthermore, we present evidence for two separate binding sites in SMN for Sm/Lsm proteins. One recognizes Sm domains and the second one, the sDMA-modified RG-tails, which are present only in a subset of these proteins.

The U7 small nuclear ribonucleoprotein (snRNP)[5] is an essential factor mediating the endonucleolytic 3′-end processing of the replication-dependent, non-polyadenylated animal histone mRNAs (reviewed in Ref. 1). Although this cleavage reaction is biochemically distinct from the transesterifications involved in pre-mRNA splicing (reviewed in Ref. 2), the U7 snRNP particle resembles spliceosomal U snRNPs in various aspects of its structure and biogenesis.

The spliceosomal small nuclear RNAs (snRNAs) contain a conserved single-stranded sequence element, the Sm binding site, which interacts with the seven Sm proteins, B/B′, D1, D2, D3, E, F, and G, to form an ring-shaped heteroheptamer, the so-called Sm core (3–5). In contrast, U7 snRNA has a somewhat degenerate Sm binding site. Earlier studies indicated that, when this U7-specific Sm binding site (AAUUUGUCUAG; U7 Sm WT) was changed to resemble the spliceosomal Sm binding sequence (AAUUUUUGGAG, Sm OPT), the resulting snRNPs were non-functional in histone RNA processing (6, 7). Later, when U7 snRNPs were purified to homogeneity, they were found to contain two U7-specific Sm-like proteins, Lsm10 and Lsm11, which replace Sm D1 and D2 in a U7-specific Sm core (8, 9). The non-functionality of U7 Sm OPT snRNPs could be explained by the findings that U7 Sm OPT RNA forms a standard Sm core containing Sm D1 and D2 and that Lsm11 plays an essential functional role in histone RNA 3′-processing (8–10).

The assembly of Sm core structures occurs in the cytoplasm and is mediated by the multisubunit SMN complex (9, 11, 12). This complex consists of intrinsic components, often referred to as Gemins, and substrate proteins, *i.e.* the Sm/Lsm proteins, which are transferred onto the U snRNA during assembly. We have obtained evidence that the spliceosomal and U7-specific kinds of Sm cores are assembled by separate SMN complexes that contain either Sm D1/D2 or Lsm10/11 along with the five common Sm proteins (9). The presence of these separate and specialized SMN complexes raised the important question of how D1/D2 and Lsm10/11 gain access to the SMN complex and whether they occupy corresponding, mutually exclusive positions within this complex.

Another complex involved in the assembly of at least a subset of Sm proteins into Sm core structures is the PRMT5 complex (also termed methylosome). One of its subunits, the protein methyl transferase PRMT5, catalyzes symmetrical dimethyl arginine (sDMA) modifications within RG repeats found in the Sm proteins B/B′, D1, and D3. It has also been shown that sDMA modifications are essential for binding of these proteins to the SMN complex (13, 14). *In vivo*, the PRMT5 complex may temporarily or permanently associate with the SMN complex to form a larger assembly engine (15).

Here, we have addressed the question of whether the PRMT5 complex is involved in U7 snRNP assembly and how Lsm10/11 are incorporated into the SMN complex prior to U7 snRNP assembly. We find that both Lsm10 and Lsm11 associate with the pICln subunit of the PRMT5 complex *in vitro* and *in vivo*. The basis for this association appears to be a direct binding of both proteins to pICln through their respective Sm domains. However, neither Lsm10 nor Lsm11 appear to be substrates for sDMA modification, and their binding to SMN is independent of any methylation activity. Furthermore, we present evidence that the SMN protein has two binding modes for Sm/Lsm proteins. It can interact with the Sm domains of unmethylated Sm/Lsm proteins such that different members of this protein family can compete with each other for binding. Moreover, the methylated Sm/Lsm proteins exhibit another, strong binding mode, presumably through their sDMA-modified RG repeats, which cannot be competed by Sm domains.

* This work was supported by the State of Bern, by Swiss National Science Foundation Grants 31-65225.01 and 3100A0-105547 (to D. S.), and by Grant SFB581 of the German Research Foundation (to U. F.). 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.

[1] Present address: Friedrich Miescher Institute, Maulbeerstrasse 66, 4058, Basel, Switzerland.
[2] Present address: Institute of Anatomy, University of Bern, Baltzerstrasse 2, 3012 Bern, Switzerland.
[3] Present address: Max-Planck-Institut für Biochemie, Am Klopferspitz 18, 82152 Martinsried, Germany.
[4] To whom correspondence should be addressed. Tel.: 41-31-6314675; Fax: 41-31-6314616; E-mail: daniel.schuemperli@izb.unibe.ch.

[5] The abbreviations used are: snRNP, small nuclear ribonucleoprotein; SMN, survival of motoneurons; sDMA, symmetrical dimethyl arginine; PRMT5, protein methyl transferase 5; Lsm, Sm-like protein; snRNA, small nuclear ribonucleic acid; GST, glutathione *S*-transferase; HA, hemagglutinin antigen tag; NP40, Nonidet P40; ECL, enhanced chemiluminescence; SAH, *S*-adenosyl homocysteine; WT, wild type.

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

EXPERIMENTAL PROCEDURES

Plasmids—Plasmids derived from pcDNA3-HA (8) used for the expression of HA-tagged human Sm B, D1, D2, human Lsm10, or murine Lsm11 and of various truncations of the latter have been described elsewhere (8–10). For the in vitro methylation reactions shown in Fig. 5, a previously described SmB clone (3) and Lsm10 and Lsm11 cDNAs cloned in pET28a were used. For bacterial expression of fusion protein with an N-terminal GST tag, the coding region for the SMN N terminus and tudor domain (amino acids 1–160) was cloned into the pGex4T3 vector (Amersham Biosciences) to yield a plasmid termed GST-SMN-(1–160). The full-length open reading frame of pICln was similarly cloned into pcDNA3-HA and pGEX6P-1 (Amersham Biosciences). All plasmid constructs were verified by DNA sequencing. Details of the constructs are available on request.

GST Pull-down and Competition Assays—To study protein-protein interactions, recombinant proteins were isolated from Escherichia coli BL21 Gold transfected with pGex4T3 (negative control) or with pGex-derived plasmids containing SMN-(1–160) or pICln (see above). Unless indicated otherwise, 1–3 nmol of purified GST or GST fusion proteins were coupled to glutathione-Sepharose 4B beads and incubated with [^{35}S]methionine-labeled proteins obtained by coupled *in vitro* transcription/translation (see below). For competition experiments, 3.7 nmol of bacterially produced heterodimers of SmD1/D2 (16) were used additionally. The beads and proteins were incubated in phosphate-buffered saline supplemented with 0.1% Nonidet P-40 at 4 °C for 2 h while slowly turning on a rotating wheel, washed with phosphate-buffered saline/Nonidet P-40, and the bound and input materials were analyzed by 12% high TEMED SDS-PAGE (17) and detected on a Storm 820 PhosphorImager (Amersham Biosciences).

Coupled in Vitro Transcription/Translation—Protein substrates used for binding were produced by coupled in vitro transcription/translation of pcDNA3-HA-derived plasmids for Lsm10, Lsm11, or Sm B, D1, or D2 in rabbit reticulocyte lysate (TNT kit, Promega) and labeled with [^{35}S]methionine (Hartmann Analytic). To produce unmethylated Sm D1 or Sm B proteins, the reactions contained ~100 nM of the methylation inhibitor S-adenosyl-homocysteine (SAH). Coupled *in vitro* transcription/translation/methylation reactions were performed in rabbit reticulocyte lysate in the presence of 1 μM L-methionine and 2 μCi of *S*-[^{3}H]adenosyl methionine. Reactions were analyzed by 15% high TEMED SDS-PAGE, and dried gels were subjected to fluorography using a Biomax transcreen LE (Kodak).

Interaction Studies in Cell Extracts—Human 293-T cells were cultured as described (8). For transfection, they were grown in 10-cm dishes to 50–60% confluency and transfected with 10 μg of the pcDNA3-HA-derived plasmids complexed with Lipofectamine (Invitrogen, Life Technologies). Cells were harvested 48-h post-transfection. The preparation of small scale whole cell extracts, nuclear or cytoplasmic extracts, and precipitations with biotinylated oligonucleotides complementary to the 5′-ends of U7 or U1 snRNA were performed as described (8, 9). Proteins were analyzed by 12% high-TEMED SDS-PAGE (17), blotted, and probed with appropriate antibodies and developed by the ECL method (Amersham Biosciences).

Antibodies and Immunoprecipitation Experiments—The following antibodies were used: Monoclonal antibodies 7B10 for SMN (18) or Y12 for Sm proteins B/B′, D1, and D3 (19); rabbit antisera or affinity-purified antibodies specific for Lsm11 (9), pICln (14), bacterially produced, non-methylated Sm D1/D2 heterodimer (9), PRMT5 (14), or for WDR77 (alias WD45 or MEP50). For indirect immunodetection in Western blots anti-rabbit or anti-mouse antibodies coupled to horseradish peroxidase (Promega) were used. The HA epitope was detected with anti-HA antibody (Roche Applied Science) directly coupled to horseradish peroxidase.

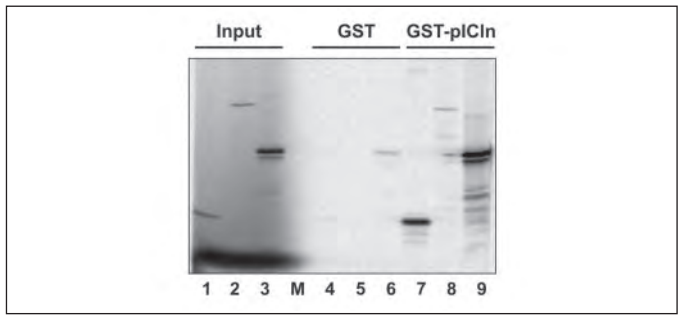


FIGURE 1. **Interactions of Lsm10 and Lsm11 with pICln *in vitro*.** *In vitro* translated [^{35}S]methionine-labeled Lsm10 (*lanes 1*, *4*, *7*), Lsm11 (*lanes 2*, *5*, *8*), and Sm B (*lanes 3*, *6*, *9*) were incubated with GST-pICln (*lanes 7–9*) or GST alone (*lanes 4–6*), immobilized on glutathione-Sepharose beads. *Input* (*lanes 1–3*) show 10% of the amount of *in vitro* translation reactions used in the binding experiments. Proteins binding to the immobilized recombinant proteins were resolved by SDS-PAGE and visualized by a phosphorimaging device. (Molecular Dynamics).

RESULTS

Interaction of Lsm10 and Lsm11 with pICln through Their Respective Sm Domains—In a yeast two-hybrid screen of 174,000 primary transformants using human Lsm10 as the bait, we isolated two full-length cDNA clones encoding the pICln subunit of the PRMT5 complex. These clones produced very strong β-galactosidase signals and did not transactivate the reporter genes in combination with four unrelated proteins used as baits or with the DNA binding domain alone (data not shown). To confirm this interaction and to analyze whether the other U7-specific Lsm protein, Lsm11, might also bind to pICln, we performed in vitro binding experiments. Lsm10, Lsm11, and Sm B were translated in the presence of [^{35}S]methionine in rabbit reticulocyte lysate and then incubated either with GST or with a GST-pICln fusion protein attached to glutathione-Sepharose beads. The beads were precipitated, washed, and their protein content was analyzed by SDS-PAGE and autoradiography. All three proteins clearly bound to GST-pICln (Fig. 1, *lanes 7–9*). For Lsm10 and Lsm11 this binding was caused by the pICln moiety, as no protein bound to GST alone (*lanes 4* and *5*). A trace of Sm B was precipitated by GST (*lane 6*), but the amount precipitated by GST-pICln was much higher, indicating that this previously described interaction (14) also preferentially occurred through the pICln moiety.

To study the relevance of these interactions, we analyzed whether they can also be detected in cell extracts. To this end, extracts from human 293-T cells expressing HA-tagged Lsm10, Lsm11 or various deletion mutants of the latter were subjected to immunoprecipitation by anti-pICln antibodies, and the precipitated material was analyzed by SDS-PAGE and Western blotting with anti-HA antibody. Both Lsm10 (Fig. 2*A*) and full-length Lsm11 (Fig. 2*B*, *top*) were specifically precipitated by the anti-pICln antibody but not by the beads alone. In the large Lsm11 protein, the first Sm motif is preceded by an N-terminal extension of 170 amino acids, and the Sm motifs 1 and 2 that form the Sm domain are separated by 138 amino acids (9). Two different deletions of the N terminus (ΔN104, ΔN140) or a deletion of 77 amino acids from the spacer separating the two Sm motifs (Δsp77) still allowed the interaction with pICln *in vivo* to occur (Fig. 2*B*). Note that all of these deleted proteins are also incorporated into U7 snRNPs in these cells (9). In contrast, an HA-tagged fragment of amino acids 1–136, which does not contain any of the Sm motifs did not associate with pICln *in vivo* (Fig. 2*B*, *N136*). Taken together, these results indicate that Lsm10 and Lsm11

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

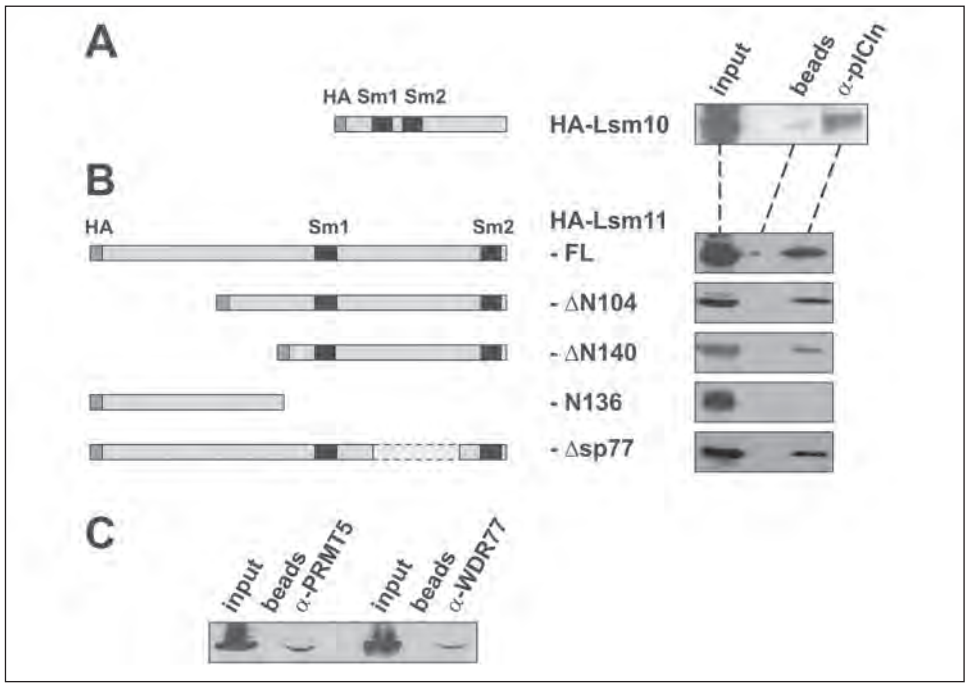


FIGURE 2. **Association of Lsm10 and Lsm11 with pICln, PRMT5, and WDR77 in cell extracts.** Expression plasmids encoding HA-tagged versions of Lsm10 (*A*) or of various parts of Lsm11 (*B*) were transfected into 293-T cells. Whole cell extracts were prepared 48-h post-transfection and subjected to immunoprecipitation with polyclonal anti-pICln antibodies coupled to protein G-Sepharose beads. *C*, similar immunoprecipitation of extract from cells expressing full-length HA-Lsm11 with antibodies against two other components of the PRMT5 complex, PRMT5, and WDR77. The bound proteins were resolved by SDS-PAGE and analyzed by Western blotting with anti-HA antibodies. Each panel shows samples of original extracts from the transfected cells (*input*), control precipitations with protein G-Sepharose beads lacking antibody (*beads*), and precipitations with the specific antibody indicated. An *empty lane* in the Lsm10 panel was loaded with a protein size marker.

interact with pICln both *in vitro* and *in vivo* and that, for Lsm11, this interaction occurs primarily if not exclusively through the Sm domain.

Interaction of Lsm11 with Other Members of the PRMT5 Complex—To test whether the U7-specific proteins interact with the complete PRMT5 complex or only with pICln, we performed a similar immunoprecipitation experiment with extracts from 293-T cells expressing full-length HA-Lsm11 and with antibodies specific for PRMT5 and WDR77. As shown in Fig. 2C, Lsm11 interacts not only with pICln but also with these other subunits of the PRMT5 complex.

Binding of pICln to Free Sm/Lsm Proteins Prior to snRNP Assembly—Because pICln binds Sm and Lsm proteins through their Sm motifs, we wanted to know whether it interacts with these proteins prior to or after their assembly into snRNPs. To address this question, HA-tagged pICln was transiently transfected into human 293-T cells. Nuclear and cytoplasmic extracts from these cells were then precipitated with magnetic streptavidin beads after the addition of biotinylated oligonucleotides complementary to either U7 or U1 snRNAs or without oligonucleotide. The presence or absence of HA-tagged pICln in the precipitates was analyzed by SDS-PAGE and Western blotting with anti-HA antibody. No precipitation of HA-pICln was observed with either the U7- or U1-specific oligonucleotide (Fig. 3A, *top panels*), although the method readily precipitated the Sm B/B′ that is associated with both kinds of snRNPs (*lower panels*). Note that the U7 snRNP is less abundant and that, therefore, less B/B′ was precipitated.

In control experiments, the expressed HA-pICln was readily precipitated from the cytoplasmic extract with antibodies against Sm proteins (Y12) or against Lsm11 (Fig. 3*B*, *right panel*), indicating that it shows the same interactions as observed for endogenous pICln (see Fig. 2 above). As the Y12 antibody is known to react primarily with the sDMA-modified repeats of Sm B, D1 and D3 (20), we also tested an antibody raised against a bacterially produced, unmethylated heterodimer of Sm D1 and D2. This antibody also precipitated HA-pICln (data not shown).

We note that, although HA-pICln is mostly cytoplasmic, a small fraction of it is also present in the nucleus. However, only traces of the nuclear HA-pICln (which may reflect a low level of cytoplasmic contamination) are associated with Sm proteins or with Lsm11. Taken

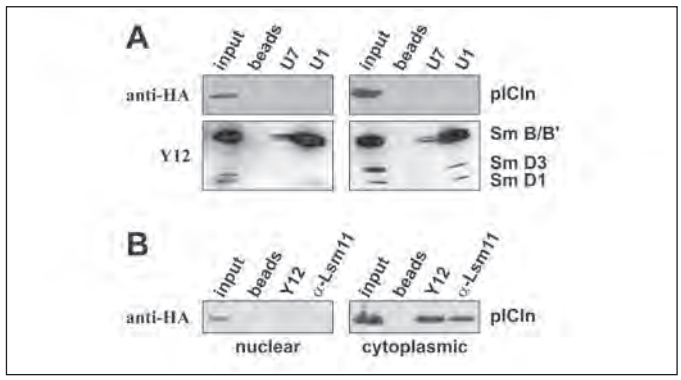


FIGURE 3. **pICln is not associated with mature snRNPs.** HA-tagged pICln was expressed by transient transfection in 293-T cells. Nuclear (*left panels*) and cytoplasmic (*right*) extracts were prepared 48-h post-transfection. *A*, precipitation of the extracts with magnetic streptavidin beads containing biotinylated oligonucleotides complementary to the 5′-ends of either U7 or U1 snRNA. The bound proteins were resolved by SDS-PAGE and analyzed by Western blotting with anti-HA antibodies (*top*) or Y12 anti-Sm antibodies (*bottom*) to reveal the proteins indicated on the *right*. Each panel shows control samples of original extracts from the transfected cells (*input*), as well as control precipitations with streptavidin beads lacking oligonucleotide (*beads*). Note that because of the lower abundance of U7 snRNPs much less Sm B/B′ is precipitated, and Sm D3 is undetectable. *B*, precipitation of the extracts with protein G-Sepharose beads containing monoclonal Y12 anti-Sm antibody or polyclonal anti-Lsm11 serum.

together with the previous results, these findings indicate that pICln (or the PRMT5 complex) does not bind mature assembled snRNPs but only free Sm or Lsm proteins before their assembly into mature snRNPs.

Binding of Lsm10 and Lsm11 to SMN Is Independent of Symmetrical Arginine Dimethylation—Having found that Lsm10 and Lsm11 interact with pICln, which is part of the PRMT5 complex that mediates sDMA modification of RG repeats in certain Sm proteins (13, 14), we asked whether the two U7-specific proteins are substrates for this kind of modification. Human/mouse Lsm11 contain 10/6 RGs and 9/6 GR dipeptides, respectively (9). For human/mouse Lsm10, there are 0/1 RGs and 4/3 GRs (8). However there is only one cluster of RG or GR dipeptides in the N terminus of Lsm11, which is removed by both of the N-terminal deletions described in Fig. 2. It is possible to determine

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

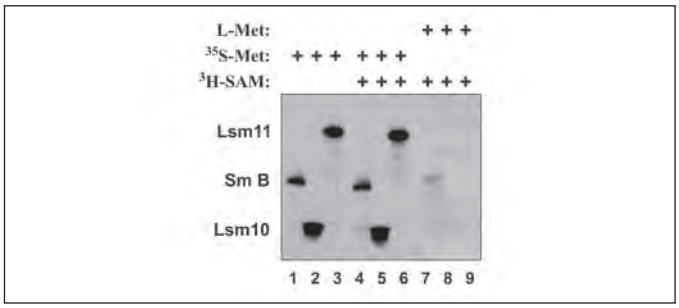


FIGURE 4. **Lack of evidence for symmetrical dimethylarginine modification of Lsm10 and Lsm11.** Sm B (*lanes 1, 4, 7*), Lsm11 (*lanes 2, 5, 8*), and Lsm10 (*lanes 3, 6, 9*) were synthesized by coupled *in vitro* transcription/translation in rabbit reticulocyte lysate. The translation products were either labeled with [^{35}S]methionine (*lanes 1–6*) or synthesized as unlabeled proteins (*lanes 7–9*). Methylation was traced with ^{3}H-labeled *S*-adenosine methionine (*^{3}H-SAM*; *lanes 4–9*) and revealed by fluorography. Note the lack of methylation of the two U7-specific Lsm proteins (*lanes 8–9*), whereas Sm B gets methylated (*lane 7*).

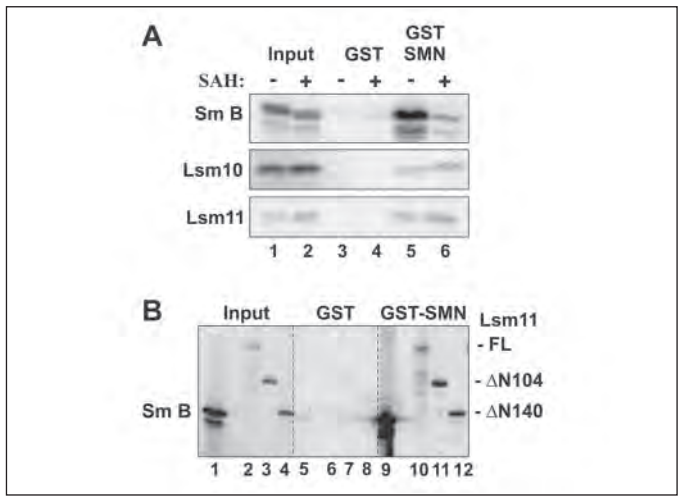


FIGURE 5. **Interaction of Lsm10 and Lsm11 with SMN *in vitro* is independent of sDMA modification.** *A*, interactions of Sm B, Lsm10, and Lsm11 *in vitro* translated in the absence or presence of the methylation inhibitor SAH with GST-SMN-(1–160) (*lanes 5* and *6*) or GST alone (*lanes 3* and *4*). Proteins binding to the immobilized recombinant proteins were analyzed as in Fig. 1. *Input* (*lanes 1* and *2*), show 10% of the amount of *in vitro* translation reactions used in the binding experiments. *B*, comparative binding to GST-SMN-(1–160) of Sm B and various fragments of Lsm11 translated in the absence of SAH.

whether proteins are substrates for methylation by coupled transcription/translation in the presence of tritiated S-adenosyl methionine in rabbit reticulocyte lysate which contains the PRMT5 methylation complex (and possibly PRMT7, which also catalyzes sDMA modifications (21, 22). Such assays were carried out with plasmids encoding Sm B, Lsm10, and Lsm11. Control reactions carried out in the presence of ^{35}S-labeled methionine either without (Fig. 4, *lanes 1–3*) or with *S*-[^{3}H]adenosyl methionine (*lanes 4–6*) proved that all three proteins were efficiently synthesized in the rabbit reticulocyte system. However, only Sm B (*lane 7*), but not Lsm10 (*lane 8*) or Lsm11 (*lane 9*), were detectable by fluorography after translation in the presence of *S*-[^{3}H]adenosyl methionine and unlabeled methionine. These results confirm that Sm B is a methylation substrate. Most importantly, however, they strongly suggest that Lsm10 and Lsm11 are not substrates for sDMA modification, although a weak signal from only one or a few methylated arginines might have escaped detection.

In the case of the Sm/Lsm proteins that are substrates for sDMA modification by the PRMT5 complex (*i.e.* Sm B/B′, D1 and D3, as well as Lsm4), the methylations are important for efficient binding to SMN and the SMN complex (13, 14). This conclusion was reached, among others, by experiments in which the proteins were translated either in the absence or presence of the methylation inhibitor *S*-adenosyl homocysteine (SAH). We therefore analyzed whether translation of Lsm10 and Lsm11 in the presence of SAH had similar effects on the subsequent binding to SMN as was the case for Sm B. As a test for binding, precipitation experiments were performed with recombinant proteins GST-SMN (1–160) or GST (as negative control) that had been immobilized on glutathione beads. Note that the SMN fragment used consists of amino acids 1–160 and contains the binding site for Sm proteins (23).

In such binding assays, all three Sm/Lsm proteins interacted with GST-SMN-(1–160) (Fig. 5*A*, *lane 5*) but not with GST alone (*lane 3*). As expected, the binding of Sm B to SMN was strongly reduced when the *in vitro* translation had been carried out in the presence of SAH (*lane 6*). We also observed a shift in electrophoretic mobility of Sm B that presumably reflected the difference in sDMA modification (*compare lanes 1* and *2 or 5* and *6*). In contrast, the binding of Lsm10 and Lsm11 to GST-SMN-(1–160) was not affected by the inhibition of methylation (*lane 6*), and there also was no detectable mobility shift. Thus methylation of Lsm10 and Lsm11, if it occurs at all, is not important for the interaction of these proteins with SMN.

Next we analyzed whether the two N-terminal truncations described in Fig. 2 could still interact with GST-SMN-(1–160). In these truncation mutants, all RG/GR dipeptide repeats have been deleted. Both Lsm11Δ^{N104} and Lsm11Δ^{N140} specifically bound to GST-SMN-(1–160) (Fig. 5*B*, *lanes 10–12*), but not to the GST control (*lanes 6–8*). Moreover, in immunoprecipitation experiments with extracts from transfected cells similar to those shown in Fig. 2, we found that the same HA-tagged Lsm11 deletions could be precipitated by anti-SMN antibody upon transient expression in 293-T cells (data not shown). Furthermore, we have previously shown that the N-terminally deleted Lsm11 proteins still assemble into U7 snRNPs *in vivo* (9). Therefore these experiments indicate that the N terminus of Lsm11 is not required for interactions with SMN protein or the SMN complex, be it *in vitro* or *in vivo*. Considering all the above results, binding of Lsm10 or Lsm11 to SMN occurs independently of arginine methylation.

Evidence for Separate Binding Sites on SMN for Sm Domains and Methylated RG-rich Tails of Sm/Lsm Proteins—Both the spliceosomal and U7-specific Sm core structures are assembled in the cytoplasm with the help of the SMN complex (9, 11, 12). We have previously obtained evidence for two separate kinds of SMN complexes that contain either Sm D1/D2 or Lsm10/11 and which are devoted to spliceosomal or U7 snRNP assembly, respectively (9). To begin to address the question how these different types of SMN complexes are formed, we analyzed whether the relevant proteins, Sm D1, D2, and Lsm10, Lsm11, bind to SMN in the same way. In particular, we wanted to know whether Sm D1, D2 can compete with Lsm10, Lsm11 for binding to SMN or vice versa.

For this purpose, the binding of these proteins to GST-SMN-(1–160) was analyzed in the absence or presence of an excess of purified recombinant Sm D1/D2 heterodimer used as competitor. As expected, in the absence of competitor all four proteins bound to the immobilized GST-SMN-(1–160) (Fig. 6*A*, *lane 3*) but not to the GST negative control (*lane 2*). Interestingly, however, the recombinant Sm D1/D2 heterodimer competed for the binding to GST-SMN-(1–160) with the *in vitro* translated Lsm11, Lsm10, and Sm D2 proteins, but not with Sm D1 (compare *lanes 4 to 3*).

An important difference between the *in vitro* translated Sm D1 used as binding substrate and the recombinant D1 present in the heterodimer used as competitor resides in the sDMA modification of the C-terminal RG tail. It is possible that SMN may have two binding sites for Sm/Lsm proteins, one interacting with the Sm domain and the other

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

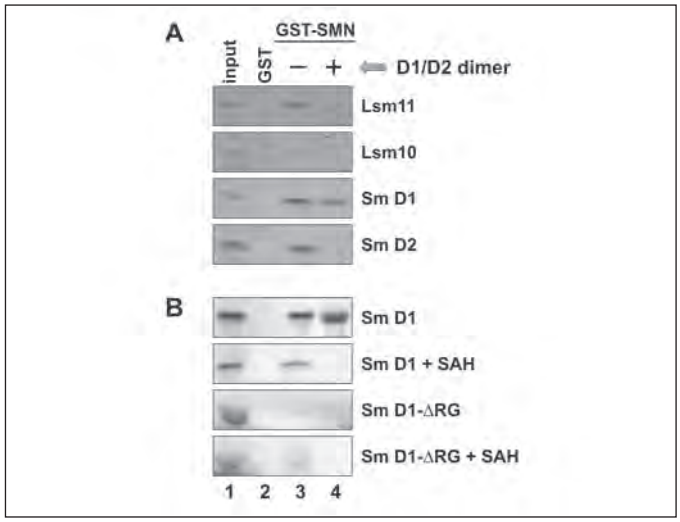


FIGURE 6. **Different binding modes of unmethylated and methylated Sm/Lsm proteins to SMN.** *A*, interactions of *in vitro* translated ^{35}S-labeled Lsm11, Lsm10, SmD1, and Sm D2 with GST (*lane 3;* negative control) or GST-SMN-(1–160) and competition experiments in the absence or the presence of unmethylated Sm D1/D2 heterodimer (*lanes 3* and *4, respectively*). *B*, binding of Sm D1 and Sm D1 deleted of its C-terminal RG-rich tail (*Sm D1-ΔRG*) translated in the absence or presence of the methylation inhibitor SAH to GST or GST-SMN-(1–160) and binding competition using the unmethylated Sm D1/D2 dimer. Proteins binding to the immobilized recombinant proteins were analyzed as in Fig. 1. *Input* (*lane 1*), show 10% of the amount of *in vitro* translation reactions used in the binding experiments.

with the methylated RG-tail (23). Based on this hypothesis, the unmethylated Sm D1/D2 dimer should compete with unmethylated Sm/Lsm proteins, which interact with SMN-(1–160) solely through the Sm domain mode. In contrast, the methylated *in vitro* translated Sm D1 should bind through both the Sm domain and RG-tail modes and therefore be resistant to competition by the D1/D2 dimer.

If this interpretation were correct, the residual binding to SMN-(1–160) of Sm D1 translated in the presence of the methylation inhibitor SAH should be competed by the Sm D1/D2 dimer. The same should be true for the binding to SMN-(1–160) of Sm D1 deleted of its RG-rich C-terminal tail. The experimental verification of these predictions is shown in Fig. 6*B*. The binding of Sm D1 *in vitro* translated in the presence of SAH to GST-SMN-(1–160) is indeed competed by an excess of the unmethylated SmD1/D2 heterodimer (Fig. 6*B*, *second panel*). Moreover, the very weak binding of Sm D1-ΔRG to GST-SMN-(1–160) is efficiently competed by the unmethylated SmD1/D2 heterodimer, irrespective of whether Sm D1-ΔRG has been translated in the presence or absence of SAH (Fig. 6*B*, *third* and *fourth panels*).

Taken together, these data provide strong support for the notion that SMN has two interaction sites for Sm/Lsm proteins. The first one involves the Sm domains and the second involves the methylated RG repeat tails contained in some of these proteins. However, the second interaction appears to be stronger and more important for binding to SMN of those proteins that contain sDMA-modified tails.

DISCUSSION

We have previously shown that the spliceosomal and U7 snRNA-specific Sm core structures are assembled by separate SMN complexes that contain either Sm D1/D2 or Lsm10/11 along with the five common Sm proteins (9). These findings raised the question how D1/D2 and Lsm10/11 gain access to the SMN complex and if their binding to the SMN protein could provide an explanation for these two mutually exclusive SMN complexes. This has been addressed here by analyzing the interactions of Lsm10, Lsm11, and various Sm proteins with pICln and SMN.

Evidence for a Methylation-independent Function of the PRMT5 Complex in snRNP Assembly—Several Sm/Lsm proteins, as well as other substrates acquire sDMA modifications of RG-rich repeats through the action of the PRMT5 complex (13, 14, 20, 24–32). For the Sm proteins B/B′, D1 and D3, these modifications are important for their efficient interaction with SMN, the lead component of the SMN complex involved in snRNP assembly (13, 14). Here we have analyzed, whether Lsm10 and Lsm11, the two proteins that substitute Sm D1 and D2 in the U7-specific core, also interact with the PRMT5 complex and might be substrates for sDMA modification.

Our results clearly demonstrate that both Lsm10 and Lsm11 interact with the pICln subunit of the PRMT5 complex in vitro (Fig. 1) and in cell extracts (Fig. 2). For the large Lsm11 protein, this interaction requires neither the N terminus nor the spacer between the two Sm motifs, but rather seems to depend solely on the Sm domain (Fig. 2*B*). At least Lsm11 also interacts with PRMT5 and WDR77 (Fig. 2*C*) indicating that the interaction is with the entire PRMT5 complex rather than with pICln alone. Moreover, this interaction occurs prior to snRNP assembly (Fig. 3). However, we were unable to detect any sDMA modification of Lsm10 or Lsm11 (Fig. 4), and the binding of both proteins to SMN was not affected by the methylation inhibitor SAH (Fig. 5*A*). Furthermore, the N-terminal part of Lsm11 that contains several RG dipeptides does not contribute significantly to the interaction with SMN *in vitro* (Fig. 5*B*) or in cell extracts (data not shown). We have previously shown that this interaction is not a dead-end product, since these N-terminally truncated proteins get assembled onto U7 snRNA to form snRNPs (9). Taken together, these findings imply that the PRMT5 complex plays a previously unsuspected role in the assembly of Lsm10 and Lsm11 into U7 snRNPs that is independent of its methylation activity.

It is not known how the Sm proteins that are not subject to sDMA modification are recruited to the SMN complex. However our findings may be related to recent evidence demonstrating that the PRMT5 and SMN complexes form a larger structure, which more efficiently assembles U snRNPs than the SMN complex alone (15). Perhaps all Sm/Lsm proteins must bind sequentially to pICln and SMN in this superstructure during the assembly reaction. It is also conceivable that pICln plays an obligatory role in preventing premature assembly of Sm/Lsm oligomers. Recently discovered phosphorylations of pICln and SMN may be involved in this presumptive quality control function (33).

SMN Has Separate Binding Sites for Sm Domains and Methylated RG-rich Repeats—In our studies of the binding of SMN to the Sm D1, D2 and Lsm10, Lsm11 proteins, we have obtained evidence for two separate kinds of interaction. On the one hand, the Sm/Lsm proteins that are not substrates for sDMA modification or whose methylation is prevented by SAH interact with SMN through their Sm domains. This notion was deduced from the binding of the N-terminally truncated Lsm11 proteins to SMN (Fig. 5B). Moreover, the binding of unmethylated Sm D1 was not affected by deletion of the RG-rich C terminus (Fig. 6*B*). In every case analyzed, this Sm-domain-dependent binding could be competed by unmethylated recombinant Sm D1/D2 dimer (Fig. 6, *A* and *B*).

On the other hand, the Sm proteins containing RG-rich tails mainly interact with SMN by virtue of these sDMA-modified structures. This was demonstrated previously by a strongly reduced binding when the tails were deleted or their methylation was prevented (13, 14). Some of our control experiments confirm this for Sm B (Fig. 5*A*) and for Sm D1(Fig. 6*B*). Moreover, methylated RG tail peptides, but not unmethylated ones, have previously been shown to be sufficient for binding to SMN (34, 35).

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

Relevant for this sDMA-dependent interaction, we have shown that the unmethylated D1/D2 dimer does not compete with methylated Sm D1 for binding to SMN, but competes with unmethylated D1 or D1 deleted of its RG-rich C terminus (Fig. 6*B*). However, because unmethylated D1 and D1-ΔRG still bind to SMN, albeit weakly, this does imply that SMN has two binding sites, one for the Sm domain and the other for methylated RG repeats.

The basis for the interaction of SMN with the methylated RG tails, a negatively charged surface of the protein, has previously been analyzed at the structural level by NMR spectroscopy and x-ray crystallography (23, 36). Interestingly, these studies also revealed a structural similarity between the SMN tudor domain and Sm domains (23). It was therefore speculated that the tudor domain may form a similar intermolecular β4–β5 interface with Sm proteins as is observed between neighboring Sm domains in oligomeric Sm/Lsm structures (5, 16, 37–39). However, this postulated interaction of SMN with Sm domains had not been studied experimentally. Only in one study it was reported that, even though the Sm-like proteins Lsm2, Lsm6, and Lsm7 lack an RG-rich tail, they are still able to bind SMN (34). As discussed above, we have now obtained experimental evidence for this interaction of SMN with Sm domains. Thus, the formation of a high affinity SMN-Sm complex could require multiple and cooperative interactions, involving Sm core and tail binding to the SMN tudor domain, and, possibly, interactions with additional regions of the SMN protein (40, 41) or with other members of the SMN complex.

Although our studies have not yet provided an explanation how the two mutually exclusive SMN complexes dedicated to the assemblies of the U7-specific and spliceosomal Sm core structures are formed, they provide important new insights into the interactions of Sm and Lsm proteins with the PRMT5- and SMN complexes. A full understanding of the specificity of U7 *versus* spliceosomal snRNP assembly will most likely require a deeper insight into the relative arrangements and stoichiometries of the Sm/Lsm proteins with respect to the other members of the SMN complex.

Acknowledgment—We thank K. Schranz for technical help.

REFERENCES

1. Müller, B., and Schümperli, D. (1997) *Semin. Cell Dev. Biol.* **8,** 567–576
2. Moore, M. J., Query, C. C., and Sharp, P. A. (1993) in *The RNA World* (Gesteland, R. F., and Atkins, J. F., eds) pp. 303–357, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
3. Raker, V. A., Plessel, G., and Lührmann, R. (1996) *EMBO J.* **15,** 2256–2269
4. Will, C. L., and Lührmann, R. (2001) *Curr. Opin. Cell Biol.* **13,** 290–301
5. Kambach, C., Walke, S., and Nagai, K. (1999) *Curr. Opin. Struct. Biol.* **9,** 222–230
6. Grimm, C., Stefanovic, B., and Schümperli, D. (1993) *EMBO J.* **12,** 1229–1238
7. Stefanovic, B., Hackl, W., Lührmann, R., and Schümperli, D. (1995) *Nucleic Acids Res.* **23,** 3141–3151
8. Pillai, R. S., Will, C. L., Lührmann, R., Schümperli, D., and Müller, B. (2001) *EMBO J.* **20,** 5470–5479
9. Pillai, R. S., Grimmler, M., Meister, G., Will, C. L., Lührmann, R., Fischer, U., and Schümperli, D. (2003) *Genes Dev.* **17,** 2321–2333
10. Azzouz, T. N., Gruber, A., and Schumperli, D. (2005) *Nucleic Acids Res.* **33,** 2106–2117
11. Meister, G., Eggert, C., and Fischer, U. (2002) *Trends Cell Biol.* **12,** 472–478
12. Paushkin, S., Gubitz, A. K., Massenet, S., and Dreyfuss, G. (2002) *Curr. Opin. Cell Biol.* **14,** 305–312
13. Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001) *Mol. Cell. Biol.* **21,** 8289–8300
14. Meister, G., Eggert, C., Bühler, D., Brahms, H., Kambach, C., and Fischer, U. (2001) *Curr. Biol.* **11,** 1990–1994
15. Meister, G., and Fischer, U. (2002) *EMBO J.* **21,** 5853–5863
16. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J., and Nagai, K. (1999) *Cell* **96,** 375–387
17. Will, C. L., Kastner, B., and Lührmann, R. (1994) in *RNA Processing*, Vol. 1 (Higgins, S. J., and Hames, B. D., eds) pp. 141–177, Oxford University Press, Oxford
18. Meister, G., Bühler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (2000) *Hum. Mol. Genet.* **9,** 1977–1986
19. Lerner, E. A., Lerner, M. R., Janeway, C. A., and Steitz, J. A. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2737–2741
20. Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Lührmann, R. (2000) *J. Biol. Chem.* **275,** 17122–17129
21. Miranda, T. B., Miranda, M., Frankel, A., and Clarke, S. (2004) *J. Biol. Chem.* **279,** 22902–22907
22. Lee, J. H., Cook, J. R., Yang, Z. H., Mirochnitchenko, O., Gunderson, S. I., Felix, A. M., Herth, N., Hoffmann, R., and Pestka, S. (2005) *J. Biol. Chem.* **280,** 3656–3664
23. Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U., and Sattler, M. (2001) *Nat. Struct. Biol.* **8,** 27–31
24. Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Lührmann, R. (2001) *RNA* **7,** 1531–1542
25. Baldwin, G. S., and Carnegie, P. R. (1991) *Biochem. J.* **123,** 69–74
26. Branscombe, T. L., Frankel, A., Lee, J. H., Cook, J. R., Yang, Z., Pestka, S., and Clarke, S. (2001) *J. Biol. Chem.* **276,** 32971–32976
27. Fabbrizio, E., El Messaoudi, S., Polanowska, J., Paul, C., Cook, J. R., Lee, J. H., Negre, V., Rousset, M., Pestka, S., Le Cam, A., and Sardet, C. (2002) *EMBO Rep.* **3,** 641–645
28. Kwak, Y. T., Guo, J., Prajapati, S., Park, K. J., Surabhi, R. M., Miller, B., Gehrig, P., and Gaynor, R. B. (2003) *Mol. Cell* **11,** 1055–1066
29. Pal, S., Yun, R., Datta, A., Lacomis, L., Erdjument-Bromage, H., Kumar, J., Tempst, P., and Sif, S. (2003) *Mol. Cell. Biol.* **23,** 7475–7487
30. Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P., and Sif, S. (2004) *Mol. Cell. Biol.* **24,** 9630–9645
31. Amente, S., Napolitano, G., Licciardo, P., Monti, M., Pucci, P., Lania, L., and Majello, B. (2005) *FEBS Lett.* **579,** 683–689
32. Richard, S., Morel, M., and Cleroux, P. (2005) *Biochem. J.* **388,** 379–386
33. Grimmler, M., Bauer, L., Nousiainen, M., Korner, R., Meister, G., and Fischer, U. (2005) *EMBO Rep.* **6,** 70–76
34. Friesen, W. J., and Dreyfuss, G. (2000) *J. Biol. Chem.* **275,** 26370–26375
35. Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001) *Mol. Cell* **7,** 1111–1117
36. Sprangers, R., Groves, M. R., Sinning, I., and Sattler, M. (2003) *J. Mol. Biol.* **327,** 507–520
37. Collins, B. M., Harrop, S. J., Kornfeld, G. D., Dawes, I. W., Curmi, P. M., and Mabbutt, B. C. (2001) *J. Mol. Biol.* **309,** 915–923
38. Mura, C., Cascio, D., Sawaya, M. R., and Eisenberg, D. S. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 5532–5537
39. Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. (2001) *EMBO J.* **20,** 2293–2303
40. Bühler, D., Raker, V., Lührmann, R., and Fischer, U. (1999) *Hum. Mol. Genet.* **8,** 2351–2357
41. Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11167–11172

Downloaded from www.jbc.org at Univ Bibliothek Wurzburg on July 28, 2009

Cell

An Assembly Chaperone Collaborates with the SMN Complex to Generate Spliceosomal SnRNPs

Ashwin Chari,[1] Monika M. Golas,[2] Michael Klingenhäger,[1] Nils Neuenkirchen,[1] Bjoern Sander,[2] Clemens Englbrecht,[1] Albert Sickmann,[3] Holger Stark,[2] and Utz Fischer[1,*]
[1]Department of Biochemistry, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany
[2]Research Group of 3D Electron Cryomicroscopy, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37070 Göttingen, Germany
[3]Rudolf Virchow Centre for Experimental Medicine, Versbacher Str. 9, D-97078 Würzburg, Germany
*Correspondence: utz.fischer@biozentrum.uni-wuerzburg.de
DOI 10.1016/j.cell.2008.09.020

SUMMARY

Spliceosomal small nuclear ribonucleoproteins (snRNPs) are essential components of the nuclear pre-mRNA processing machinery. A hallmark of these particles is a ring-shaped core domain generated by the binding of Sm proteins onto snRNA. PRMT5 and SMN complexes mediate the formation of the core domain in vivo. Here, we have elucidated the mechanism of this reaction by both biochemical and structural studies. We show that pICln, a component of the PRMT5 complex, induces the formation of an otherwise unstable higher-order Sm protein unit. In this state, the Sm proteins are kinetically trapped, preventing their association with snRNA. The SMN complex subsequently binds to these Sm protein units, dissociates pICln, and catalyzes ring closure on snRNA. Our data identify pICln as an assembly chaperone and the SMN complex as a catalyst of spliceosomal snRNP formation. The mode of action of this combined chaperone/catalyst system is reminiscent of the mechanism employed by DNA clamp loaders.

INTRODUCTION

Molecular chaperones interact transiently with polypeptide chains, prevent or reverse misfolding, and thereby promote the adoption of functional tertiary structures. The concept of chaperone action also applies to the assembly of multisubunit macromolecular complexes. In this case, the primary function is to orchestrate the joining of individual components into a higher-order complex. In addition, these chaperones prevent nonspecific or premature interactions. Examples for this type of chaperones (also termed assembly chaperones) include the biogenesis factors of nucleosomes, proteasomes, or Rubisco (Hirano et al., 2006; Laskey et al., 1978; Le Tallec et al., 2007; Saschenbrecker et al., 2007).

The enormous complexity of some ribonucleoprotein particles (RNPs) has implied that these macromolecular structures may, likewise, require assisting factors for their formation. In keeping with this notion, an active and factor-mediated process leads to the formation of the Sm class of spliceosomal small nuclear ribonucleoproteins (snRNPs) (Fischer et al., 1997; Meister et al., 2001a; Pellizzoni et al., 2002). This class, as an essential part of the spliceosome, catalyzes the removal of noncoding sequences from pre-mRNAs. SnRNPs are composed of a small eponymous snRNA, as well as common (Sm) and specific proteins. Prior to their involvement in the splicing cycle, snRNPs undergo a segmented biogenesis pathway, starting with the nuclear export of the snRNAs. Within the cytoplasm, the seven Sm proteins B-B′ (B′ is an alternatively spliced product of the Sm B gene), D1, D2, D3, E, F, and G bind to a single-stranded region of the RNA, termed Sm site. As a result, a toroidal Sm core domain is generated that interacts with the RNA in its central region (Kambach et al., 1999; Stark et al., 2001; Urlaub et al., 2001). After additional modification and processing steps, the RNP is targeted to its nuclear site of function (reviewed in Will and Lührmann, 2001).

The formation of the Sm core domain has been recapitulated in vitro with purified Sm proteins and snRNA (Raker et al., 1996). These studies have established a hierarchical maturation pathway in which Sm proteins initially form the three hetero-oligomers D3/B, D1/D2, and E/F/G independent of snRNA. Next, binding of D1/D2 and E/F/G onto the snRNA leads to the formation of an Sm subcore intermediate, which is finally converted into the mature Sm core upon the addition of the D3/B heterodimer. Even though self-recognition of RNA and protein counterparts is sufficient for Sm core assembly in vitro, two functional units, termed PRMT5 and SMN complexes (protein arginine methyltransferase 5 and survival motor neuron complexes), are required in vivo (Meister et al., 2002; Paushkin et al., 2002). The PRMT5 complex is composed of PRMT5, pICln, and WD45 (Mep50) and recruits Sm proteins via the pICln subunit (Friesen et al., 2001; Meister et al., 2001b). PRMT5 symmetrically dimethylates arginines (sDMA) within the C-terminal tail domains of Sm proteins B-B′, D1, and D3, which is believed to enhance their transfer onto the SMN complex (Brahms et al., 2001; Friesen and Dreyfuss, 2000). The SMN complex composed of

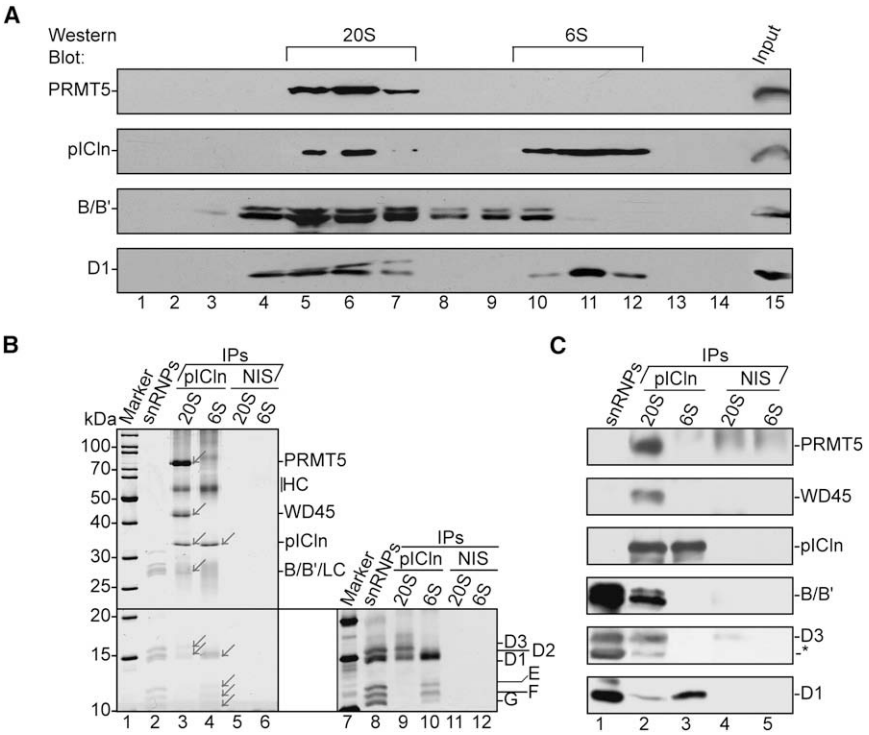


Figure 1. pICln Forms Two Complexes with Distinct Subsets of Sm Proteins

(A) Gel filtration fractions of HeLa cytosolic extract, active in snRNP assembly, were resolved by SDS-PAGE, and the indicated proteins were detected by western blotting. pICln was found in fractions corresponding to 20S (lanes 5–7) and 6S (lanes 10–12), respectively.

(B) Coomassie blue-stained gel of immunoaffinity-purified pICln complexes from the 20S (lane 3) and 6S (lane 4) fractions. Purifications with control antibodies are depicted in lanes 5 and 6, and a molecular size marker and affinity-purified snRNP cores are shown in lanes 1 and 2. The inset (lanes 7–12) shows a silver-stained portion of the same gel for a better visualization of low-molecular weight proteins.

(C) Western blot analysis of the immunoprecipitations (lanes 2–5) and affinity-purified snRNPs. The respective antigens were either detected with polyclonal antibodies (PRMT5, WD45, pICln) or the sDMA-specific monoclonal antibody Y12 (B/B′, D1, and D3). Asterisks mark nonspecific bands or degradation; HC and LC indicate heavy and light chains of the antibody used for affinity purification.

SMN and Gemins2–8 then facilitates the loading of Sm proteins onto snRNA, resulting in the formation of the Sm core (Gubitz et al., 2004; Meister et al., 2001a; Meister and Fischer, 2002; Neuenkirchen et al., 2008; Pellizzoni, 2007; Pellizzoni et al., 2002). Reduced expression of the SMN protein is the underlying cause for spinal muscular atrophy, providing an interesting link between the biogenesis of snRNPs and disease (Winkler et al., 2005; Zhang et al., 2008).

Although the factors that mediate Sm core formation have been identified, the mechanism of this reaction remains elusive. Here, we have dissected the assembly process by a combination of biochemical and structural studies. We identify a ring-shaped 6S particle composed of the Sm proteins D1/D2, E/F/G, and pICln, which does not allow RNA binding. SnRNP core formation depends on the activity of the SMN complex, which contacts the outer surface of this 6S intermediate and displaces pICln. As a consequence, the Sm proteins adopt an arrangement ready for snRNA binding and RNP formation. Our data identify pICln as an Sm-specific assembly chaperone and the SMN complex as a catalyst for the transfer of Sm proteins from the 6S complex onto snRNA.

RESULTS

pICln Is a Constituent of Two Complexes, Each Containing Distinct Subsets of Sm Proteins

Previous studies have demonstrated a direct interaction between pICln and Sm proteins that occurs in Sm core formation before the SMN complex is involved (Friesen et al., 2001; Meister and Fischer, 2002). To understand its role in the biogenesis pathway, we set out to identify complexes formed between pICln and Sm proteins in steady state. When cytosolic extract from HeLa cells was size fractionated by gel filtration, pICln eluted in two peaks that correspond to approximate sedimentation values of 20S and 6S, respectively (Figure 1A; see Supplemental Data available online for transformation of molecular weight into sedimentation values). The Sm protein D1 had an elution profile very similar to pICln, whereas PRMT5 and B-B′ were predominantly found in the 20S peak. Both peak fractions were individually subjected to immunoaffinity purification with anti-pICln antibodies and analyzed by SDS-PAGE, western blotting, and mass spectrometry. These approaches resolved the 20S peak as a complex formed by pICln, PRMT5, WD45, and all Sm proteins (Figures 1B, lanes 3 and 9, and 1C, lane 2). Of note, E/F/G in this peak was consistently underrepresented when compared to the other Sm proteins and, in many cases, could only be detected by mass spectrometric analyses (see also Figure S1). The 6S peak fraction, aside from pICln, contained the Sm proteins D1/D2 and E/F/G in apparently stoichiometric amounts while being devoid of D3/B (Figures 1B, lanes 4 and 10, and 1C, lane 3). Furthermore, immunoblots with the symmetrical dimethylarginine-specific Sm antibody Y12 revealed that not only the Sm proteins D1, D3, and B of the 20S complex were methylated, but also D1 in the 6S complex (Figures 1A, 1C, and data not shown). Thus, upon size fractionation, pICln resolves into two peaks, one containing only the Sm proteins of the snRNP subcore (6S) and one that is equivalent to the previously identified PRMT5 complex (20S).

An Ordered Sequence of Events Leads to the Formation of the 6S Complex

Isolated 6S complex contains equimolar quantities of D1, D2, E, F, G, and pICln (Figure 1B, lanes 4 and 10). However, the Sm

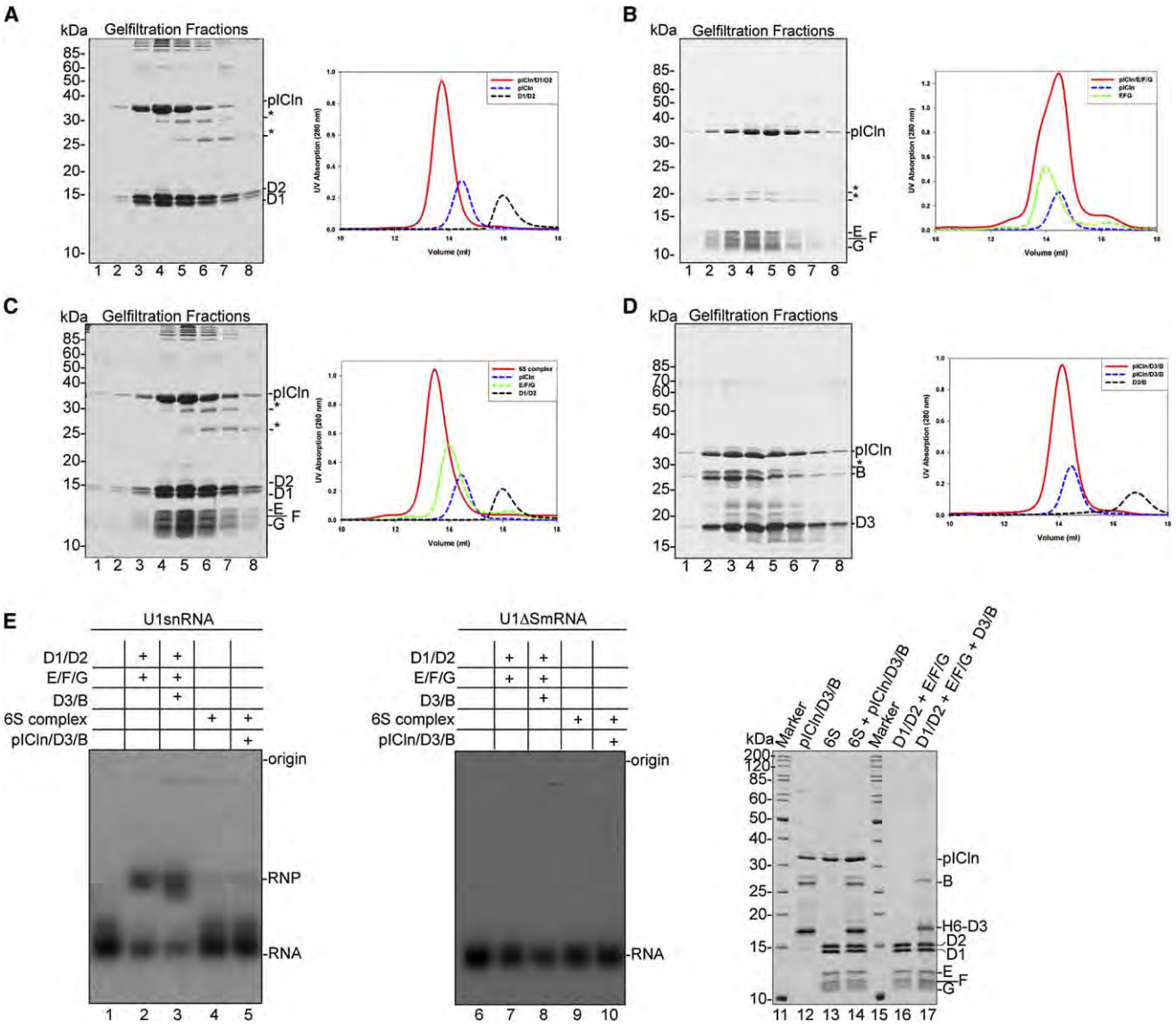


Figure 2. Cooperative Binding of Sm Proteins to pICln Leads to the Formation of an Assembly-Inactive 6S Complex

Recombinant pICln was incubated with equimolar amounts of D1/D2 (A), E/F/G (B), a combination of D1/D2 and E/F/G (C), or D3/B (D). The mixtures were separated by gel filtration, and individual fractions were analyzed by SDS-PAGE (left panel). Chromatograms of the mixtures and of individual building blocks, respectively, are shown (right panel). Asterisks denote impurities. (E) In vitro assembly of ^{32}P-labeled U1snRNA with either isolated Sm protein hetero-oligomers (lanes 2 and 3) or pICln-Sm complexes (lanes 4 and 5). RNP formation was analyzed by native gel electrophoresis. U1snRNA without proteins is shown in lane 1. In lanes 6–10, the same experiment was performed with U1ΔSm. Lanes 12–14, 16, and 17 depict the complexes used for these experiments.

protein units D1/D2 and E/F/G alone are incapable of forming a stable complex in the absence of snRNA (Raker et al., 1996). Therefore, we determined in detail how the five Sm proteins interact with pICln to form the 6S complex. For this, we incubated pICln with equimolar amounts of recombinant and unmethylated D1/D2, E/F/G, or a combination of both. Complex formation was monitored by gel filtration and subsequent SDS-PAGE. pICln formed a homogenous complex with D1/D2, as evident by the strict coelution of the three proteins in identical fractions from the column (Figure 2A). In contrast, in the presence of pICln, E/F/G eluted as a separate entity, indicating no or inefficient complex formation (Figure 2B; see also the chromatograms of individual components in the right panel). However, when pICln was incubated with a mixture of the Sm hetero-oligomers D1/D2 and E/F/G, a stable complex was formed (Figure 2C) with an identical elution profile to the 6S complex presented in Figure 1A. These findings are corroborated by immunoprecipitation of individual gel filtration fractions with anti-pICln antibodies (data not shown). Thus, the 6S complex forms in a stepwise manner, and the binding site for the E/F/G heterotrimer is generated by the preceding formation of a pICln/D1/D2 trimer. These data also suggest that methylation of Sm D1 appears to be dispensable for 6S formation.

pICln formed a second stable unit with the unmethylated D3/B hetero-oligomer, i.e., the Sm proteins missing in the 6S complex (Figure 2D). Interestingly, the pICln/D3/B complex failed to interact with any of the other two Sm hetero-oligomers and, hence, is a separate unit (data not shown). This pICln complex is not detected at steady state in size-fractionated cytosolic extracts (see Figure 1A and data not shown) and, hence, is likely to be either short lived or to exist only as part of the 20S PRMT5 complex.

Sm Proteins Do Not Associate with SnRNA in a pICln-Bound Form

The 6S complex (pICln/D1/D2/E/F/G) contains the five Sm proteins required for the assembly of the snRNP subcore, i.e., the intermediate formed in vitro on snRNA in the absence of the D3/B heterodimer (Raker et al., 1996). This raised the question of whether the Sm proteins of the 6S complex were able to associate with snRNA. By native gel electrophoresis, we demonstrate that neither preformed 6S complex alone nor in conjunction with the pICln/D3/B heterotrimer forms RNPs on ^{32}P-labeled U1snRNA (Figure 2E, lanes 4 and 5). In contrast and consistent with a previous report (Raker et al., 1996), spontaneous formation of the Sm subcore was observed upon incubation of D1/D2 and E/F/G with U1snRNA, which is matured into the snRNP core upon addition of D3/B (Figure 2E, lanes 2 and 3). No RNPs were formed on RNA lacking a functional Sm site either with pICln-bound or free Sm proteins (Figure 2E, lanes 6–10; see lanes 12–14, 16, and 17 for the protein complexes used). Therefore, Sm proteins appear to be incompetent to form snRNPs when they are associated with pICln. This could either mean that the 6S complex is a dead-end product in the assembly pathway or that it acts as a kinetic trap for the bound Sm proteins. If the latter scenario was correct, pICln would be a reversible inhibitor of snRNP core formation. This calls for another factor to liberate Sm proteins from the pICln-imposed inhibition and render them competent to form snRNPs.

Concurrent Transfer of Sm Proteins onto the SMN Complex and Dissociation of pICln

As the SMN complex is known to promote snRNP assembly in vivo (Massenet et al., 2002; Meister et al., 2001a; Meister and Fischer, 2002; Pellizzoni et al., 2002; Shpargel and Matera, 2005), we speculated that it might relieve Sm proteins from the pICln-induced inhibition. We addressed this hypothesis by incubating preformed recombinant 6S and pICln/D3/B complexes with immobilized, affinity-purified SMN complexes (see Figure 3A for protein compositions of the complexes involved). After extensive washing, the retained proteins were analyzed by SDS-PAGE and western blotting. As shown in Figure 3B, Sm protein transfer occurred efficiently in both cases, as evident by an increase of the respective Sm protein levels on the SMN complex (Figure 3B, lanes 3, 4, 7, and 8, and S2; note that Sm B comigrates with the light chain of the antibody). This reaction was strictly dependent on the SMN complex and could not be observed on a control column (Figures 3B, lanes 23 and 24, and S3, lanes 11 and 12). Furthermore, we found that the transfer reaction did not require prior methylation of Sm proteins, as these were obtained by bacterial expression.

Next, we addressed whether the tail domains of Sm proteins, i.e., the sites of sDMA modification were required for SMN complex binding. For this, we performed the transfer reaction with a reconstituted pICln/D3/B complex in which the tail domains of both Sm proteins were absent (referred to as pICln/D3-Sm/B-Sm). The SMN complex efficiently accepted this truncated D3/B heterodimer, comparable to the Sm proteins of the 6S complex (Figure S3, lanes 3, 4, 7, and 8). Having established that Sm proteins are transferred from pICln complexes to the SMN complex irrespective of their methylation status, we analyzed the fate of pICln in this reaction. By western blotting, pICln was shown to be absent from the recipient SMN complex, suggesting its dissociation from Sm proteins during transfer (Figures 3B, lanes 3 and 4, and S3, lanes 3 and 4, pICln immunoblot panel in the bottom of the figure).

We next asked which subunits of the SMN complex participated in the Sm protein transfer reaction. As a prerequisite to address this question, we expressed two recombinant SMN complexes that lack Gemins 3, 4, and 5 (SMNΔGemin3–5) or Gemins 3, 4, 5, 6, and 7 (SMNΔGemin3–7) and purified them to near homogeneity (Figure 3A; see Supplemental Data for details). These units were immobilized onto an anti-SMN antibody resin, and transfer reactions were performed as described. Surprisingly, both units took over the Sm proteins in a manner indistinguishable from the endogenous complex, concurrent with the dissociation of pICln (Figure 3B, lanes 11, 12, 15, and 16). Hence, loading of the SMN complex with Sm proteins does not require Gemins 3–7.

However, when the SMN complex was further reduced to a heterodimer composed of Gemin2 and an N-terminal fragment of SMN containing residues 1–160 (termed Gemin2/SMNΔC), the transfer reaction was stalled at an intermediate stage. Although Gemin2/SMNΔC still interacted with the 6S complex, the extent of pICln dissociation was reduced (see arrow in Figure 3B, lane 19 and pICln Immunoblot panel). In fact, this stalled transfer intermediate was stable and could be reconstituted with the recombinant proteins pICln, Gemin2/SMNΔC, D1/D2, and E/F/G (Figure S4, hereafter referred to as the 8S complex). Furthermore, transfer of the Sm proteins D3/B from pICln onto the Gemin2/SMNΔC heterodimer was severely impaired (Figure 3B, lane 20). Taken together, these experiments reveal that a heterotrimeric complex composed of SMN, Gemin2, and 8 is sufficient to accept all seven Sm proteins from pICln. However, the simultaneous removal of Gemin8 and the C-terminal part of SMN from this minimal system reduces the capacity to dissociate pICln from Sm proteins and recruit the Sm proteins D3/B.

Reversal of a Kinetic Trap upon Transfer of Sm Proteins from 6S onto the SMN Complex

The findings above show that Sm proteins are incapable of associating with RNA while bound to pICln (Figure 2E). The activity of the SMN complex, in turn, leads to the displacement of pICln (Figure 3B). Hence, we considered that this activity is a major prerequisite for the subsequent formation of snRNPs. In a first approach, we analyzed the assembly of U1snRNP in the presence of endogenous SMN complex affinity purified from cytosolic extract. As this complex is active per se (Figure 4A,

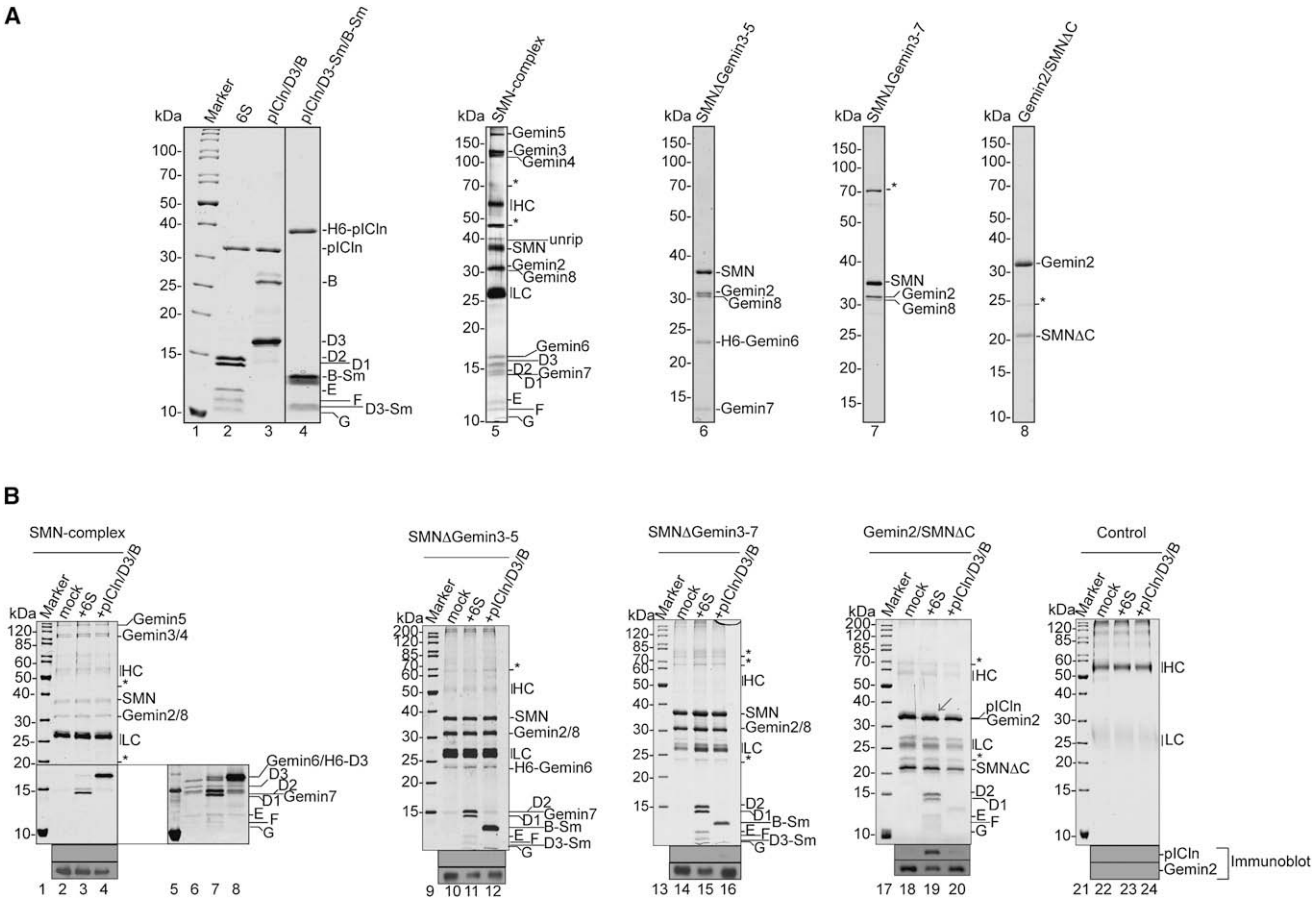


Figure 3. Transfer of Sm Proteins from pICln onto the SMN Complex

(A) SDS-PAGE of protein complexes used for the transfer reactions: recombinant 6S (lane 2), pICln/D3/B (lane 3), and pICln/D3-Sm/B-Sm (lane 4) complexes. Lane 5 shows affinity-purified endogenous SMN complex. Lanes 6–8 show the recombinant SMN subcomplexes SMNΔGemin3–5 (lane 6), SMNΔGemin3–7 (lane 7), and Gemin2/SMNΔC. Bands marked with an asterisk indicate nonspecific proteins; HC and LC indicate heavy and light chains of the antibody used for affinity purification.

(B) Transfer of Sm proteins from pICln complexes to the SMN complex. The respective, immobilized complexes (labeled on top) were incubated with either the 6S complex (lanes 3, 7, 11, 15, and 19), pICln/D3/B (lanes 4, 8, 12, 16, and 20), or treated with buffer only (lanes 2, 6, 10, 14, and 18). A control purification from HeLa extract was incubated with the 6S complex (lane 23), pICln/D3/B (lane 24), or buffer (lane 22) as a control. Proteins retained on the resin were analyzed by SDS-PAGE and western blotting with anti-pICln and anti-Gemin2 antibodies.

lane 4), a preincubation with a large molar excess of nonlabeled U1snRNA was performed to exhaust its inherent assembly activity (see Figure S5 for details). SMN complex treated this way failed to promote the formation of snRNPs on ^{32}P-labeled U1snRNA (Figure 4A, lane 2). However, a simultaneous preincubation of the activity depleted SMN complex with 6S complex, and pICln/D3/B led to its reactivation and enabled assembly of the Sm core domain on U1snRNA (Figure 4A, lane 3). No RNPs were formed on U1snRNA lacking the Sm site (U1ΔSm) (Figure 4A, lanes 5–8).

Next, the same reaction was performed with recombinant SMNΔGemin3–5 and SMNΔGemin3–7 complexes. As shown in Figures 4B and 4C, both failed to assemble U1snRNPs when Sm proteins were lacking (lanes 2). However, when loaded with 6S and pICln/D3/B or with 6S alone, they induced assembly of Sm cores and subcores, respectively (Figures 4B and 4C, lanes 3 and 4). No assembly was observed when the complexes were loaded with either pICln/D3/B alone (Figures 4B and 4C, lanes 5) or when any of the complexes were incubated with U1ΔSm (Figures 4B and 4C, lanes 6–10). Thus, along with experiments described in Figure 3B, we conclude that Sm proteins regain the competence to form snRNPs as a consequence of their transfer onto the SMN complex and the concomitant dissociation of pICln.

In contrast, we failed to activate the minimal Gemin2/SMNΔC dimer for snRNP assembly upon incubation with either 6S or pICln/D3/B complex (Figure 4D, lanes 2–4). As described in Figure 3B, the incubation of Gemin2/SMNΔC with the 6S complex results in a stalled transfer reaction intermediate. We hypothesized that removal of pICln would resolve this stalled state and enable Sm subcore formation. To address this, a complex composed of Gemin2/SMNΔC/D1/D2/E/F/G was reconstituted from recombinant proteins (Figure S6, hereafter referred to as the 7S complex). Indeed, this complex promoted efficient

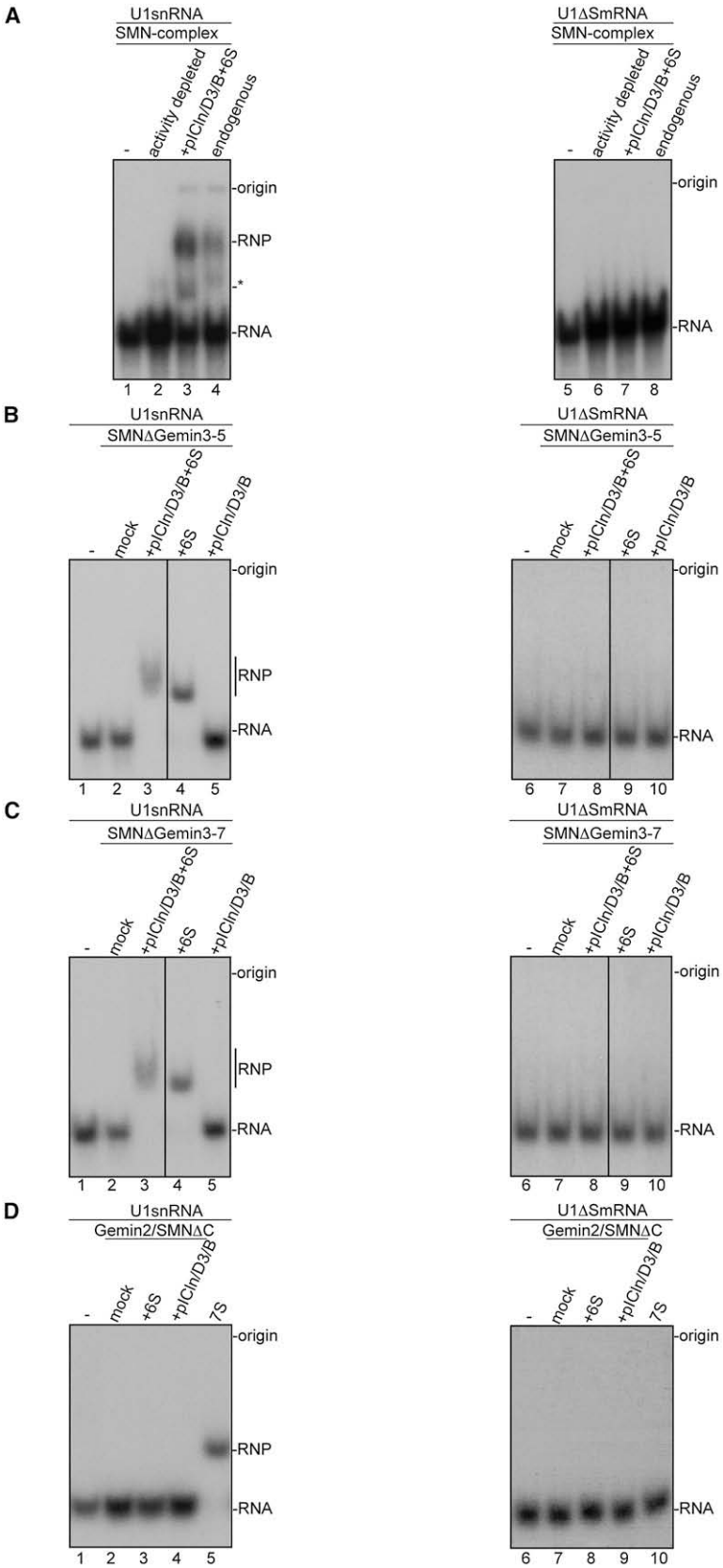


Figure 4. SnRNP Core Formation Correlates with the Capability of the SMN Complex to Dissociate pICln from Sm Proteins

(A) Immobilized, affinity-purified SMN complex was either mock treated ("endogenous") or incubated with an excess of unlabeled U1snRNA to exhaust its assembly capacity ("activity-depleted"). After washing, both complexes were incubated with ^{32}P-labeled U1snRNA, and complex formation was analyzed by native gel electrophoresis (lanes 2 and 4). In lane 3, activity-depleted SMN complex was incubated with 6S and pICln/D3/B complexes and washed before addition of ^{32}P-labeled U1snRNA. Lanes 5–8 show control reactions on ^{32}P-labeled U1ΔSmRNA; lanes 1 and 5 show RNA only. The band marked with "RNP" corresponds to the Sm core; the asterisk denotes an RNP lacking Sm proteins.

(B) Assembly reactions on ^{32}P-labeled U1snRNA with recombinant SMNΔGemin3–5 complex without bound Sm proteins (lane 2) or loaded with the indicated pICln-Sm complexes (lanes 3–5). Lane 1 shows RNA only; lanes 6–10 show control assembly reactions on ^{32}P-labeled U1ΔSmRNA.

(C) Assembly reactions as in (B) but with SMNΔGemin3–7.

(D) Assembly reaction as in (B) using the Gemin2/SMNΔC dimer without Sm proteins (lanes 2 and 7) or preincubated with 6S (lanes 3 and 8) and pICln/D3/B (lanes 4 and 9), respectively. An in vitro reconstituted complex composed of Gemin2/SMNΔC/D1/D2/E/F/G was incubated with U1snRNA (lane 5) or U1ΔSmRNA (lane 10) prior to native gel electrophoresis.

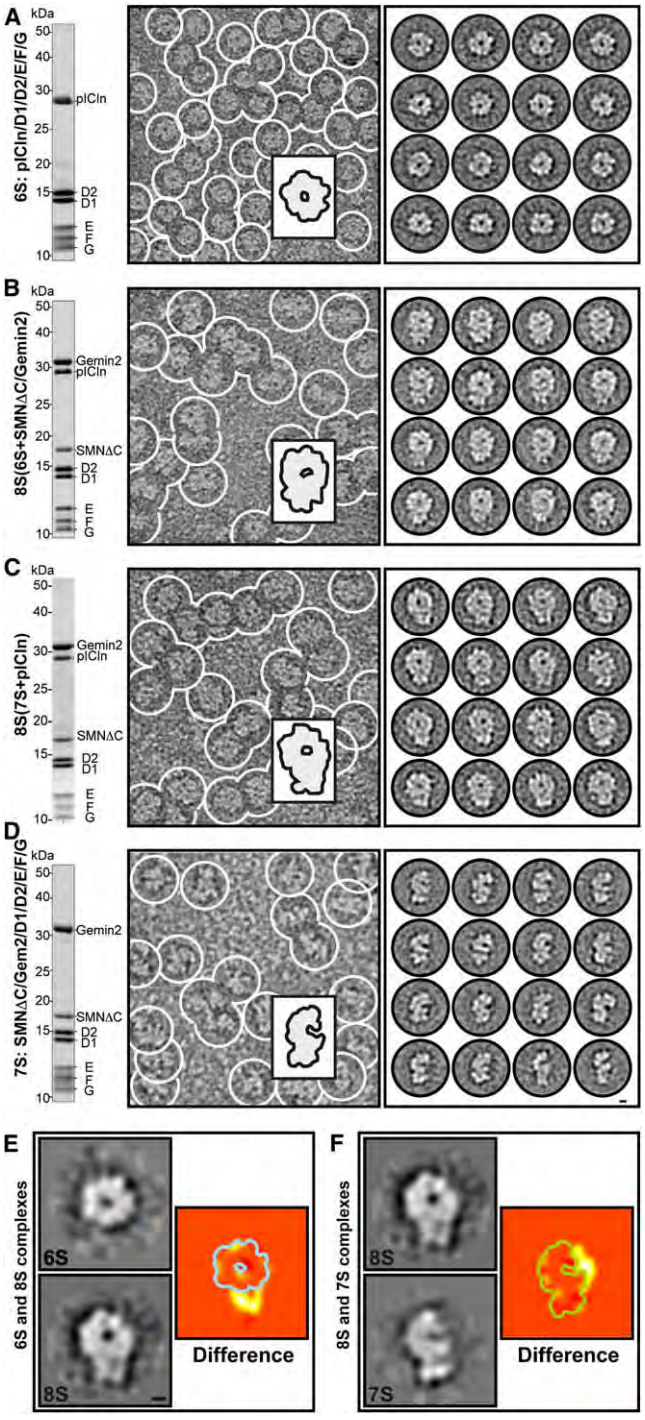


Figure 5. Electron Microscopic Analysis of Assembly Intermediates

(A–D) Reconstituted assembly intermediates (for SDS-PAGE of the respective complexes, see left panel) were subjected to EM analysis: (A) 6S, (B) and (C) 8S, and (D) 7S complexes. Note that the 8S complex was either reconstituted by incubating 6S complexes with SMNΔC/Gemin2 (B) or adding pICln to 7S complexes (C). Characteristic raw images and class averages of assembly intermediates are shown in the middle and right panel, respectively. The insets depict a schematic representation of particle morphology.

assembly of the Sm subcore in a strictly Sm site-dependent manner (Figure 4D, lanes 5 and 10). From these experiments we conclude that Sm core formation directly correlates with the capability of the SMN complex to dissociate pICln from the Sm proteins. Furthermore, assembly of the Sm core per se requires only a minimal set of subunits of the SMN complex in vitro, namely SMN and Gemin2. Our recent finding of a minimal SMN complex in *Drosophila melanogaster* composed of the same two proteins further corroborates the latter conclusion (Kroiss et al., 2008 and Supplemental Discussion).

Structural Analysis of Assembly Intermediates Suggests a Mechanism for SnRNP Core Formation

As both the 6S (pICln/D1/D2/E/F/G) and 7S (Gemin2/SMNΔC/D1/D2/E/F/G) complexes contain the same set of Sm proteins, it was intriguing to ask what makes the first one incompetent and the latter suitable for Sm subcore formation. We reasoned that the transition from the inactive to an activated state is accompanied by structural rearrangements. To address this, we conducted electron microscopic (EM) studies of the reconstituted 6S complex; the stalled Sm protein transfer intermediate from 6S to SMN complexes, designated as the 8S complex (Gemin2/SMNΔC/pICln/D1/D2/E/F/G); and the 7S complex as a subunit of the SMN complex functional in RNP formation. Due to their decreased susceptibility to proteolysis, we turned to the highly homologous *Drosophila* counterparts of the respective complexes. As shown in Figures 5A–5D, left panel, these complexes could be efficiently reconstituted from single proteins and purified to homogeneity. Gel filtration of these complexes indicated molecular masses consistent with a monomeric stoichiometry of each constituent. Moreover, the *Drosophila* complexes conform to the identical biochemical characteristics described above for their respective human counterparts (data not shown).

Typical negatively stained EM fields of all three purified, reconstituted complexes revealed a monodisperse population of asymmetric single particles with well-defined, distinct structural features (Figures 5A–5D, middle panels). All four analyzed data sets (6S, 8S[6S+SMNΔCGemin2], 8S[7S+pICln], 7S particles) showed preferential binding orientation of the particles on the carbon film (87%–94% of the particle views; see Figure S7 for details). Class averages of the 6S complex exhibit a ring-like architecture with a central accumulation of stain (Figure 5A, middle

(E and F) Difference maps identify shared structural elements and lacking domains in the Sm-containing complexes. The shapes of the respective complexes are shown on the left side and the difference maps on the right side of each panel. Differences are visible as bright yellow regions within the map. For comparison, the contours of the 6S (blue) and 7S (green) complexes are also drawn in the difference maps.

(E) A comparison of the 6S complex and the 8S transfer intermediate complex reveals shared features in the ring-shaped domain, while the foot-like domain of the 8S intermediate complex is lacking in the 6S complex. This suggests that SMN and Gemin2 form the foot-like domain.

(F) Difference map of the 8S transfer intermediate and the 7S complex shows common features in most parts of both ring-shaped and foot-like domains, while a minor portion of the ring-like domain is missing in the 7S complex. This suggests that pICln is located within the ring-like density element. Scale bar, 2 nm.

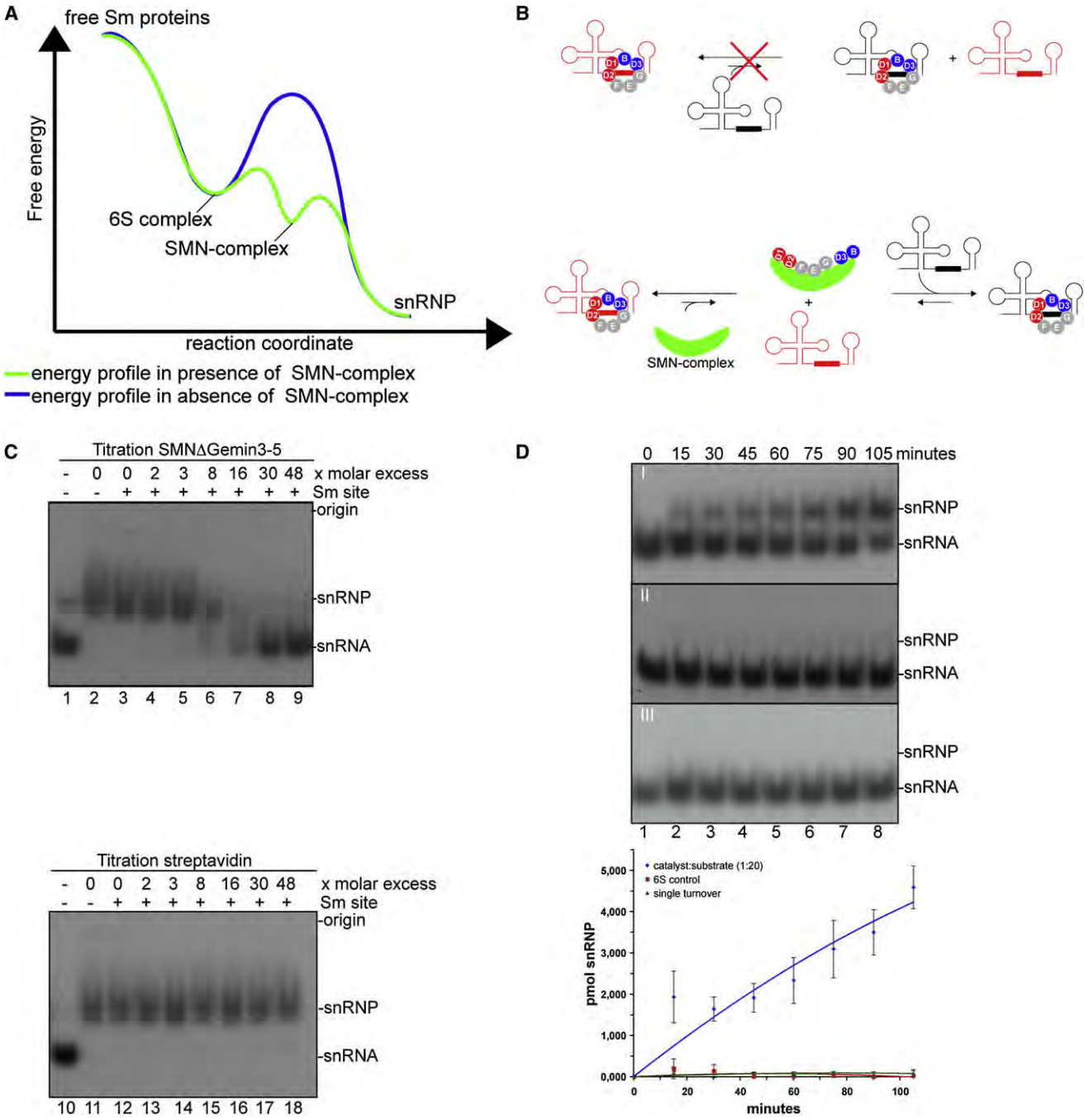


Figure 6. The SMN Complex Is a Catalyst of SnRNP Core Formation

(A) Hypothetical energy profile of snRNP core formation in vivo in the absence (blue line) and presence (green line) of the SMN complex. The model suggests that formation of the 6S complex results in a kinetic trap for Sm proteins and that the SMN complex acts as a catalyst for the transition from the 6S complex to the snRNP.

(B) Schematic of the experimental setup to test whether the SMN complex is a catalyst. Due to the stability of the Sm core domain, transfer of Sm proteins from an assembled core (red RNA) onto unlabeled competitor snRNA (black) is unlikely (upper panel). The hypothesized role of the SMN complex as a catalyst is illustrated in the lower panel.

(C) ^{32}P-labeled U1snRNA (lane 1) was incubated with Sm proteins to allow formation of the Sm core domain (lane 2). To the preformed core, a 330-fold molar excess of Sm-site RNA and increasing amounts of SMNΔGemin3–5 were added (lanes 3–9). The reactions were resolved on native gels 2.5 hr later. The same reactions were performed in the presence of increasing amounts of streptavidin instead of the SMNΔGemin3–5 complex (lanes 10–18).

(D) Panel I depicts RNP formation at the indicated time points with 10 pmol each of the 6S complex and RNA and 0.5 pmol of the SMNΔGemin3–5 complex (a catalyst:substrate ratio of 1:20). Panel II shows assembly of 10 pmol U1snRNA and 0.5 pmol SMNΔGemin3–5 complex preloaded with the Sm proteins D1/D2

and right panels). These particles had a maximum diameter of about 8–8.5 nm. Notably, these dimensions are comparable to those reported for the Sm core domain (Kastner and Lührmann, 1989). In contrast, the shape of assembly active 7S complexes was strikingly different (Figure 5D, middle and right panels). Class averages showed elongated particles with a length of about 11.5 nm. Importantly, the closed ring observed in 6S complexes was now converted into an open clamp (upper domain), which is packed against a foot-like protuberance (lower domain). Thereby, the opening of the clamp (oriented to the right side in Figure 5D) was found in proximity to the foot-like domain. The transfer intermediate 8S complex still contained the ring shape of the 6S complex (upper domain) but displayed, additionally, a foot-like protuberance as the 7S complex (lower domain) (Figures 5B and 5C, middle and right panels). Notably, the ring domain of this particle had a diameter of about 8–8.5 nm and, hence, has very similar dimensions as the 6S complex. The overall length of these particles was about 11.5 nm, similar to that of 7S complexes. To investigate whether the closed and open states are the result of different conformations or of different orientations on the carbon film, we also tilted the specimen in the electron microscope. In support of our model, this did not change the open or closed appearances of individual particles (data not shown).

Since the 8S complex comprises both structural features of 6S and 7S complexes and all three complexes contain the five Sm proteins D1, D2, E, F, and G as a common compositional denominator, we determined the positions of accessory factors by structural comparison. In particular, the similarity of distinct domains is not only restricted to the overall shape, but is also seen in fine structural details. All four samples were analyzed independently of the respective other three to enable an unbiased structural characterization of the complexes. Two-dimensional difference maps of the complexes elucidated shared structural elements as well as differences (Figures 5E and 5F). 6S and 8S complexes share the globular ring-like domain (Figure 5E; for comparison, the shape of the 6S complex is shown as a blue line in the difference map). We noted minor changes in the ring-like domain, which most likely represent conformational changes or slightly varied projections and large deviances in the foot-like protuberance of the 8S complex (yellow regions). We attribute the latter to a compositional difference and, therefore, suggest that the foot-like domain is the position of Gemin2 and SMNΔC, as these two proteins are the only ones lacking in the 6S complex.

Difference maps of the 8S intermediate and the 7S complex showed common structural elements in the majority of the ring-like upper domain and foot-like domain in the lower region of the complex (Figure 5F). A minor portion of the circular domain, however, is missing in the 7S complex, suggesting that this is the position occupied by pICln, which is the compositional difference between the 8S and 7S complex (Figure 5F; for comparison, the contour of the 7S complex is shown as a green line in the difference map). We also observed small differences along the body of the particle, which we yet again attribute to conformational changes or slightly different adsorption orientations on the EM grid. Therefore, we suggest that pICln is an integral component of the ring in the 6S complex, while Gemin2 and SMN occupy more distal positions on the convex face of an opened ring in the 7S complex.

The SMN Complex Is a Catalyst of SnRNP Core Formation

SnRNP core formation, albeit a thermodynamically favorable reaction (Raker et al., 1996), depends on PRMT5 and SMN complexes in vivo (Meister et al., 2002; Paushkin et al., 2002). We reasoned that these *trans*-acting factors are required to modulate the energetics of key steps of this reaction. To address this hypothesis, we deduced the energy profile of the in vivo Sm core assembly reaction, taking the reaction intermediates described in Figures 1–5 into account. First, we considered the energy levels of newly synthesized Sm proteins (the substrates of the reaction) and the assembled Sm core (the product of the reaction). Assuming that snRNP assembly in cells is, likewise, a spontaneous reaction, we suggest that free Sm proteins must occupy high energy levels along the reaction coordinate, whereas the Sm core would reflect a global energetic minimum. The 6S complex must reside in a local minimum along the reaction coordinate since it is stable under steady-state conditions in extracts. We further postulate that Sm proteins are inhibited from forming Sm cores out of the pICln-bound state due to a very high activation energy barrier (Figure 6A, blue line). On the basis of experiments shown in Figures 3 and 4, we propose that the SMN complex helps Sm proteins surmount this unfavorable energetic obstacle (Figure 6A, green line). A consequence of these arguments is that, first, the SMN complex loaded with Sm proteins occupies another local yet transient energetic minimum and, second, that the SMN complex acts as a catalyst for snRNP assembly.

Two strategies were employed to test this assumption experimentally. The first relied on the observation that snRNP cores are exceedingly stable structures, capable of withstanding treatment with 7 M urea or 2.5 M NaCl (Liautard et al., 1982). Thus, spontaneous exchange of Sm proteins from an assembled snRNP core to a free snRNA is unlikely (Figure 6B, top panel). However, if the prediction of our energetic considerations holds true, the exchange of Sm proteins should be possible upon addition of a molar excess of the SMN complex over the Sm proteins of the snRNP (Figure 6B, bottom panel). As concentrated solutions of catalyst are required for this type of experiment, we used recombinant SMNΔGemin3–5 complex rather than the endogenous SMN complex. Indeed, addition of an 8-fold molar excess of SMNΔGemin3–5 caused a half-maximal exchange of Sm proteins from the preformed snRNP to the unlabeled RNA (Figure 6C, lane 6), with a total exchange at a 30-fold molar excess (Figure 6C, lane 8). This effect was strictly dependent on the SMN and Gemin2 subunits of the SMN complex, as incubation with unlabeled RNA alone (Figure 6C, lanes 3 and 12) or with nonrelevant proteins in the presence of unlabeled RNA had no effect on the stability of the Sm core

and E/F/G, corresponding to single-turnover conditions. A control assembly reaction with 10 pmol 6S complex and RNA incubated in the absence of the SMNΔGemin3–5 complex is shown in panel III. The bottom panel shows quantification of the experiments. Data are represented as mean ± SD (n = 3).

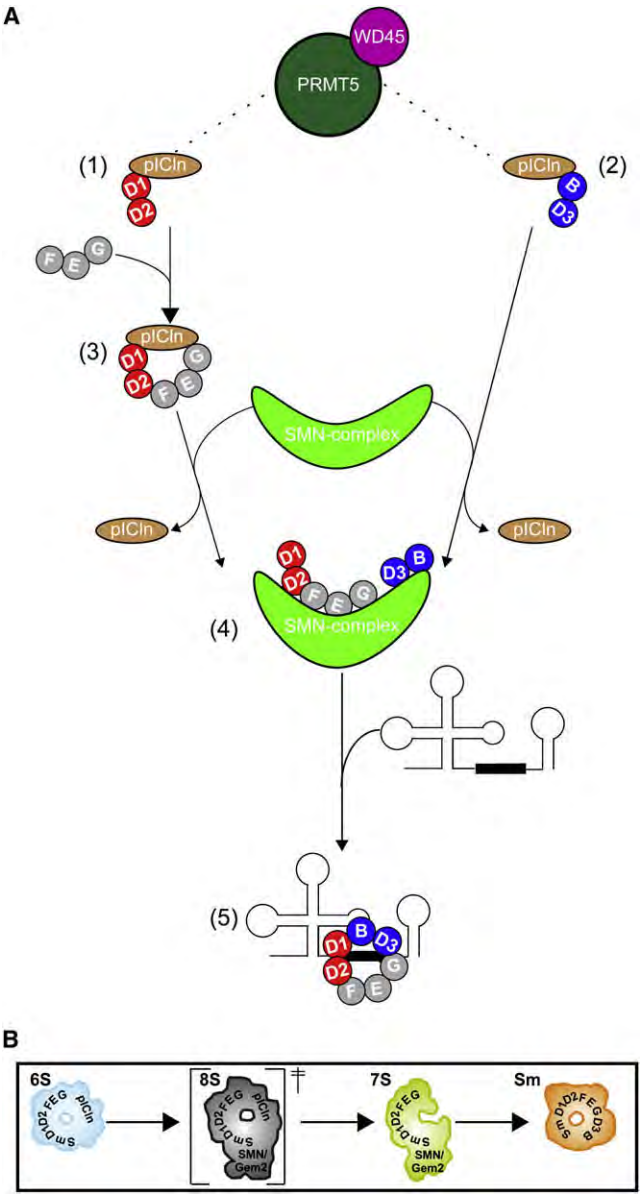


Figure 7. Model of the Assisted Assembly of SnRNPs and Structures of Key Intermediates

(A) The chaperone pICln initially binds D1/D2 and D3/B (1 and 2). EFG is then recruited to pICln/D1/D2 to form the 6S complex (3). Transfer of pICln-bound Sm proteins onto the SMN complex coincides with displacement of pICln and forms the loaded SMN complex (4). The SMN complex then allows RNA binding, ring closure, and RNP release (5).

(B) Structural EM models of intermediates of the assembly reaction. 6S (see also A3) forms a closed ring, 7S holds the Sm proteins in an open configuration (part of A4), and 8S represents a trapped transition complex from A3 to A4. The assembled core is derived from EM models of negatively stained snRNP cores.

(Figure 6C, lanes 13–18; see also Figure S8 for additional experiments). Of note, this reaction followed dose-dependent saturation kinetics (data not shown).

In the second approach, we asked whether the SMN complex is a catalyst of assembly reaction even when present in substoichiometric amounts and is, thus, capable of multiple turnover reactions. To test this possibility, we initially incubated a small amount of SMN complex (0.5 pmol) loaded with Sm proteins with a 20-fold molar excess of U1snRNA. To visualize the assembly reaction in native gels, 0.5% of the RNA substrate was radiolabeled. Under these conditions, the SMN complex delivered its Sm proteins onto the U1snRNA. However, due to the small amount of substrate (i.e., Sm proteins) used in this assay and the low specific activity of the U1snRNA, only marginal assembly could be observed (Figure 6D, panel II and lower panel green line). However, upon addition of a 20-fold molar excess (10 pmol) of the 6S complex, assembly was stimulated in a time-dependent manner (Figure 6D, panel I and lower panel blue line). Under these conditions, 2 pmol RNPs were generated per hour, directly showing multiple turnover of the substrate (Sm proteins and snRNA) to product. No assembly was observed when 6S was incubated with U1snRNA alone, showing the dependence of this reaction on the SMN complex (Figure 6D, panel III and lower panel red line). We conclude that the SMN complex facilitates snRNP assembly by (1) lowering the activation energy of the reaction from 6S to snRNP and (2) loading Sm proteins onto snRNA in a multiple turnover reaction. These results are consistent with a function of the SMN complex as a catalyst of snRNP formation.

DISCUSSION

In this study, we have applied a combination of biochemical and structural approaches to decipher key steps in the assisted assembly of snRNPs. We show that an assembly chaperone, pICln, initially induces the formation of otherwise unstable higher-order Sm protein complexes. In this state, a kinetic trap is imposed on the Sm proteins by bound pICln, which actively prevents their premature association with snRNA. Sm protein progression toward the snRNP core is mediated by the SMN complex that dissociates the inhibitory chaperone and catalyzes ring closure on snRNA. Our postulated mechanism is schematically depicted in Figure 7.

pICln Is an Assembly Chaperone for Sm Proteins

pICln has been described as a component of the PRMT5 complex that recruits Sm protein substrates to the methyltransferase PRMT5 (Friesen et al., 2001; Meister et al., 2001b). Our data reveal a second function for pICln as an organizer of assembly intermediates. The most stable complex contains the Sm proteins D1/D2 and E/F/G, which are tethered upon pICln binding into an RNA-free ring structure (Figure 7A, complex 3). Importantly, the resulting Sm-heteropentamer cannot form in the absence of pICln, as D1/D2 and E/F/G have little affinity for each other in the absence of snRNA (Raker et al., 1996). How does pICln induce the formation of this otherwise unstable Sm-heteropentamer? The solution structure of an N-terminal fragment of pICln shows that it adopts a fold with several antiparallel β strands, designated as a pleckstrin homology domain (Furst et al., 2005). Sm proteins have been reported to associate with each other by antiparallel β strands (Kambach et al., 1999). Thus, we suggest that pICln in the RNA-free 6S complex acts as a structural counterpart of D3/B in the assembled Sm core domain

(Kambach et al., 1999; Stark et al., 2001). The architecture of the 6S complex as determined by electron microscopy is consistent with this hypothesis; however, the validation of this model awaits the elucidation of high-resolution structures.

Pulse-chase experiments performed by G. Blobel and coworkers demonstrated an RNA-free 6S precursor of snRNP cores more than two decades ago (Fisher et al., 1985). This precursor was reported to form early in biogenesis and contained the Sm proteins E, F, G, and D (only one D protein was known at that time). Based on its apparent sedimentation coefficient, composition, and role in assembly, the pICln/D1/D2/E/F/G complex described here might be identical to this long-sought precursor.

pICln formed a second complex containing D3/B in vitro (Figure 7A, complex 2), which could not be isolated from cellular extracts as a separate entity. However, the Sm proteins D3/B are abundantly present in the isolated 20S PRMT5 complex (Figure 1B). Hence, we believe that pICln/D3/B is also formed in vivo but cannot be detected in extracts in steady state because it is either short lived or remains stably bound to the methyltransferase complex.

Consistent with studies in *Xenopus laevis* oocytes (Pu et al., 1999), we find Sm proteins incompetent of forming snRNPs in a pICln-bound state. The architecture of the 6S particle suggests that pICln sterically prevents the RNA from gaining access to its binding site on the inside of the Sm ring (Kambach et al., 1999; Stark et al., 2001; Urlaub et al., 2001). Based on the capability to promote higher-order assemblies and to simultaneously prevent premature RNA binding, pICln is an assembly chaperone specific for Sm proteins. Although similar factors have been described for the assembly of protein complexes (Hirano et al., 2006; Laskey et al., 1978; Le Tallec et al., 2007; Saschenbrecker et al., 2007), to our knowledge, pICln is the first protein of this class implicated in the assembly of RNPs. We predict that other RNPs containing Sm proteins in their cores, such as the U7snRNP, likewise require pICln as an assembly chaperone (Pillai et al., 2003).

The SMN Complex Acts as a Catalyst to Resolve a Kinetic Trap in Assembly

We have previously reported that the SMN complex is copurified upon immunopurification of pICln from assembly-active extracts, albeit in a substoichiometric manner (Meister and Fischer, 2002). This finding suggested Sm proteins to be transferred directly from pICln to the SMN complex. Indeed, here, we show that the SMN complex accepts Sm proteins from both 6S and pICln/D3/B complexes and simultaneously dissociates pICln. As a consequence of this reaction, Sm proteins regain the competence to form snRNP particles. Of note, neither the two other subunits of the PRMT5 complex nor arginine methylation of Sm proteins appears to be essential for Sm protein transfer onto the SMN complex in vitro. However, we do not exclude at this stage that this posttranslational modification might modulate the kinetics or efficiency of this reaction.

We suggest that the failure of Sm proteins to interact with snRNA as part of the 6S complex is associated with a very high activation energy barrier in the transition from 6S to the assembled snRNP (Figure 6A). The capability of the SMN complex to accept Sm proteins from this state and transfer them onto RNA implies that it lowers this energetic hurdle and, therefore, is a catalyst of the reaction (Figure 6A). In favor of this idea, a moderate molar excess of the SMN complex was capable of melting a preformed snRNP core, an exceedingly thermodynamically stable entity (Liautard et al., 1982). Furthermore, the SMN complex acts substoichiometrically in snRNP formation and, hence, exhibits multiple turnover (Figure 6D). These findings indicate that snRNP assembly is a catalyzed reaction in vivo, with the SMN complex bound to Sm proteins, representing a transition-state intermediate (Figure 7A, complex 4). Considering the high abundance of Sm proteins in the cytoplasm, the nuclear export of snRNA is likely to be the global rate-limiting step of snRNP assembly. This unbalanced availability of both substrates may explain the stability of the SMN complex under steady-state conditions.

An SMN complex composed of only SMN and Gemins 2 and 8 appears to be sufficient for both aforementioned activities (Figures 3 and 4). However, performing the transfer reaction with Gemin2/SMNΔC has allowed us to uncouple 6S complex binding and pICln release (Figure 3B). Also, this complex failed to efficiently accept pICln/D3/B (Figure 3B). From these experiments, we conclude that either Gemin8 and/or the C-terminal portion of SMN affects pICln release and, simultaneously, constitutes the binding site for Sm D3/B.

We obtained further mechanistic insight into the role of the SMN complex in snRNP core formation by electron microscopy of assembly intermediates (Figure 5). When the architecture of the stalled 8S transfer intermediate was compared to 6S and 7S complexes, we found that SMN and Gemin2 occupy positions on the outer face of the pICln-Sm protein ring. Thus, the transfer reaction relieves Sm proteins of the pICln-imposed steric hindrance and enables the RNA to access its binding site on the inner face of the Sm protein ring. How is pICln dissociation affected? Binding of the SMN complex onto the outer surface of the 6S ring may induce a conformational change in the Sm proteins of the 6S complex that results in pICln release. The SMN complex then holds Sm proteins in an open-ring configuration, poised for transfer onto snRNA. These findings also suggest that the spatial organization of Sm proteins in the snRNP is attained already at the level of the 6S complex, i.e., in an early stage of assembly (Figure 7B).

Structural Similarity of SnRNP Assembly with DNA-Clamp Loading

Based on our studies, we hypothesize that the SMN complex-assisted formation of snRNPs is structurally similar to the clamp-loading reaction, i.e., the process by which ring-shaped clamps are placed onto DNA to tether poorly processive DNA replication enzymes (Ellison and Stillman, 2001; Indiani and O'Donnell, 2006). Clamp loaders are hetero-oligomeric protein assemblies that bind and open hexameric clamp rings in an ATP-dependent manner. ATP hydrolysis then induces a conformational change in the loader, which leads to ring closure and ejects the clamp onto DNA (Davey et al., 2002; Johnson and O'Donnell, 2005). What then are the similarities to the formation of snRNP cores? The Sm core is a toroidal RNP (Kambach et al., 1999; Stark et al., 2001) and is thus structurally similar to

a replication clamp around DNA (Georgescu et al., 2008). The SMN complex holds the Sm proteins in an open state poised for transfer onto RNA (Figure 7B), reminiscent of the clamp-loading complex. Although the eight subunits making up the SMN complex and ATP are required in cells, a minimal complex composed of SMN and Gemin2 is sufficient to catalyze formation of the snRNP core in vitro in the absence of ATP. Similarly, a minimal *E. coli* clamp loader consisting of the δ subunit only transfers clamps onto DNA without a nucleotide requirement. Thus, it is intriguing to postulate that cells have evolved two separate machineries to deal with a similar topological problem in the joining of nucleic acids with proteins. Future studies will elucidate fine-tuning and regulation of this reaction.

EXPERIMENTAL PROCEDURES

Sm Protein Transfer Assays

Fifty picomoles of the indicated SMN complex were immobilized to 20 μl of an anti-SMN resin. One hundred picomoles of either 6S or pICln/D3/B complex were added and incubated at 37°C for 1 hr. After washing with 3 × 1 ml PBS, the retained proteins were eluted with SDS-PAGE loading buffer. 50% of this eluate was analyzed by SDS-PAGE. The rest was subjected to western blotting with both anti-pICln and anti-Gemin2 as a loading control.

Sm Protein Strand Exchange Assays

Sm cores were preformed using 3 pmol Sm proteins and 0.3 pmol ^{32}P-labeled snRNA as described (Raker et al., 1996). To this, 100 pmol of unlabeled Sm site oligonucleotide (AAUUUUUGA) and increasing amounts of SMNΔGemin3–5 complex were added. After incubation for 2.5 hr at 37°C, the mixtures were analyzed by native gel electrophoresis as described (Meister et al., 2001a). The indicated SMNΔGemin3–5 molar excesses over the Sm proteins were used.

Reconstitution of Sm Protein-Containing Complexes

Due to the strongly contrasting pI values of the assembly factors (pICln, SMN, and Gemin2) and Sm proteins, direct mixture of these proteins resulted in irreversible aggregation. To circumvent this, the respective proteins were mixed in 20 mM HEPES-NaOH (pH 7.5), 1 M NaCl, and 5 mM DTT and dialyzed overnight to 20 mM HEPES-NaOH (pH 7.5), 0.2 M NaCl, and 5 mM DTT at 4°C. Complex formation was then monitored by gel filtration run in 20 mM HEPES-NaOH (pH 7.5), 0.2 M NaCl, and 5 mM DTT, using a Superdex200 10/300 column (GE Healthcare) and subsequent SDS-PAGE of individual fractions.

Electron Microscopy and Single-Particle Image Processing

For EM sample preparation, purified complexes were subjected to the GraFix approach using a 5%–20% glycerol gradient (Kastner et al., 2008), and specimens were negatively stained with uranyl formate according to a previously published protocol (Golas et al., 2003). Images were taken on a Philips CM200 electron microscope (Philips/FEI, Eindhoven, The Netherlands) using a 4-fold binned 4 k × 4 k CCD camera (TemCam-F415, TVIPS, Gauting, Germany) at a magnification of 343,000- and 380,000-fold (Sander et al., 2005). 42014, 11126, 31150, and 27131 single-particle images were selected for the data sets of the 6S, 8S (6S+SMN/Gemin2), 8S (7S+pICln), and 7S complex, respectively. Using iteratively refined class averages, single-particle images were aligned via an exhaustive polar alignment (Sander et al., 2003) and subsequent multivariate statistical analysis (MSA)-based hierarchical ascendant classification (HAC) in the context of IMAGIC-5 (van Heel et al., 1996) into classes of 35–40 class members in average. All four data sets were processed independently of the other three; at no state of image analysis were the results of one data set used to refine another. After four iterations, the results were stable. Representative class averages were selected for presentation, and difference maps were calculated from the normalized, bandpass-filtered class averages using IMAGIC-5.

Additional Experimental Procedures can be found in the Supplemental Data.

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Discussion, Experimental Procedures, eight figures, and two tables and can be found with this article online at http://www.cell.com/supplemental/S0092-8674(08)01177-X.

ACKNOWLEDGMENTS

We thank T. Müller, P. Stayton, C. Kambach, J. Steitz, and I. Mattaj for reagents; J. Ohmer for technical assistance; and B. Laggerbauer, M. Kroiss, G. Matera, S. Jentsch, S. Rospert, F.U. Hartl, and R. Lührmann for comments. This work was supported by grants of the DFG (SFB581, TP B18) and AFM (ID11730) to U.F. and grants of the BMBF and European "3D Repertoire" to H.S.

Received: April 25, 2008
Revised: July 14, 2008
Accepted: September 8, 2008
Published: October 30, 2008

REFERENCES

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Lührmann, R. (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.

Davey, M.J., Jeruzalmi, D., Kuriyan, J., and O'Donnell, M. (2002). Motors and switches: AAA+ machines within the replisome. Nat. Rev. Mol. Cell Biol. *3*, 826–835.

Ellison, V., and Stillman, B. (2001). Opening of the clamp: an intimate view of an ATP-driven biological machine. Cell *106*, 655–660.

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

Fisher, D.E., Conner, G.E., Reeves, W.H., Wisniewolski, R., and Blobel, G. (1985). Small nuclear ribonucleoprotein particle assembly *in vivo*: demonstration of a 6S RNA-free core precursor and posttranslational modification. Cell *42*, 751–758.

Friesen, W.J., and Dreyfuss, G. (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.

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

Furst, J., Schedlbauer, A., Gandini, R., Garavaglia, M.L., Saino, S., Gschwentner, M., Sarg, B., Lindner, H., Jakab, M., Ritter, M., et al. (2005). ICln159 folds into a pleckstrin homology domain-like structure. Interaction with kinases and the splicing factor LSm4. J. Biol. Chem. *280*, 31276–31282.

Georgescu, R.E., Kim, S.S., Yurieva, O., Kuriyan, J., Kong, X.P., and O'Donnell, M. (2008). Structure of a sliding clamp on DNA. Cell *132*, 43–54.

Golas, M.M., Sander, B., Will, C.L., Lührmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. Science *300*, 980–984.

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

Hirano, Y., Hayashi, H., Iemura, S., Hendil, K.B., Niwa, S., Kishimoto, T., Kasahara, M., Natsume, T., Tanaka, K., and Murata, S. (2006). Cooperation of multiple chaperones required for the assembly of mammalian 20S proteasomes. Mol. Cell *24*, 977–984.

Indiani, C., and O'Donnell, M. (2006). The replication clamp-loading machine at work in the three domains of life. Nat. Rev. Mol. Cell Biol. *7*, 751–761.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu. Rev. Biochem. *74*, 283–315.

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

Kastner, B., and Lührmann, R. (1989). Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5′ RNA terminus. EMBO J. *8*, 277–286.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods *5*, 53–55.

Kroiss, M., Schultz, J., Wiesner, J., Chari, A., Sickmann, A., and Fischer, U. (2008). Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *105*, 10045–10050.

Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature *275*, 416–420.

Le Tallec, B., Barrault, M.B., Courbeyrette, R., Guerois, R., Marsolier-Kergoat, M.C., and Peyroche, A. (2007). 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and in mammals. Mol. Cell *27*, 660–674.

Liautard, J.P., Sri-Widada, J., Brunel, C., and Jeanteur, P. (1982). Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar structural domain of U1, U2, U4 and U5 small nuclear RNAs. J. Mol. Biol. *162*, 623–643.

Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W., and Dreyfuss, G. (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. Mol. Cell. Biol. *22*, 6533–6541.

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

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

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

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

Neuenkirchen, N., Chari, A., and Fischer, U. (2008). Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett. *582*, 1997–2003.

Paushkin, S., Gubitz, A.K., Massenet, S., and Dreyfuss, G. (2002). The SMN complex, an assemblyosome of ribonucleoproteins. Curr. Opin. Cell Biol. *14*, 305–312.

Pellizzoni, L. (2007). Chaperoning ribonucleoprotein biogenesis in health and disease. EMBO Rep. *8*, 340–345.

Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002). Essential role for the SMN complex in the specificity of snRNP assembly. Science *298*, 1775–1779.

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Lührmann, R., Fischer, U., and Schümperli, D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. *17*, 2321–2333.

Pu, W.T., Krapivinsky, G.B., Krapivinsky, L., and Clapham, D.E. (1999). pICln inhibits snRNP biogenesis by binding core spliceosomal proteins. Mol. Cell. Biol. *19*, 4113–4120.

Raker, V.A., Plessel, G., and Lührmann, R. (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.

Sander, B., Golas, M.M., and Stark, H. (2003). Corrim-based alignment for improved speed in single-particle image processing. J. Struct. Biol. *143*, 219–228.

Sander, B., Golas, M.M., and Stark, H. (2005). Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. J. Struct. Biol. *151*, 92–105.

Saschenbrecker, S., Bracher, A., Rao, K.V., Rao, B.V., Hartl, F.U., and Hayer-Hartl, M. (2007). Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. Cell *129*, 1189–1200.

Shpargel, K.B., and Matera, A.G. (2005). Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. Proc. Natl. Acad. Sci. USA *102*, 17372–17377.

Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature *409*, 539–542.

Urlaub, H., Raker, V.A., Kostka, S., and Lührmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. EMBO J. *20*, 187–196.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17–24.

Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. *13*, 290–301.

Winkler, C., Eggert, C., Gradl, D., Meister, G., Giegerich, M., Wedlich, D., Laggerbauer, B., and Fischer, U. (2005). Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. Genes Dev. *19*, 2320–2330.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585–600.

Cell, *Volume 135*

Supplemental Data

An Assembly Chaperone Collaborates with the SMN Complex to Generate Spliceosomal SnRNPs

Ashwin Chari, Monika M. Golas, Michael Klingenhäger, Nils Neuenkirchen, Bjoern Sander, Clemens Englbrecht, Albert Sickmann, Holger Stark, and Utz Fischer

Supplemental Data

Supplemental Discussion

Most Gemins are dispensable for Sm proteins uptake and formation of UsnRNP cores

The vertebrate SMN-complex is composed of eight subunits (SMN and Gemins 2-8). Recently, by performing exhaustive bioinformatic analyses using databases containing whole genome assemblies of diverse organisms, we have been able to delineate an ancestral complex composed of SMN and Gemin2 only (Kroiss et al., 2008). Furthermore, in the same report we were able to show by biochemical purification that certain arthropods such as *Drosophila melanogaster* have functional SMN-complexes composed only of these two subunits. This is even more surprising in the light that these organisms still contain genes encoding for distantly related putative orthologs of Gemins 3 and 5 in their genomes (Kroiss et al., 2008). Consistent with these findings, we report here the biochemical definition of a minimal functional unit in the human SMN-complex composed of SMN, Gemin2 and Gemin8 sufficient for UsnRNP core formation *in vitro* (Figures 3&4). Whether this complex can be reduced even further is currently unclear, as removal of Gemin8 results in the pronounced susceptibility of SMN to C-terminal proteolysis, and hence to a non-functional unit (data not shown and see also characterization of the SMNΔC/Gemin2 complex in Figures 3 and 4). The majority of Gemins hence appear not to be essentially required for Sm protein uptake from pICln-complexes and Sm protein delivery onto RNA *in vitro*. Instead, we suggest that they serve additional functions in other aspects of the biogenesis pathway. In support of this hypothesis, the binding and recruitment of the UsnRNA to the SMN-complex has been recently ascribed to the Gemin5 subunit in cooperation with Gemins 3 and 4 (Battle et al., 2007; Battle et al., 2006). What steps in biogenesis remain that could demand other subunits of the SMN-complex? Two essential steps in the biogenesis pathway follow UsnRNP core formation, hypermethylation of the m^7G-cap structure to the m_3G-cap and the nuclear import of the assembled UsnRNP (reviewed in Will and Lührmann, 2001). The first reaction generates one part of the bipartite nuclear-localization signal of UsnRNPs (Fischer and Lührmann, 1990; Hamm et al., 1990), while the second ensures the subcellular transport of the assembled UsnRNP to its site of function. Both steps have been convincingly shown to be dependent on UsnRNP core formation (Fischer et al., 1993; Mattaj, 1986), and we propose that the other subunits of the SMN-complex function to recruit the trans-acting factors necessary for these. Consistent with this

latter notion, the SMN-complex has been reported to be associated with the entire cytoplasmic biogenesis phase of UsnRNPs (Massenet et al., 2002). Moreover, the nuclear import of UsnRNPs has been shown to be coupled to that of the SMN-complex (Narayanan et al., 2004).

Supplemental Experimental Procedures

Table S1: Plasmids used in this Study

Plasmid Name	Comments
T7-U1RNA	Kind gift from Dr. Ian Mattaj, EMBL Heidelberg, linearized with *Bam*HI for run-off transcription (Hamm et al., 1987).
T7-U1ΔSmRNA	Kind gift from Dr. Ian Mattaj, EMBL Heidelberg, linearized with *Bam*HI for run-off transcription (Hamm et al., 1987).
pQE30-2Z	Kind gift from Dr. Christian Kambach, PSI Villigen, a pQE30 (Qiagen) derivative containing two N-terminal Z tags, followed by a His_6 tag and a TEV cleavage site (Zaric et al., 2005).
pET21a	Novagen.
pET28a	Novagen.
pETM30	Kind gift from Dr. Ario de Marco, EMBL Heidelberg.
pET28b*	Kind gift from Dr. Thomas Müller, University of Würzburg, a pET28b (Novagen) derivative with swapped *Nco*I and *Nde*I sites.
D1D2:pRK172	Kind gift from Dr. Christian Kambach, PSI Villigen (Kambach et al., 1999).
D3(fl)B(fl):pQE30	Kind gift from Dr. Christian Kambach, PSI Villigen (Kambach et al., 1999).
D3(75)B(91):pQE30	Kind gift from Dr. Christian Kambach, PSI Villigen (Kambach et al., 1999). Expression plasmid for a D3/B heterodimer containing only the Sm fold domains.
EFG:pET15b	Kind gift from Dr. Christian Kambach, PSI Villigen (Kambach et al., 1999).

Plasmid Name	Comments
pICln:pET28a	The human pICln open reading frame (ORF) was subcloned into the *Eco*RI and *Xho*I sites of pET28a from the pICln:pGEX-6P-1 construct (Meister et al., 2001b).
pICln:pQE30-2Z	This work, the pICln ORF was PCR amplified as a *Bam*HI-*Hin*dIII fragment and ligated in frame to the two N-terminal Z tags, the His_6 tag, and the TEV cleavage site of pQE30-2Z.
SMN(1-160):pGEX-6P-1	As described (Meister et al., 2000).
Gemin2:pET28a	As described (Meister et al., 2000).
SMN:pETM30	This work, the internal *Nco*I site in the human SMN ORF was removed by silent mutagenesis. The ORF was then PCR amplified as an *Nco*I-*Not*I fragment and cloned into the *Nco*I and *Not*I sites of pETM30. This yielded His_6GSTTEV-tagged SMN. Both tags can be removed by TEV protease digestion.
Gemin2:pET28b*	This work, the human Gemin2 ORF was PCR amplified as an *Nde*I-*Not*I fragment and cloned into *Nde*I and *Not*I sites of pET28b*.
Gemin8:pET28b*	This work, the human Gemin8 ORF was PCR amplified as an *Nde*I-*Not*I fragment and cloned into *Nde*I and *Not*I sites of pET28b*.
Gemin6:pET28a	As described (Grimmler et al., 2005).
Gemin7:pET21a	This work, the human Gemin7 ORF was PCR amplified as an *Nde*I-*Not*I fragment and cloned into the *Nde*I and *Not*I sites of pET21a.
Gemin7-Gemin6:pET21a	This work, see cloning of the bicistronic Gemin7-Gemin6:pET21a expression plasmid for details (below). The Gemin6 ORF is in frame to a His_6-tag in this variant, while Gemin7 is untagged.

Plasmid Name	Comments
Gemin2-Gemin8- His_6GSTTEV-SMN:pET28b*	This work, see cloning of the tricistronic Gemin2-Gemin8-SMN:pET28b* expression plasmid for details (below). As described above SMN is in frame to His_6-GST tag and both tags are removable by TEV digestion. Gemin2 and Gemin8 are both untagged.
dpICln:pET28a	This work, the *Drosophila melanogaster* pICln cDNA was obtained from the Drosphila Genomics Resource Center (DGRC), Bloomington, Indiana. The dpICln ORF was PCR amplified as an *Nco*I-*Not*I fragment and cloned into *Nco*I and *Not*I sites of pET28a.
dSMN(1-122):pETM30	This work and Kroiss et al. (2008), the *Drosophila melanogaster* SMN cDNA was obtained from the DGRC, Bloomington, Indiana. The nucleotide sequence encoding for dSMN(1-122) was PCR amplified as an *Nco*I-*Not*I fragment and cloned into the *Nco*I and *Not*I sites of pETM30.
dGemin2:pETM30	This work and Kroiss et al. (2008), the *Drosophila melanogaster* Gemin2 cDNA was obtained from the DGRC, Bloomington, Indiana. The dGemin2 ORF was PCR amplified as an *Nco*I-*Not*I fragment and cloned into *Nco*I and *Not*I sites of pETM30.
SA(WT):pUC18	Kind gift from Dr. Patrick Stayton, University of Washington. Contains the coding sequence for the Streptavidin core.
SA(WT):pET28b*	This work, the Streptavidin core ORF was subcloned as an *Nde*I-*Hin*dIII fragment from the SA(WT):pUC18 construct into *Nde*I and *Hin*dIII sites of pET28b*.

Table S2: Antibodies Used in this Study

Antibody	Comments
Anti-PRMT5 (rabbit polyclonal)	Antibodies were generated by immunizing rabbits with an N-terminal His_6- tagged fragment of PRMT5 (residues 1-291; Immunoglobe, Himmelstadt). Antibodies were affinity purified with the cognate GST-tagged antigen covalently coupled to Glutathione Sepharose (GE Healthcare).
Anti-pICln (rabbit polyclonal)	Antibodies were generated by immunizing rabbits with untagged full-length pICln (Immunoglobe, Himmelstadt). Antibodies were affinity purified with GST-pICln covalently coupled to Glutathione Sepharose (GE Healthcare).
Y12 (mouse monoclonal)	A kind gift from Dr. J. Steitz, Yale University.
Anti-Gemin2 (mouse monoclonal)	Obtained from Transduction Laboratories (clone 2E17).
Anti-WD45 (rabbit polyclonal)	Antibodies were generated by immunizing rabbits with His_6- tagged WD45, the diluted serum was used for Western Blots.

Table S3: Protein Purification Buffers

Buffer A:	20 mM Bis-Tris (pH 6.0)
	5 mM β- Mercaptoethanol
Buffer B:	20 mM HEPES-NaOH (pH 7.5)
	5 mM DTT
Buffer C:	20 mM HEPES-NaOH (pH 8.0)
	5 mM β- Mercaptoethanol

Cloning of the Bicistronic Gemin7-Gemin6:pET21a Expression Plasmid

This dicistronic expression plasmid was constructed as follows. Gemin7:pET21a was linearized by digestion with *Not*I-*Xho*I, to this an annealed double-stranded polylinker (pl, see below) was ligated to introduce an unique *Nhe*I site, destroy the upstream *Not*I site, introduce a downstream *Not*I site, while retaining the *Xho*I site. The resulting Gemin7-pl:pET21a plasmid was digested with *Nhe*I-*Xho*I and purified. The Gemin6 ORF with a preceding ribosomal binding site (RBS) was obtained by digestion of Gemin6:pET28a with *Xba*I-*Xho*I and gel purification of the fragment. This fragment was then ligated with *Nhe*I-*Xho*I digested Gemin7-pl:pET21a to yield Gemin7-Gemin6:pET21a.

pl-*Not*I-upr: 5′- GGCCAGCTAGCGCGGCCGCC- 3′
pl-*Xho*I-lwr: 5′- TCGAGGCGGCCGCGCTAGCT- 3′

Cloning of the Tricistronic Gemin2-Gemin8:pET28b* Expression Plasmid

This tricistronic expression plasmid was generated by the linearization of Gemin2:pET28b* by digestion with *Not*I and *Xho*I, and the ligation of an annealed double-stranded polylinker, as described for the Gemin7-Gemin6:pET21a plasmid. The resulting Gemin2-pl:pET28b* plasmid was digested with *Nhe*I-*Xho*I. The Gemin8 ORF with a preceding ribosomal binding sequence (RBS) was obtained by the digestion of Gemin8:pET28b* with *Xba*I and *Xho*I and gel purification of the fragment. This fragment was then ligated with *Nhe*I-*Xho*I digested Gemin2-pl:pET28b* to obtain Gemin2-Gemin8:pET28b*. Another round of linker ligation as described above yielded Gemin2-Gemin8-pl:pET28b*. This plasmid was digested with *Nhe*I-*Xho*I. The His_6GSTTEV-SMN ORF with a preceding RBS was obtained by digestion of SMN:pETM30 with *Xba*I and *Xho*I and gel purification of the fragment. This fragment was then ligated with *Nhe*I-*Xho*I digested Gemin2-Gemin8-pl:pET28b* to yield Gemin2-Gemin8-SMN:pET28b*.

HeLa Extracts and Size Fractionation

Assembly-active HeLa cytosolic extracts were prepared essentially as described (Meister et al., 2001a). In brief, washed packed cells were suspended in an equal volume of 1×PBS, supplemented with 0.01% NP-40 and protease inhibitors. The suspension was homogenized by douncing with 30 strokes of a Type B pestle. The

cytoplasm was separated from nuclei by centrifugation at 1,000×g for 10min at 4°C. The supernatant (cytoplasm) was clarified by centrifugation at 25,000×g for 1h at 4°C, resulting in a cytosolic extract. For size fractionation, 5 ml of the resulting supernatant was directly applied to a Superdex200 26/60 column (GE Healthcare) equilibrated in 1×PBS + 0.01% NP-40. Proteins were eluted with 1×PBS + 0.01% NP-40 at a flow-rate of 2 ml/min, and 5ml fractions were collected.

Immuno-Affinity Purification of Protein Complexes

Immunoprecipitations were conducted essentially as described (Meister et al., 2001a; Meister et al., 2001b). For the immunoprecipitation of 20S and 6S complexes with anti-pICln antibodies (coupled to Protein G Sepharose, GE Healthcare, at 1mg/ml) fractions with an approximate molecular weight of 500kDa (20S) and 100kDa (6S) were supplemented with BSA to a final concentration of 1mg/ml. The determined molecular weight values were transformed into approximate sedimentation values in this entire study using the formula $S = 0.00242 \times \text{mol. weight}^{0.67}$ (Eason, 1984). For the immuno-affinity purification of the endogenous SMN-complex cytosolic extract (see above) was diluted two-fold with 1×PBS + 0.01% NP-40 and isolated with anti-SMN antibodies (coupled to Protein G Sepharose, GE Healthcare, at 5mg/ml).

Native Gel Electrophoresis of RNA-Protein Complexes

Band shift assays were performed essentially as described (Meister et al., 2001a). In brief, 3pmol proteins were added to 0.3pmol radio-labeled RNA in 10µl reactions, containing 10% glycerol, 0.1U/µl and 0.1µg/µl tRNA. The mixtures were incubated for 15min at 30°C and 45min at 37°C. After incubation the mixtures were briefly centrifuged, supplemented with heparin to 1µg/µl and separated on 6% native polyacrylamide gels (acrylamide/bisacrylamide ratio 80:1) containing 4% glycerol and 0.5×TBE buffer. Gels were pre-run for 1h at 4°C at a constant current of 15 mA in 1×TBE, and run with samples for 2h at 4°C at a constant current of 35mA. Gels were exposed wet at -80°C.

Expression and Purification of Streptavidin

Streptavidin was expressed in *E.coli* and inclusion bodies were isolated. Inclusion bodies were solubilized, proteins refolded and purified as described previously (Chilkoti et al., 1995).

Expression and Purification of the Sm Proteins D1/D2, D3/B and EFG

D1/D2, D3/B, the Sm-fold domain D3-Sm/B-Sm heterodimers, and the E/F/G heterotrimer were expressed in *E. coli* and purified as described (Kambach et al., 1999).

Expression and Purification of His_6-pICln

pICln:pET28a was transformed into *E. coli* BL21(DE3) cells (Novagen). Transformed cells were grown at 37°C in Superbroth containing 25µg/ml Kanamycin to an OD_{600} of 0.4. At this stage cultures were cooled to 25°C and induced with 1mM IPTG for an additional 5h. Cells were harvested by centrifugation and resuspended in Buffer A(0; number in parentheses denotes the molar concentration of NaCl) with protease inhibitors (Complete-EDTA free, Roche). Resuspended cells were broken by sonication and a cleared lysate prepared by ultracentrifugation in a 45Ti rotor (Beckman), 40,000 rpm, 1h, 4°C. The cleared lysate was applied to a 5ml HiTrapQ column (GE Healthcare). After washing with Buffer A(0) proteins were eluted with a linear gradient to Buffer A(1). pICln containing fractions were identified by SDS-PAGE, pooled, and applied without dialysis to a 5ml HiTrapChelating column (GE Healthcare). The column was washed with Buffer A(0.5) supplemented with 0.01M imidazole, and eluted with Buffer A(0.5) supplemented with 0.3M imidazole. pICln fractions were pooled, additional purification and buffer exchange was performed by gel filtration with a Superdex200 column (GE Healthcare) run in Buffer B(0.2). Purified protein was concentrated by ultrafiltration (Vivascience) to 20mg/ml and stored at -80°C until further use.

Expression and Purification of Untagged pICln

M15(pREP4) cells (Qiagen) were transformed with pICln:pQE30-2Z. Transformed cells were grown in Superbroth containing 100µg/ml Ampicillin and 25µg/ml Kanamycin at 37°C to an OD_{600} of 1. The temperature was reduced to 25°C and cultures induced with 1mM IPTG for 3h. Cultures were harvested, resuspended,

broken, and a cleared lysate prepared as described for His_6-pICln above. The cleared lysate was applied to a 5ml HiTrapQ column (GE Healthcare), the column was washed and proteins eluted by a linear gradient as described above. pICln containing fractions were identified by SDS-PAGE, pooled, diluted with Buffer A(0) to a conductivity of Buffer A(0.1) and resubjected to HiTrapQ chromatography. Proteins were eluted as above and pICln fractions were pooled and combined with 1:50 ratio (protease: protein) of TEV protease and cleavage allowed to proceed for 48h at room temperature. The sample was simultaneously dialyzed against Buffer A(0.5). The cleaved sample was then applied to a 5ml HiTrapChelating column (GE Healthcare) and the flow-through collected. Additional purification and buffer exchange of cleaved pICln was achieved by gel filtration with a Superdex200 column (GE Healthcare) run in Buffer B(0.2). Purified protein was concentrated by ultrafiltration (Vivascience) to 20mg/ml and stored at -80°C until further use.

Expression and Purification of the Gemin2/SMNΔC Heterodimer

BL21(DE3) (Novagen) cells were cotransformed with both Gemin2:pET28a and SMN(1-160):pGEX-6P-1. Transformed cells were cultured in Superbroth containing 100µg/ml Ampicillin and 25µg/ml Kanamycin at 37°C to an OD_{600} of 0.8 and cultures cooled to 15°C. Protein expression was induced with 1mM IPTG for 20h at 15°C. Cells were harvested by centrifugation and resuspended in Buffer C(0.5) supplemented with 0.01M imidazole and broken by sonication. A cleared lysate was prepared by ultracentrifugation as described for pICln and applied to a NiNTA column (Qiagen). The column was washed extensively with Buffer C(0.5) containing 0.01M imidazole and eluted with Buffer C(0.5) supplemented with 0.3M imidazole. Gemin2/GST-SMNΔC containing fractions were identified by SDS-PAGE and supplemented with a 1:100 ratio (protease:protein) PreScission protease (GE Healthcare), and cleavage allowed to proceed over night at 4°C. The cleavage mixture was then passed through a Glutathion Sepharose column (GE Healthcare), the flow-through collected and dialyzed to Buffer C(0.1) and applied to a 5ml HiTrapQ column (GE Healthcare). After washing with Buffer C(0.1), bound protein was eluted with a linear gradient to Buffer C(1). The Gemin2/SMNΔC fractions were identified by SDS-PAGE and pooled. The final purification and buffer exchange step was Superdex200 gel filtration chromatography (GE Healthcare) in Buffer B(0.2). Purified

Gemin2/SMNΔC heterodimer was concentrated by ultrafiltration (Vivascience) to 15mg/ml and stored at -80°C until further use.

Expression and Purification of the SMNΔGemin3-5 Complex

For expression and purification of the SMNΔGemin3-5 complex *E. coli* Rosetta2(DE3) cells (Novagen) were cotransformed with both Gemin7-Gemin6:pET21a and Gemin2-Gemin8-GSTHis$_6$TEV-SMN:pET28b* plasmids. Transformed cells were cultured in Superbroth containing 100µg/ml Ampicillin, 25µg/ml Kanamycin, and 34µg/ml Chloramphenicol. Cultures were incubated at 37°C to an OD_{600} of 0.4, cooled to 15°C and induced with 0.5mM IPTG for an additional 20h. Cells were harvested by centrifugation, resuspended in Buffer C(0.5) containing protease inhibitors, and broken by sonication. A cleared lysate was prepared by ultracentrifugation in a 45Ti rotor (Beckman), 25,000 rpm, 1h at 4°C. The cleared lysate was incubated with Glutathione Sepharose (GE Healthcare) in batch for 2h at 4°C. After extensive washing with Buffer C(0.5), bound proteins were eluted with Buffer C(0.5) containing 0.02M Glutathione. The eluate was pooled, supplemented with a 1:50 ratio (protease:protein) of TEV protease and digested for 48h with simultaneous dialysis to Buffer C(0.1). The cleaved SMNΔGemin3-5 complex was liberated from the cleaved His$_6$-GST tag and the protease by anion-exchange chromatography on a 5ml HiTrapQ column (GE Healthcare) equilibrated in Buffer C(0.1). Proteins were eluted by a linear gradient to Buffer C(1). SMNΔGemin3-5 complex containing fractions were identified by SDS-PAGE and pooled. Finally a purification step and buffer exchange was performed by Superdex200 gel filtration (GE Healthcare) in Buffer B(0.2) without reducing agent. SMNΔGemin3-5 complex fractions were pooled, concentrated to 60µM (7.5mg/ml) and stored at 4°C until further use.

Expression and Purification of the SMNΔGemin3-7 Complex

Expression and purification of the SMNΔGemin3-7 complex was essentially as detailed for the SMNΔGemin3-5 complex above. The exception being that Rosetta2(DE3) cells (Novagen) were transformed with the Gemin2-Gemin8-GSTHis$_6$TEV-SMN:pET28b* expression plasmid only and that cells were cultured in Superbroth containing 25µg/ml Kanamycin and 34µg/ml Chloramphenicol.

Expression and Purification of *Drosophila Melanogaster* pICln (dpICln)

Rosetta2(DE3) cells (Novagen) were transformed with dpICln:pET28a. Transformed cells were cultured in Superbroth containing 25µg/ml Kanamycin and 34µg/ml Chloramphenicol to an OD_{600} of 0.8. Cultures were cooled to 15°C, induced with 1mM IPTG, and further incubated for 4h. Cells were harvested, broken and a cleared lysate prepared as described for human pICln. The cleared lysate was applied to HiPrepQ column (GE Healthcare) equilibrated in Buffer A(0). The column was washed extensively with Buffer A(0) and bound proteins eluted with a linear gradient to Buffer A(1). dpICln containing fractions were identified by SDS-PAGE and dialyzed over night to Buffer A(0.1) at 4°C. The dialysate was subjected again to HiPrepQ chromatography and bound proteins eluted with a linear gradient to Buffer A(1). dpICln fractions were pooled and concentrated by ultrafiltration (Vivascience) and gel filtered on Superdex200 (GE Healthcare) in Buffer B(0.2). dpICln elutes from the column as soluble aggregates in the void volume of the column and in fractions corresponding to monomeric protein (data not shown). The monomeric protein fractions were pooled and subjected to anion-exchange chromatography using a 5ml HiTrap Q column (GE Healthcare). dpICln eluted as a single peak in a linear gradient to Buffer B(1). dpICln fractions were pooled, dialyzed to Buffer B(0.2), concentrated to 15mg/ml and stored at -80°C until further use. Rechromatography of purified dpICln on Superdex200 revealed that the monomeric protein fractions do not form soluble aggregates after purification (data not shown).

Expression and Purification of the *Drosophila Melanogaster* Gemin2/SMNΔC Heterodimer

Rosetta2(DE3) cells (Novagen) were singly transformed with dGemin2:pETM30 and dSMN(1-122):pETM30 expression plasmids respectively. Both cells transformed with dGemin2:pETM30 and dSMN(1-122):pETM30 were cultured in Superbroth containing 25µg/ml Kanamycin and 34µg/ml Chloramphenicol to an OD_{600} of 1. Both cultures were cooled to 15°C, induced with 1mM IPTG, and incubated for an additional 20h. Both cultures were harvested and resuspended separately. Equal cell culture volumes of resuspended cells of both expression cultures were then combined and broken by sonication to enable complex formation. Cleared lysates were prepared by ultracentrifugation in a 45Ti rotor (Beckman), 40,000 rpm, 1h hat 4°C. The cleared

lysate was then applied to a 5ml HiTrapChelating column (GE Healthcare) equilibrated in Buffer C(0.5) supplemented with 0.01M imidazole. After extensive washing with the same buffer, bound proteins were eluted with Buffer C(0.5) containing 0.5M imidazole. The His_6GSTTEV-Gemin2/His_6GSTTEV-SMNΔC heterodimer containing fractions were pooled, supplemented with an 1:50 ratio (protease: protein) of TEV protease and cleaved over night at 4°C, with simultaneous dialysis to Buffer C(0.1). The dialysate was passed through a 5ml HiTrapChelating column and the flow-through was directly subjected to anion- exchange chromatography on a 5ml HiTrapQ column (GE Healthcare). The cleaved *Drosophila* Gemin2/SMNΔC heterodimer was eluted by a gradient to Buffer C(1), and appropriate fractions pooled. The *Drosophila* Gemin2/SMNΔC heterodimer was additionally purified and the buffer exchanged to Buffer B(0.2) by Superdex200 gel filtration chromatography (GE Healthcare). Pure heterodimer fractions were pooled, concentrated by ultrafiltration (Vivascience) to 20mg/ml, and stored at -80°C until further use.

Supplemental References

Battle, D.J., Kasim, M., Wang, J., and Dreyfuss, G. (2007). SMN-independent subunits of the SMN complex. Identification of a small nuclear ribonucleoprotein assembly intermediate. J Biol Chem *282*, 27953-27959.

Battle, D.J., Lau, C.K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006). The Gemin5 protein of the SMN complex identifies snRNAs. Mol Cell *23*, 273-279.

Chilkoti, A., Tan, P.H., and Stayton, P.S. (1995). Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. Proc Natl Acad Sci U S A *92*, 1754-1758.

Eason, R. (1984). Analytical ultracentrifugation. In Centrifugation: a practical approach, D. Riekwood, ed. (Oxford and Washington, IRL Press Ltd.), pp. 1-286.

Fischer, U., and Lührmann, R. (1990). An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. Science *249*, 786-790.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Lührmann, R. (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.

Grimmler, M., Otter, S., Peter, C., Müller, F., Chari, A., and Fischer, U. (2005). Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. Hum Mol Genet *14*, 3099-3111.

Hamm, J., Darzynkiewicz, E., Tahara, S.M., and Mattaj, I.W. (1990). The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. Cell *62*, 569-577.

Hamm, J., Kazmaier, M., and Mattaj, I.W. (1987). *In vitro* assembly of U1 snRNPs. EMBO J *6*, 3479-3485.

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

Kroiss, M., Schultz, J., Wiesner, J., Chari, A., Sickmann, A., and Fischer, U. (2008). Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*. PNAS *105*, 10045–10050.

Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W., and Dreyfuss, G. (2002). The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. Mol Cell Biol *22*, 6533-6541.

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

Meister, G., Bühler, D., Laggerbauer, B., Zobawa, M., Lottspeich, F., and Fischer, U. (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.

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

Meister, G., Eggert, C., Bühler, D., Brahms, H., Kambach, C., and Fischer, U. (2001b). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr Biol *11*, 1990-1994.

Narayanan, U., Achsel, T., Lührmann, R., and Matera, A.G. (2004). Coupled *in vitro* import of U snRNPs and SMN, the spinal muscular atrophy protein. Mol Cell *16*, 223-234.

Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol *13*, 290-301.

Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F.K., and Kambach, C. (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. J Biol Chem *280*, 16066-16075.

Supplemental Figures

Chari et al., Figure S1

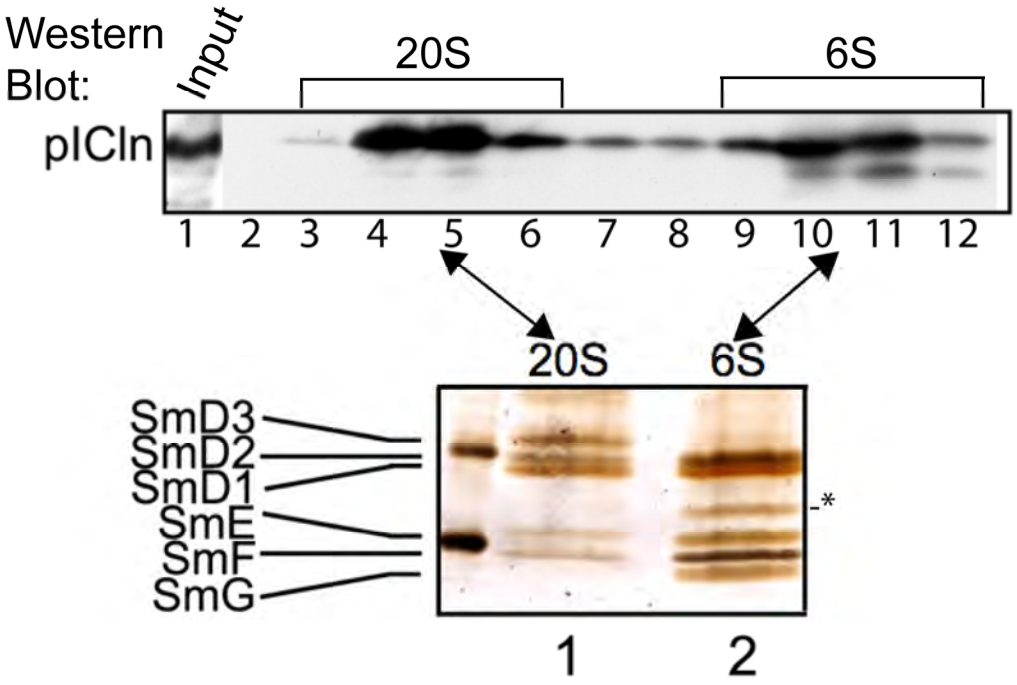


Figure S1: EFG Levels in the 20S Complex Vary among Different Purifications

HeLa cytosolic extract active in UsnRNP assembly was size fractionated, individual fractions resolved by SDS-PAGE and 20S (lanes 3-6) and 6S (lanes 9-12) peak fractions revealed by Western blotting with anti-pICln antibodies (upper panel). 20S and 6S peak fractions were pooled and separately subjected to immuno-affinity purification with anti-pICln antibodies (bottom panel, lanes 1&2). The isolates were subjected to SDS-PAGE and subsequent silver-staining. The associated Sm proteins were identified by Western blotting and mass spectrometry. The asterisk denotes a degradation product of pICln.

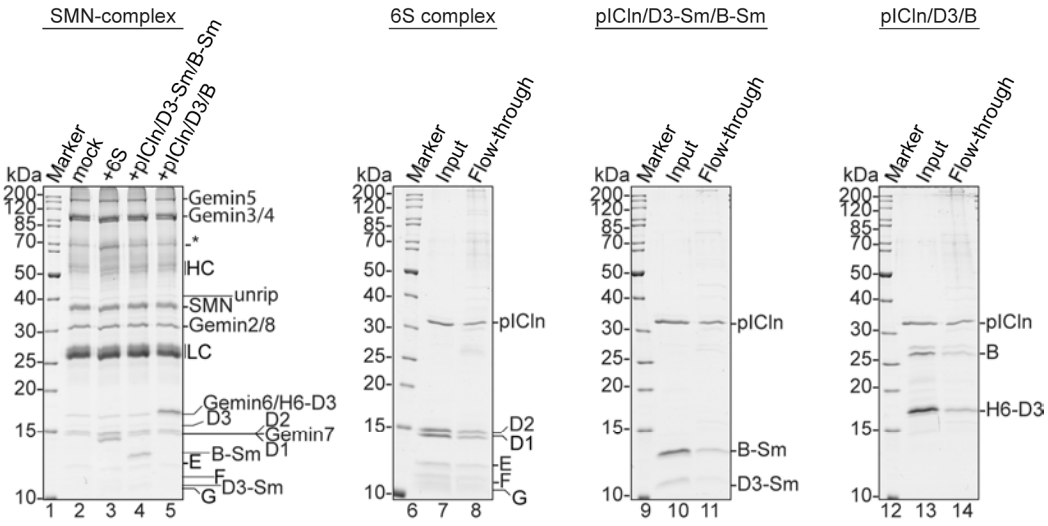


Figure S2: Efficient Transfer of Sm Proteins from the Respective pICln Complexes onto the SMN-Complex

Immobilized, affinity-purified SMN-complex was mock-treated (lane 2), or incubated with either the 6S complex (lane 3), the pICln/D3-Sm/B-Sm complex (lane 4) or pICln/D3/B (lane 5) The retained proteins were eluted by boiling in SDS-PAGE buffer and analyzed by SDS-PAGE and Coomassie staining (lanes 1-5). Also depicted are the input and flow-through material of the 6S complex (lanes 7&8), the pICln/D3-Sm/B-Sm complex and pICln/D3/B complexes (lanes 13&14) used in the binding experiment. Densitometric analyses and subsequent comparison of the ratios of Sm proteins with respect to pICln before and after the transfer experiment reveal a reduction of approx. 30-40% in Sm protein levels (data not shown). As the respective pICln complexes were offered in a two-fold molar excess over the recipient SMN-complex, Sm protein transfer occurred with an efficiency of at least 60% under the conditions employed here.

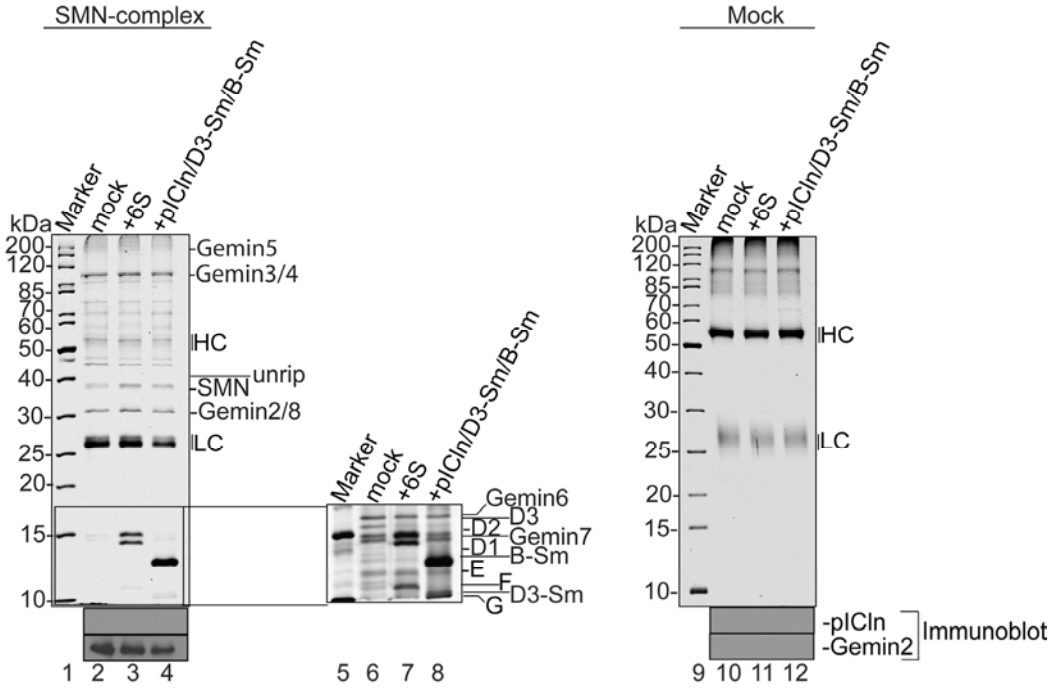


Figure S3: The C-terminal Tail Domains of the Sm Proteins D3 and B Are Dispensable for Their Transfer onto the SMN-Complex *in Vitro*

Immobilized, affinity-purified SMN-complex was either mock-treated (lanes 2&6), treated with the 6S complex (lanes 3&7), or the pICln/D3-Sm/B-Sm complex (lanes 4&8). The retained proteins were eluted by boiling in SDS-PAGE buffer and analyzed by SDS-PAGE and Coomassie Blue staining (lanes 1-4). The inset (lanes 5-8) shows a silver-stained portion of the same gel for a better appreciation of low-molecular weight proteins that stain poorly with Coomassie. Lanes 9-12 show the same mixtures treated with a control-column and processed as described for the SMN-complex samples. 50% of each eluate was analyzed by Western blotting for retained pICln (pICln immunoblot panel). Gemin2 served as a loading control on Western blots (Gemin2 immunoblot panel). HC and LC indicate heavy and light chains of the antibody used for affinity-purification.

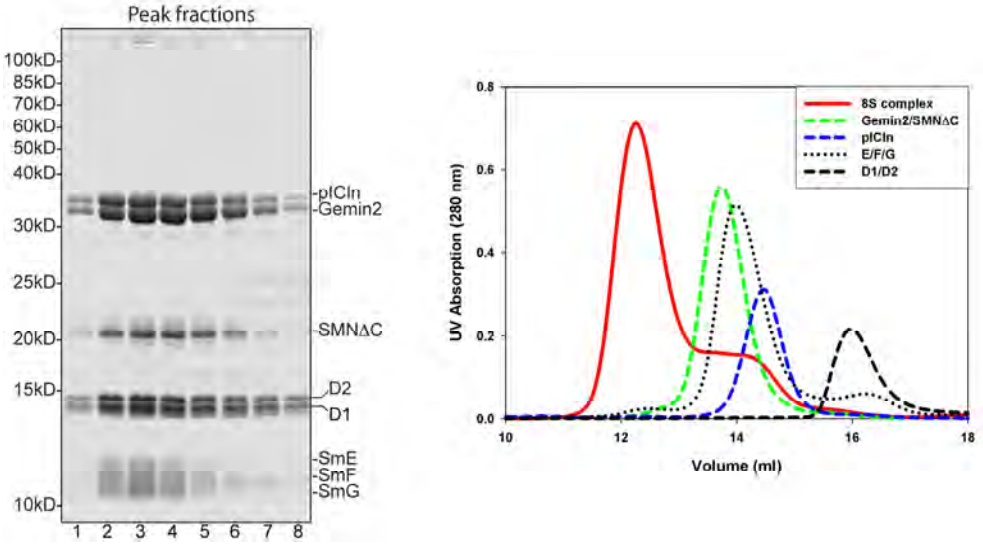


Figure S4: Reconstitution of the 8S Complex (Gemin2/SMNΔC/pICln/D1/D2/ E/F/G)

The 8S complex was reconstituted as described in the Supplemental Experimental Procedures section. Individual gel filtration fractions were resolved by SDS-PAGE and visualized by staining with Coomassie (left panel). Corresponding elution profiles of the reconstituted 8S complex and elution profiles of the individual building blocks are shown (right panel).

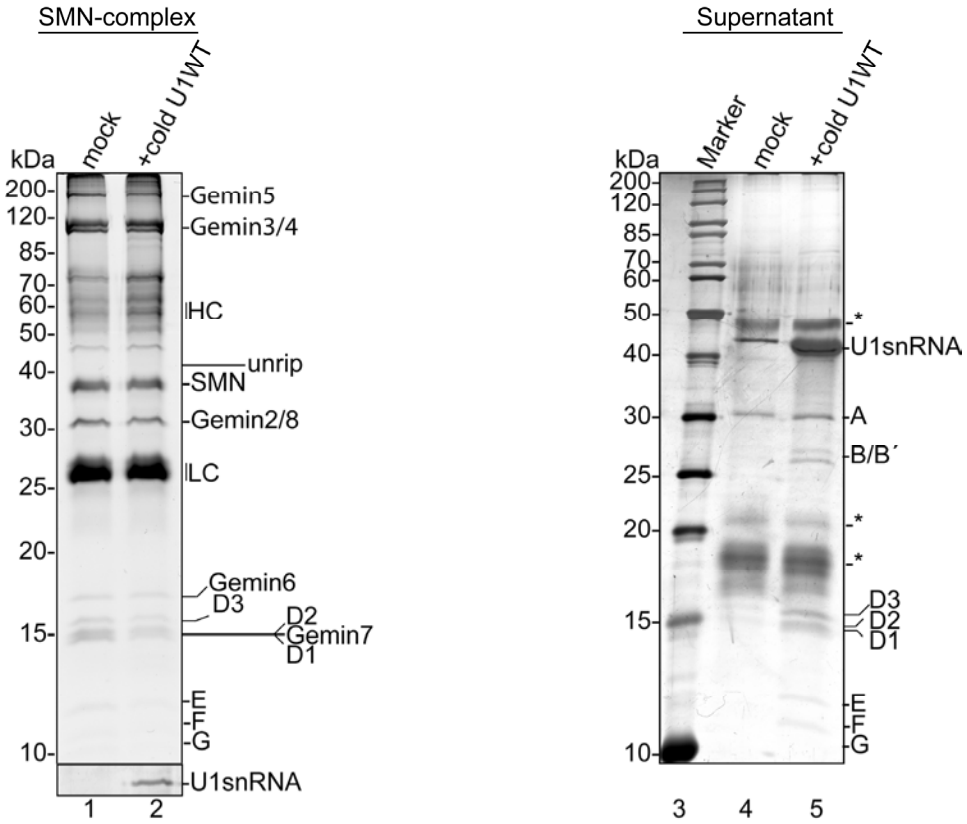


Figure S5: Incubation of the SMN-Complex with a Large Excess of U1snRNA Exhausts its Assembly Activity

Immobilized, affinity-purified SMN-complex was either mock-treated (lane 1) or mixed with 100µg non-labeled U1 snRNA (lane 2) and incubated under assembly conditions for 1 hour. Hereafter, the immobilized SMN-complexes were separated from their corresponding supernatants and analyzed by SDS-PAGE. The SMN-complex treated with U1 snRNA shows a reduction of Sm proteins of approx. 50% when compared to the control complex, whereas SMN and Gemins remained constant (compare lanes 1 and 2). Furthermore, a significant fraction of U1 snRNA remained associated with the SMN-complex under these conditions (compare lanes 1 and 2, lower panel). Assembled U1 snRNP could be identified in the supernatant of the assembly reaction (lane 5) but not in the mock control (lane 4).

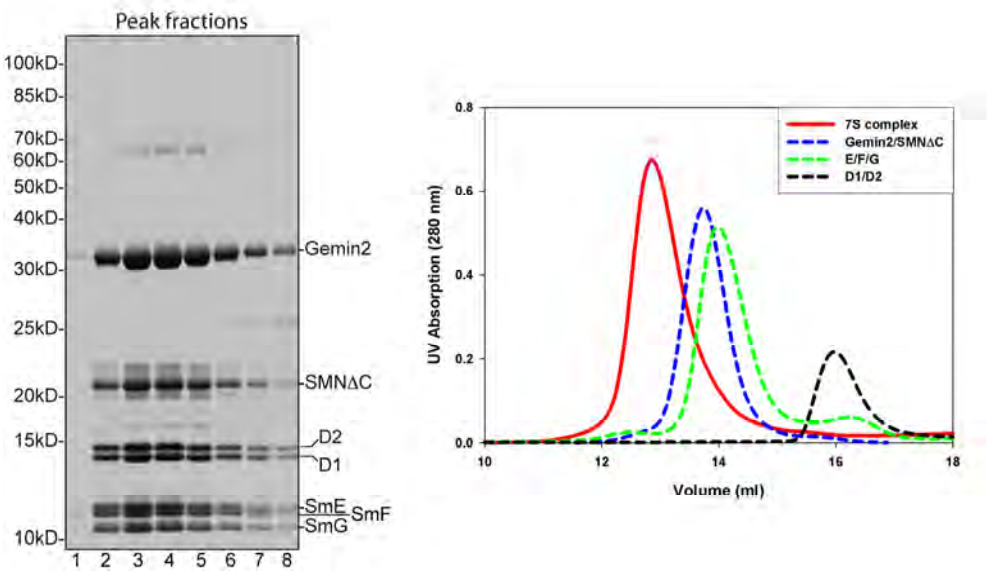


Figure S6: Reconstitution of the 7S Complex (Gemin2/SMNΔC/pICln/D1/D2/ E/F/G)

The 7S complex was reconstituted as described in the Supplemental Experimental Procedures section. Individual gel filtration fractions were resolved by SDS-PAGE and visualized by staining with Coomassie (left panel). Corresponding elution profiles of the reconstituted 8S complex and elution profiles of the individual building blocks are shown (right panel).

A

	6S views	8S views	7S views	Slim views	Globular views	Truncated views	Low SNR
6S [n = 1050]	0.907	0.000	0.000	0.025	0.031	0.029	0.009
8S (6S+SMN ΔCGem2) [n = 278]	0.014	0.878	0.011	0.000	0.032	0.025	0.040
8S (7S+pICln) [n = 779]	0.019	0.937	0.004	0.000	0.018	0.018	0.004
7S [n = 775]	0.028	0.010	0.866	0.021	0.019	0.039	0.017

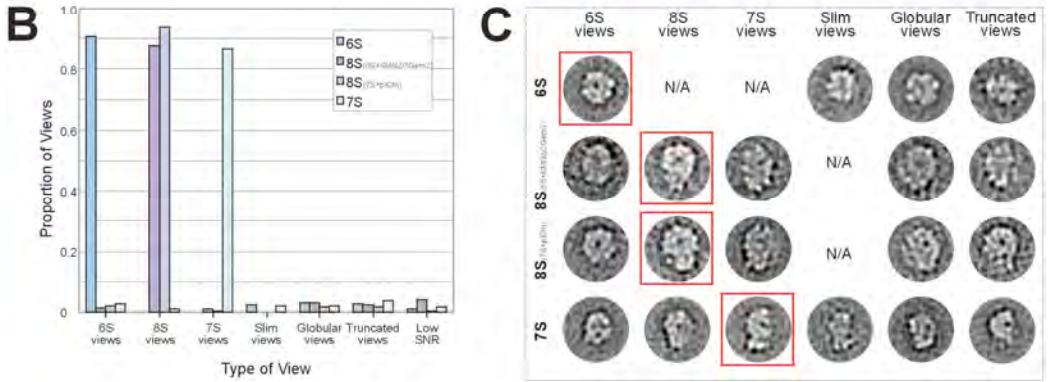


Figure S7: Distribution of Particle Views as Studied by Single-Particle Electron Microscopy.

For a comprehensive analysis, all particles on the raw electron microscopic images were included without any pre-selection and subjected to iterative rounds of multi-reference alignment and multivariate statistical analysis. All resulting class averages were subsequently assigned to one of the following types (compare also Figure S7C): (1) 6S views showing a ring-like architecture with a central accumulation of stain; (2) 8S views revealing a ring-shaped domain connected an additional foot-like protuberance; (3) 7S views comprising an open clamp domain packed against a foot-like protuberance. Furthermore, four rare additional types of views that deviate from the main views were discerned: (4) slim, (5) globular, (6), truncated, and (7) views of low signal-to-noise ratio (SNR). Note that the latter four types were defined on the basis of the shapes of the main particle population: e.g. the truncated view of the 7S data set (Figure S7C) shows the clamp domain, but lacks the foot-like protuberance, while the truncated 6S view is composed of a smaller or rotated ring-domain.

(A) The majority of particle views (87- 94%) adopt typical shapes as shown in Figure 5 of the main manuscript while all other view types were only rarely found and typically were non-redundant.

(B) Histogram depicting the distribution of particle views as summarized in Figure S7A.

(C) Typical example class averages found in the four data sets. The largest subpopulation found in each data set is highlighted by a red box. N/A, not available: no particle view adopted this shape.

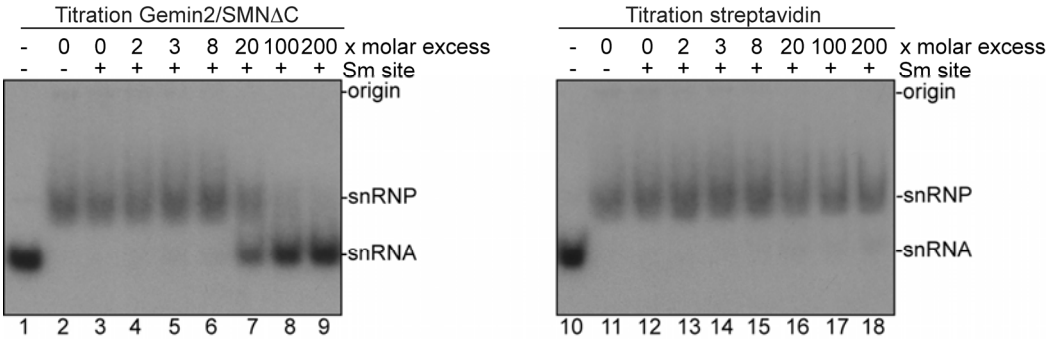


Figure S8: The Gemin2/SMNΔC Heterodimer is Sufficient to Reduce the Activation Energy Barrier for Sm Protein Strand-Exchange

Based on the architecture of the 7S complex and experiments in Figure 4D (lane 5), we predicted that the Gemin2/SMNΔC heterodimer would also be active in Sm protein strand exchange assays. However, as the binding site for D3/B is missing in this unit (compare Fig. 3B, lane 20), it should be less efficient in comparison to the SMNΔGemin3-5 complex. To address this, [^{32}P]-labeled U1snRNA (lane 1) was incubated with Sm proteins to allow formation of the Sm core domain (lane 2). To the pre-formed core, a 330-fold molar excess of Sm-site RNA and increasing amounts of the Gemin2/SMNΔC heterodimer were added (lanes 3-9). After 2.5h, the reactions were resolved on native gels. In lanes 10-18, the same reactions were performed in the presence of increasing amounts of streptavidin instead of the Gemin2/SMNΔC heterodimer. According to our hypothesis that this unit would be less active in this experiment due to the absence of the binding site for D3/B, we find a 20-fold molar excess of the dimer to be necessary for half-maximal exchange of Sm proteins from the pre-formed snRNP to the unlabeled RNA (lane 7), and a 100-fold molar excess required for total exchange (lane 8).

Zellulärer Maschinenbau

In der Zelle sind Maschinenbauer am Werk

In unseren Zellen läuft eine Vielzahl von Prozessen ab, die wie in einer großen Fabrik von vielen verschiedenen Maschinen verrichtet wird. Da diese Maschinen in der Regel nur wenige Millionstel Millimeter im Durchmesser groß sind, nennt man sie auch „Molekulare Maschinen". Ähnlich der von Menschenhand gebauten Maschinen sind sie oft sehr komplex aufgebaut, statt aus Metall oder Plastik bestehen sie aus Proteinen oder Nukleinsäuren (DNA und RNA). Deshalb war es auch kaum zu glauben, dass sich diese hoch komplexen Maschinen spontan zusammenlagern sollten – wie dies einige Forscher behaupteten. „Diese Annahme ist jetzt eindeutig widerlegt. Unsere Untersuchungen haben gezeigt, dass die Zelle ähnlich vorgeht wie wir es von der Konstruktion einer Maschine in einer Fabrik kennen. Richtige Maschinenbauer sind da am Werk", so Utz Fischer.

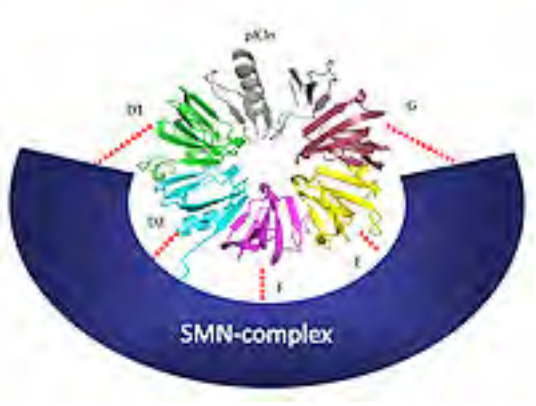
Proteine arbeiten nach einem vorgegebenen Bauplan

Fischer und sein Forscherteam untersuchten hierzu den Aufbau von RNA-Proteinkomplexen des Spleißosoms. Hierbei handelt es sich um eine enorm komplexe molekulare Maschine; ihre Aufgabe ist es, das Umschreiben der Erbinformation in Proteine zu kontrollieren. Fischers Mitarbeiter fanden eine Gruppe von Proteinen, die nicht beim Umschreiben, also dem Prozess selbst, sehr wohl aber in der Aufbauphase eine Rolle spielten. „Es lag nahe zu vermuten, dass diese Proteine als Zusammenlagerungsgehilfen agieren", so Fischer. Die Art und Weise, wie sie das machen, war überraschend: „Der Aufbau gleicht einem komplizierten Puzzle, dessen Lösung nur durch das Zusammenarbeiten von unterschiedlichen Proteinen möglich wird, die als Zulieferer und Monteure wirken", erklärt Fischer.

Zunächst werden Gruppen von Proteinen vorgefertigt, die auch im vollendeten RNA-Proteinkomplex dicht beieinander platziert werden müssen. Diesen Job verrichtet ein Helfer, der als Zulieferer agiert und die korrekte Bildung der Proteine sowohl zeitlich als auch räumlich überwacht. In der Fachsprache werden derartige Helfer „Chaperone", also Anstandsdamen, genannt. Die vorgefertigten Untereinheiten werden dann an den eigentlichen Monteur, den sogenannten SMN-Komplex, übergeben. Dieser fügt die Proteine mit der RNA zu einem funktionierenden Ganzen zusammen. „Der Aufbau dieser Partikel erfolgt also nach einem vorgegebenen Bauplan, wobei sowohl Zulieferer als auch Monteure das Voranschreiten ständig einer Qualitätskontrolle unterziehen", so Ashwin Chari, der die Studie experimentell leitete und durchführte.

Fehlerhafte Maschinen führen zu schweren Krankheiten

Funktioniert dieses System des Zusammenbauens nicht mehr, können Krankheiten die Folge sein. Ein Beispiel ist die Spinale Muskelatrophie (SMA). In den Zellen der Betroffenen schaffen es die Helfer nicht, Maschinen in ausreichender Zahl zu produzieren. Deshalb fehlt es an den wichtigen RNA-Proteinkomplexen. Bei den Erkrankten sterben die Nervenzellen im Rückenmark; schwerste Lähmungen und Tod sind die Folge.



Proteinbau im Schema (Grafik Ashwin Chari)

Ashwin Chari, Monika M. Golas, Michael Klingenhäger, Nils Neuenkirchen, Bjoern Sander, Clemens Englbrecht, Albert Sickmann, Holger Stark, and Utz Fischer (2008), An Assembly Chaperone Collaborates with the SMN Complex to Generate Spliceosomal SnRNP. Cell, DOI 10.1016/j.cell.2008.09.020

Kontakt: Tel.: +49 931 888-4029, E-Mail: utz.fischer@biozentrum.uni-wuerzburg.de

Links: Den gesamten Prozess haben die Wissenschaftler schematisch in einem Trickfilm zusammengefasst. Den Film gibt es hier zu sehen. Zum Abspielen wird der VLC Media Player benötigt.

By: Sonja Jülich / Gunnar Bartsch

06.11.2008, 08:32 Uhr

5.4 Evolution of the SMN-complex

Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster

Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U: **PNAS 2008**, 105:10045–10050.

Thesis author´s contribution:

Conception:	30 %
Experimental contribution:	10 %
Formulation of results:	10 %

Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster*

Matthias Kroiss*, Jörg Schultz†, Julia Wiesner‡, Ashwin Chari*, Albert Sickmann‡, and Utz Fischer*§

Departments of *Biochemistry and †Bioinformatics, Theodor Boveri Institute, Biocenter, Am Hubland, 97074 Würzburg, Germany; and ‡Functional Genomics Group, Rudolf Virchow Center, Versbacher Strasse 9, 97078 Würzburg, Germany

Edited by Kay E. Davies, University of Oxford, Oxford, United Kingdom, and accepted by the Editorial Board April 25, 2008 (received for review March 8, 2008)

In vertebrates, assembly of spliceosomal uridine-rich small nuclear ribonucleoproteins (UsnRNPs) is mediated by the SMN complex, a macromolecular entity composed of the proteins SMN and Gemins 2–8. Here we have studied the evolution of this machinery using complete genome assemblies of multiple model organisms. The SMN complex has gained complexity in evolution by a blockwise addition of Gemins onto an ancestral core complex composed of SMN and Gemin2. In contrast to this overall evolutionary trend to more complexity in metazoans, orthologs of most Gemins are missing in dipterans. In accordance with these bioinformatic data a previously undescribed biochemical purification strategy elucidated that the dipteran *Drosophila melanogaster* contains an SMN complex of remarkable simplicity. Surprisingly, this minimal complex not only mediates the assembly reaction in a manner very similar to its vertebrate counterpart, but also prevents misassembly onto nontarget RNAs. Our data suggest that only a minority of Gemins are required for the assembly reaction *per se*, whereas others may serve additional functions in the context of UsnRNP biogenesis. The evolution of the SMN complex is an interesting example of how the simplification of a biochemical process contributes to genome compaction.

splicing | spinal muscular atrophy | UsnRNA

Splicing of pre-mRNAs is catalyzed by the spliceosome, a macromolecular machine consisting of a large number of protein factors and the uridine-rich small nuclear ribonucleoproteins (snRNPs) U1, U2, U4/6, and U5. The biogenesis of these particles occurs in a stepwise manner. First, nuclear-transcribed, m^7G-capped snRNAs U1, U2, U4, and U5 are exported into the cytoplasm, where a conserved sequence motif in these RNAs (Sm site) serves as a binding platform for the seven Sm proteins B/B′, D1, D2, D3, E, F, and G. As a consequence, a ring-shaped Sm core domain is formed. This domain is crucial for subsequent steps in the biogenesis of UsnRNPs, such as formation of the hypermethylated $m^{2,2,7}$G cap and import of the assembled particle into the nucleus. At a yet to be defined step, additional factors are recruited to form the mature UsnRNP particles that function in splicing (1).

Previous studies have shown that Sm proteins bind spontaneously, albeit in a hierarchical manner, onto UsnRNAs *in vitro* (2, 3). However, in cellular extracts, this process depends on ATP and the activity of the multisubunit SMN complex (4–6). Recently, a systematic interaction study on the human SMN complex has established its basic architecture (7). A modular composition was deduced where the three factors SMN, Gemin2, and Gemin8 form the backbone of the entire complex. Onto this core, the peripheral building blocks Gemin3/4 and Gemin6/7/UNRIP bind to form the functional unit. In support of this modular architecture, Gemin-containing subcomplexes have been identified composed of SMN/Gemin2, Gemin3–Gemin5, and Gemin6/7/UNRIP (8).

The SMN complex not only functions in the assembly of the Sm core domain, but also influences additional steps in the biogenesis pathway of UsnRNPs. One such step is the nuclear import of the assembled UsnRNP, which is mediated by the SMN complex (or parts thereof) in conjunction with the import factor importin β (9). In addition, specific UsnRNP proteins and the cap hypermethylase Tgs1 have been found in association of the SMN complex (10). This observation indicates that the SMN complex coordinates various events during UsnRNP biogenesis by assuming the role of a binding platform for the respective assisting factors.

The multisubunit composition of the human SMN complex has impeded the mechanistic dissection of the UsnRNP assembly process. Thus, although RNA interference studies indicated essential roles of several Gemins in the assembly reaction (11–13), their precise contributions remain unclear. To facilitate mechanistic studies and to gain insight into the evolution of the SMN complex, we have mined genomic databases for organisms that lack individual Gemins and hence may contain a simpler assembly machinery. Indeed, this was the case for different organisms including dipterans. We chose to further investigate *Drosophila* because of the wealth of genetic resources. An affinity chromatography strategy has permitted us the purification of an assembly-active complex composed of SMN and Gemin2 only. Remarkably, this complex not only facilitated assembly of the Sm core domain but also discriminated between cognate and noncognate RNAs. Thus, our combined bioinformatic and biochemical approach revealed that the assembly reaction requires only two core proteins *in vitro*, even though SMN complexes from most metazoans are of considerable complexity. We speculate that Gemins 3–8 have been recruited to the SMN complex in the course of evolution to integrate assembly with additional steps in the biogenesis of UsnRNPs.

Results

An Elaborate SMN Complex Is Characteristic of Metazoans. To understand how the UsnRNP assembly machinery has evolved, we performed homology searches for all Gemin proteins constituting the human SMN complex in genomic databases of a variety of organisms (see *Experimental Procedures* for details). Because of its diverse functions (14, 15) and its transient cytoplasmic interaction

Author contributions: M.K., J.S., A.C., and U.F. designed research; M.K., J.S., J.W., and A.C. performed research; M.K. contributed new reagents/analytic tools; M.K., J.S., J.W., A.S., and U.F. analyzed data; and M.K., J.S., A.C., and U.F. wrote the paper.

Conflict of interest statement: The authors have filed a patent application for the affinity purification of complexes (European Patent Office application no. 07108779.5).

This article is a PNAS Direct Submission. K.E.D. is a guest editor invited by the Editorial Board.

§To whom correspondence should be addressed. E-mail: utz.fischer@biozentrum.uni-wuerzburg.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802287105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

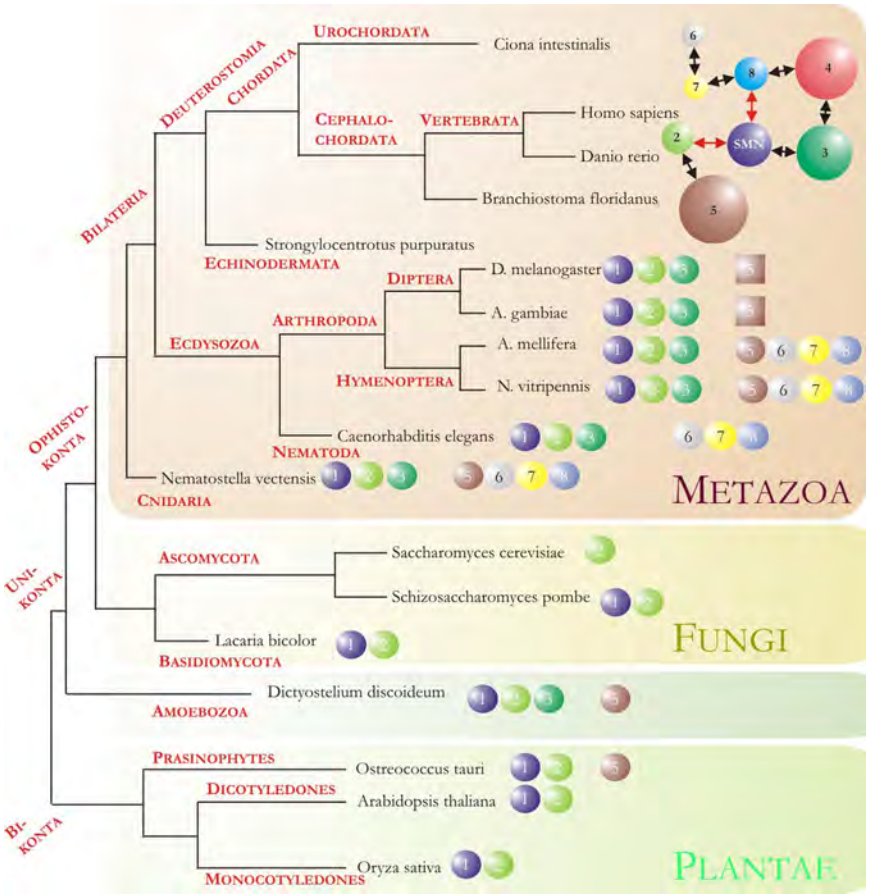


Fig. 1. Evolution of the SMN complex. Complete genome assemblies of indicated organisms have been screened for orthologs of human SMN complex components. The presence of the proteins was mapped on a published phylogenetic tree (38). A homolog of Gemin6 was found in the algae *O. tauri*, but the homology was restricted to the C-terminus and therefore is not shown. Gemin5 orthologs in dipterans are evolving significantly faster than in other organisms. Because this may indicate a change of function, they are shown as squares.

with the SMN complex, the UNRIP protein has been excluded from this analysis.

SMN and Gemin2 orthologs (termed Yab8p and Yip1p) but no other Gemins can be found in the fungus *Schizosaccharomyces pombe* [see Fig. 1; for protein identifiers see supporting information (SI) Table S1]. Importantly, both orthologs interact physically and may hence form a functional unit (16). In *Saccharomyces cerevisiae*, however, only the distantly related Gemin2 ortholog Brr1p, but no SMN ortholog, could be identified. Brr1p has been proposed to be an ortholog of human Gemin2 in an early publication (5). However, because of its limited homology to Yip1p this finding has been questioned (16). Taking advantage of the dramatically increased genome databases and novel search algorithms (PSI BLAST), Brr1p can be defined as the single significant homolog of human Gemin2 (NP_003607.1, PSI BLAST third iteration, $E = 3 \times 10^{-5}$) and *S. pombe* Yip1p (NP_594775.1, fourth iteration, $E = 3 \times 10^{-5}$). Because reciprocal searches further support this homology, Brr1p is the ortholog of human Gemin2. Thus, *S. cerevisiae* retained only one Gemin and hence is unlikely to form a functional SMN complex. Interestingly, like *S. pombe*, the plants *Arabidopsis thaliana* and *Oryza sativa* contained orthologs of SMN and Gemin2 only. We therefore conclude that SMN and Gemin2 represent the most primitive and ancestral version of the SMN complex. Surprisingly, the genome of *Dictyostelium discoideum*, a facultative multicellular organism, encoded orthologs of Gemin3 and Gemin5. Given that *D. discoideum* is basal to fungi and metazoans, the bioinformatic data suggest that both Gemins have been lost during evolution in the fungi branch but were retained in metazoans. Interestingly, the presence of a Gemin5 ortholog in *Ostreococcus tauri* but its absence in land plants indicates an independent gene loss in this phylum. Moreover, we found SMN and Gemins 2, 3, 5, 6, 7, and 8 to be present in the cnidarian *Nematostella vectensis*, a basic metazoan. Gemin4 first appeared in the sea urchin *Strongylocentrotus purpuratus*, whereas it is absent in all ecdysozoans under study. This suggests that Gemin4 has joined the SMN complex only recently in evolution, most likely with the appearance of deuterostomians. Consequently, it was found to be part of the SMN complex in vertebrates such as *Danio rerio* but also cephalochordates like *Branchiostoma floridae* and *Ciona intestinalis* (urochordates). Thus, plants and some fungi possess a core complex composed of SMN and Gemin2 only, whereas an elaborate SMN complex has developed only in animal branches by addition of Gemin proteins.

Absence of Most Gemin Orthologs in Genomes of Dipterans. The bioinformatic data indicated an evolutionary trend in the animal kingdom toward a multisubunit SMN complex. Interestingly, however, we failed to identify orthologs of most Gemins in the dipterans *Drosophila melanogaster* and *Anopheles gambiae* although they were present in closely related *Apis mellifera* and *Nasonia vitripennis*. We have restricted our further analysis to *D. melanogaster* in this study. Besides the known SMN ortholog (17, 18), we found a Gemin2 ortholog encoded by CG10419 and putative orthologs of Gemin3 (Dhh1) and Gemin5 (Rigor mortis). Dhh1 protein shows high conservation in the N-terminal DEAD box helicase domain but possesses a diverged C terminus. Rigor mortis displays moderate homology to Gemin5 over the entire protein length. A phylogenetic analysis revealed that both evolve significantly faster ($P < 0.001$, see also *Experimental Procedures*) than their orthologs in other organisms. This released evolutionary pressure might indicate the emergence of a novel function or the loss of a common one for these factors. These data suggest that *D. melanogaster* possesses a much simpler SMN complex as compared with vertebrates.

Biochemical Investigation of Dhh1 and Rigor Mortis. To investigate whether Dhh1 and Rigor mortis have retained their function in the

context of the *D. melanogaster* SMN (dSMN) complex, we made use of a novel epitope tag. This tag consists of the first 30 aa of human SMN protein, which are specifically recognized by the monoclonal antibody 7B10 (19). Importantly, competition with synthetic peptide comprising this epitope allows native elution of tagged proteins from this antibody. Because *D. melanogaster* SMN protein lacks these 30 aa, we constructed a plasmid allowing the expression and subsequent purification of a protein fused to this epitope (termed *TagIt* epitope) after stable transfection of Schneider2 cells. In a *TagIt*-Dhh1 affinity purification, only small amounts of dSMN protein could be detected under physiological conditions but not at salt concentrations exceeding 250 mM (Fig. 2*A*). Thus, Dhh1 is only weakly associated with dSMN. Similarly, we have investigated the role of Rigor mortis (Fig. 2*B*; see also Fig. S1). No binding of Rigor mortis to dSMN has been observed, arguing against a stable association of this protein with the dSMN complex. These data suggest that Rigor mortis either functions in UsnRNP core formation in a manner different from vertebrate Gemin5 or has completely lost its function in the pathway of UsnRNP biogenesis.

Biochemical Isolation of the *Drosophila* SMN Complex Using the *TagIt* Epitope. To gain detailed insight into the composition of the *D. melanogaster* SMN complex, a *TagIt*–dSMN-expressing Schneider2 cell line was generated. Importantly, *TagIt*–dSMN was incorporated into a high-molecular-weight complex [≈20 Svedberg (S) units] that also contained Gemin2 (Fig. 2*C*). This implied that the tagged dSMN protein engages in interactions similar to those of its endogenous counterpart. We next affinity-purified the SMN complex from extracts by means of 7B10 affinity chromatography (see *Experimental Procedures* for details). Affinity-purified proteins were separated by SDS-PAGE under reducing and nonreducing conditions and identified by protein mass spectrometry and Western blotting (Fig. 2 *D* and *E*). Whereas the tagged SMN protein and its interactor dGemin2 could be readily identified, neither Dhh1 nor Rigor mortis was found under the purification conditions applied here.

It is known that the human SMN complex consists of the core machinery (i.e., SMN and Gemins) as well as the transiently interacting substrates that are transferred onto the UsnRNA during assembly. These are the Sm proteins and some UsnRNP-specific proteins. Strikingly, the entire set of Sm proteins, namely SmB, SmD1 (gene snRNP69D), SmD2 (CG1249), SmD3, SmE (CG18591), SmF (DebB), and SmG (CG9742), was prominently present in the elution. Furthermore, we also found the UsnRNP-specific factors U1 70K (20), U2A′ (21), the U2B″/U1A ortholog SNF (22), and the ortholog of the U5 specific protein (CG4849) U5 116kD (23) reproducibly in the purified complex. However, the abundance of these specific proteins varied among preparations and was often substoichiometric.

During UsnRNP assembly, the SMN complex physically contacts the UsnRNAs (4). In vertebrates, this interaction has been proposed to be mediated, at least in part, by Gemin5 and to occur in the cytoplasm (12). Interestingly, despite the absence of Rigor mortis in the *TagIt*–dSMN complex, snRNAs U1, U2, U4, and U5 were specifically coprecipitated with dSMN and dGemin2 antibodies from total Schneider2 cell extract (Fig. 3*A*). Identical results were obtained when the SMN complex was purified from the cytosol, where SMN is predominantly localized (Fig. 3*B* and Fig. S1*B*). Hence, in *D. melanogaster*, the SMN complex is sufficient to recruit a set of substrate proteins similar to those in vertebrates. In addition, the complex interacts specifically with U snRNAs in the cytoplasm, which reflects a situation previously observed in *Xenopus laevis* oocytes (4).

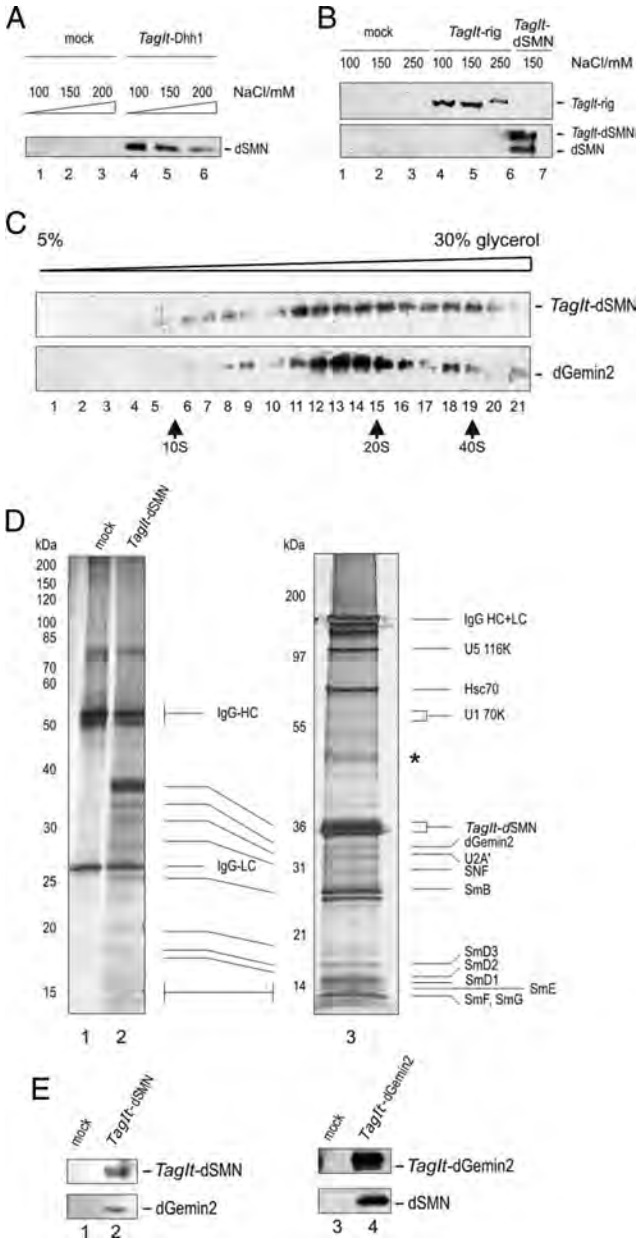


Fig. 2. *TagIt*–dSMN binds dGemin2 and Sm proteins. (*A*) Coimmunoprecipitation of *TagIt*–Dhh1 with dSMN from Schneider2 cells at increasing salt conditions (lanes 4–6). The coprecipitated dSMN protein was detected by Western blot. Lanes 1–3 show control immunoprecipitations. (*B*) Extracts from cells expressing *TagIt*–Rigor mortis (lanes 4–6), *TagIt*–dSMN (lane 7), or no tagged protein (lanes 1–3) were immunoprecipitated with antibody 7B10. Immunoprecipitates were analyzed by Western blotting with antibodies against dSMN and Rigor mortis, respectively. (*C*) Extracts from *Drosophila* Schneider2 cells stably expressing *TagIt*–dSMN were separated on glycerol gradients and analyzed by Western blotting with 7B10 (*Upper*) an anti-dGemin2 antibody (*Lower*). Estimated sedimentation value is indicated. (*D*) Isolation of *TagIt*–dSMN from Schneider2 extracts. Proteins were separated by SDS/PAGE under reducing (lanes 1 and 2) and nonreducing (lane 3) conditions and visualized by silver staining. Lane 1 shows a control elution of nontransfected cells. The indicated proteins were identified by mass spectrometry. (*E*) Immunoblot analyses of *TagIt*–dSMN (*Upper*) and *TagIt*–dGemin2 (*Lower*) purifications with indicated antibodies. Lanes 1 and 3 show mock controls.

Isolated SMN Complex Is Active in UsnRNP Assembly. A series of studies has suggested that Gemins 2, 3, 4, and 5 are essential players in the assembly reaction in vertebrates (7, 12, 13). Hence, we tested whether the minimal SMN complex in *D. melanogaster* is sufficient to mediate formation of the UsnRNP core domain. To this end, ^{32}P-labeled dU1snRNA was incubated with affinity-purified dSMN complex (Fig. 3*C*). dU1snRNA lacking the Sm site (dU1ΔSm) and an unrelated small Cajal body-associated RNA [U85scaRNA (24)]

EVOLUTION

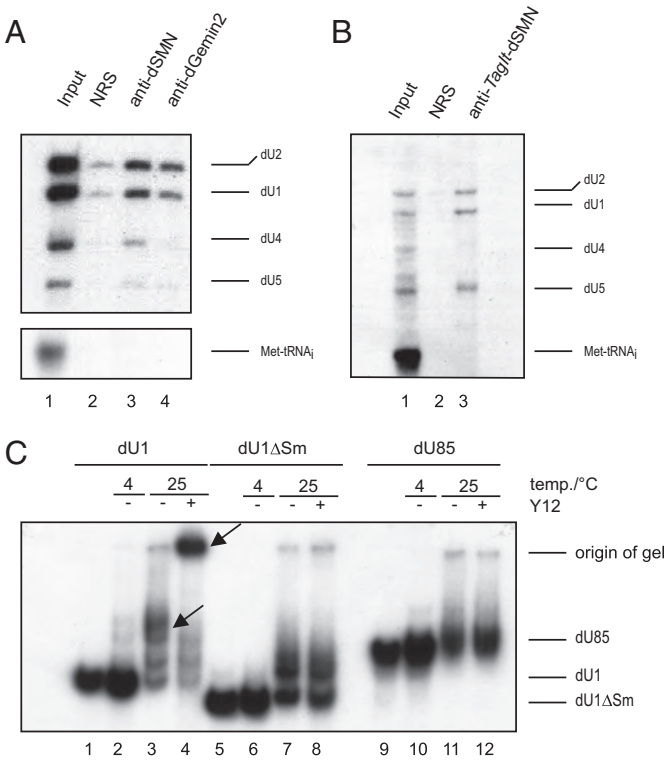


Fig. 3. dSMN complex contains UsnRNAs and is active in UsnRNP assembly. (*A*) Northern blot analysis revealed snRNAs U1, U2, U4, and U5 but not Met-tRNA$_i$ in anti-dSMN (lane 3) and anti-dGemin2 (lane 3 and 4) immunoprecipitates from Schneider2 total cell extract. Lane 1 shows the extract before immunoprecipitation, and lane 2 shows a control immunoprecipitation with a normal rabbit serum (NRS). (*B*) The same set of snRNAs was purified from cytosolic extract of *TagIt*–dSMN-expressing cells. U4 snRNA was detected after longer exposure of the film. (*C*) *In vitro* transcribed U1snRNA, U1ΔSm, and U85scaRNA were incubated with purified *TagIt*–dSMN complex at the indicated temperatures (lanes 2–4, 6–8, and 10–12) and separated by native gel electrophoresis. Lanes 1, 5, and 9 show the indicated RNAs in the absence of SMN complex; in lanes 4, 8, and 12 monoclonal antibody Y12 was added after the assembly reaction had been completed. The assembled dU1 Sm core domain and the supershift are indicated by arrows.

served as negative controls. The reaction products were analyzed by native gel electrophoresis. Incubation of dU1snRNA but not the control RNAs with the dSMN complex at 25°C led to the formation of a specific complex whose formation was inhibited at 4°C (Fig. 3*C*, lanes 2 and 3). This complex could be shifted to lower mobility upon incubation with the anti-Sm antibody Y12 (Fig. 3*C*, lane 4). Thus, isolated dSMN complex mediates UsnRNP core assembly.

Reconstituted dSMN Complex Binds Sm Proteins and Catalyzes UsnRNP Assembly. Because we could not exclude that substoichiometric factors contributed to UsnRNP core formation, we reconstituted the SMN complex from recombinant proteins, assuming that it consists of SMN and Gemin2 only. His-dSMN and His-GST-dGemin2 were coexpressed in *Escherichia coli* and purified on Ni-NTA-Sepharose. After immobilization on glutathione-Sepharose the dimer was loaded with recombinant Sm heterooligomers B/D3, D1/D2, and E/F/G. Remarkably, individual oligomers bound only weakly to the dSMN/dGemin2 dimer (Fig. 4*A*). However, binding of D1/D2 was greatly enhanced in the presence of E/F/G, indicating cooperative binding of these units. Likewise, B/D3 bound only efficiently to the complex when the other Sm proteins were present. Simultaneous incubation of all heterooligomers with the dSMN/dGemin2 dimer resulted in its efficient loading with the complete set of Sm proteins (Fig. 4*A*, lane 15).

To test its function in UsnRNP assembly, the dSMN/dGemin2 dimer was immobilized on glutathione-Sepharose beads and incubated with all Sm proteins. After washing, the protein complexes were eluted with glutathione and incubated with dU1snRNA, dU1ΔSm snRNA, or dU85scaRNA (Fig. 4 *B–D*). As a control, the same procedure was performed with His-GST protein instead of the SMN/Gemin2 complex. Sm core formation was observed only upon incubation of the SMN complex with dU1snRNA but not with the His-GST control (Fig. 4*B*). Importantly, Sm proteins bound to the SMN complex could not be transferred onto RNAs lacking a functional Sm site (Fig. 4 *C* and *D*). Thus, *in vitro* reconstituted dSMN complex is sufficient for Sm core assembly. To test whether a vertebrate SMN/Gemin2 dimer is likewise able to mediate this reaction alone, similar experiments were conducted using human SMN and Gemin2 (Fig. S2). Indeed we found not only binding of all Sm proteins to hSMN and hGemin2 (Fig. S2*A*) but also UsnRNP core assembly in an *in vitro* reaction (Fig. S2*B*).

UsnRNP assembly can occur spontaneously *in vitro* but depends on the SMN system *in vivo*. We therefore considered the possibility that the SMN complex confers specificity to the assembly reaction and hence prevents misassembly. To test this, we compared the specificity of spontaneous and SMN complex-mediated assembly. Remarkably, spontaneous binding of isolated Sm proteins occurred even under stringent conditions not only to the dU1snRNA target (Fig. 4*B*, lane 5), but also to the control RNAs lacking an Sm site (Fig. 4*B*, lanes 10 and 15). In contrast, Sm proteins bound to the SMN complex were exclusively transferred onto the cognate snRNA (Fig. 4*B*, lane 4), whereas noncognate RNAs were ignored (Fig. 4*B*, lanes 9 and 14). Hence, the minimal SMN complex of *D. melanogaster* acts not only as an assembly device but also as a chaperone that discriminates target and nontarget RNAs of Sm proteins.

Discussion

The human SMN complex belongs to a growing list of factors that assist assembly of RNA–protein complexes (25–27). However, because of its considerable complexity, only little is known about its mode of action and the contribution of individual factors to the assembly reaction. We therefore searched databases for organisms with simpler SMN complexes. This has not only allowed insight into the evolution of this assembly machinery, but has also led to the identification of the *D. melanogaster* SMN complex as a minimal unit amenable to biochemical investigation of the assembly process.

Previous studies have indicated that Gemins interact within the SMN complex in a modular manner (7, 8). Interestingly, our homology searches for components of the SMN complex in a variety of organisms recapitulated this finding on an evolutionary scale. The most simple SMN-containing complex is composed of SMN and Gemin2 only and can be found in unicellular organisms such as the fission yeast *S. pombe* (16) and in plants. The next level of complexity is characterized by the appearance of Gemin3 in *D. discoideum*, thus predating the emergence of the Fungi/Metazoa clade. The absence of Gemin5 from genomes of fungi and land plants and its presence in the green algae *O. tauri* and in *D. discoideum* indicate independent secondary gene loss in fungi and plants. This may be due to a role of Gemin5 outside of the SMN complex, which is not retained in these organisms (see also below).

Only later in evolution at the level when first metazoans developed, the building block composed of Gemins 6, 7, and 8 was added to the set of the Gemin family. From this time on, organisms had the potential to express an SMN complex similar in architecture to the human one. The only component that was not present at that point was Gemin4, which can be found only in the genomes of deuterostomians. Thus, our data suggest that the SMN complex evolved by a blockwise addition of Gemins to an ancient core complex of SMN and Gemin2 in a manner corresponding to their mutual biochemical association.

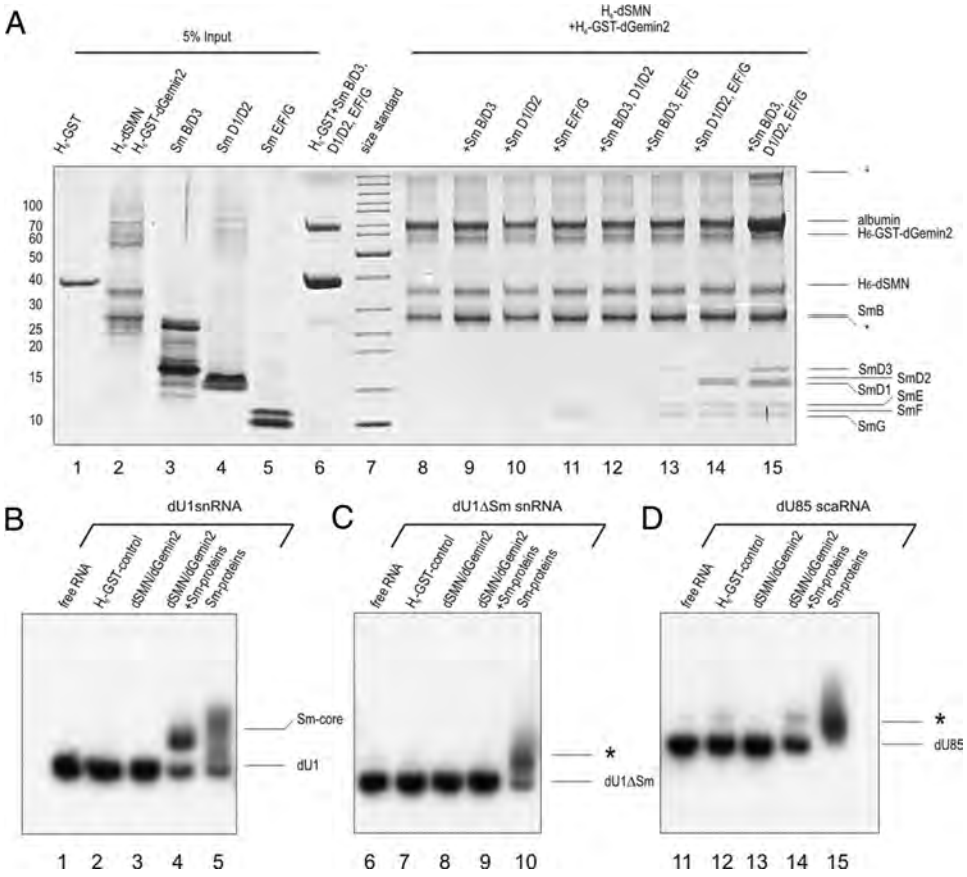


Fig. 4. Reconstitution of functional dSMN complex. (*A*) Recombinant H_6-GST-dGemin2/H_6-dSMN complex was immobilized on glutathione-Sepharose beads and incubated with purified Sm protein heterooligomers as indicated (lanes 9–15). As a specificity control, H_6-GST was immobilized on beads and mock-loaded with Sm proteins (lane 6). After removal of unbound proteins, reconstituted complexes were analyzed by SDS/PAGE. Lanes 1–5 show proteins used for the reconstitution assay. Bands indicated by an asterisk are degradation products or aggregates. *In vitro* assembly assay of reconstituted dSMN complex with dU1snRNA (*B*), dU1ΔSm (*C*), and dU85scaRNA (*D*) is shown. Radiolabeled RNAs were incubated with His-GST control (lanes 2, 7, and 12), dSMN/dGemin2 dimer (lanes 3, 8, and 13), or dSMN/dGemin2 bound to Sm proteins (lanes 4, 9, and 14), or with free Sm proteins (lanes 5, 10, and 15). Lanes 1, 6, and 11 show the RNA in the absence of protein. Reaction mixtures were separated by native gel electrophoresis, and complexes were visualized by autoradiography. The light upper band seen most prominently in lanes 12 and 14 denotes a conformer of dU85scaRNA.

In striking contrast to this overall evolutionary trend, we found a remarkable simplification of this complex in the dipterans *A. gambiae* and *D. melanogaster*. In these animals, we failed to identify orthologs of Gemin4, as expected, but also of Gemins 6–8. However, these latter Gemins were clearly present in hymenopterans. Although orthologs of Gemin3 and Gemin5 were found in dipterans, they show a significantly higher evolutionary rate in dipterans than in other clades. These computational findings have been experimentally challenged by a biochemical approach that has allowed us to isolate an assembly-active SMN complex from *D. melanogaster*. Indeed, the composition of the complex was remarkably simple and consisted of SMN and Gemin2 as the only stoichiometric components. Dhh1 (Gemin3) bound to this core complex only at low salt concentrations, and Rigor mortis (Gemin5) was not present at all. The *D. melanogaster* SMN complex therefore equals its counterpart in *S. pombe* and plants although the function of SMN and Gemin2 orthologs in these organisms has not been demonstrated. It is conceivable that Dhh1 and Rigor mortis have adopted novel functions in a different context because they rapidly diverge from their ancestors (28). Consistent with this notion, a function of Rigor mortis in ecdysone signaling has been described (29).

Despite the obvious simplicity of the SMN complex in *D. melanogaster*, we provide evidence that this unit is functionally related to the SMN complex of mammals. First, a set of UsnRNP-related substrates, namely the common Sm proteins, UsnRNP specific factors (U1 70K, U2A′, U2B″/U1A, and U5 115K), and UsnRNAs were found to be part of the complex. Most of these factors have previously been shown to bind to SMN complexes of vertebrates (30). Second, affinity-purified dSMN complex mediated the assembly of the Sm core domain *in vitro*. Similar to the situation in humans, we found a strong dependence of UsnRNP core assembly on temperature but not on ATP (30). However, at present we cannot rule out the possibility that assembly of UsnRNPs in *D. melanogaster* cytosolic extracts requires ATP hydrolysis as observed for the same reaction in vertebrates (6, 31).

The obvious simplicity of the assembly system of *D. melanogaster* allowed the reconstitution of the dSMN complex from recombinant proteins and the investigation of its mode of action. Interestingly, we observed strong cooperativity in Sm protein binding onto the complex. Heterooligomers D1/D2 and B/D3 had only little affinity for the complex, but binding was greatly enhanced in the presence of E/F/G. Further studies are required to determine the precise binding sites of all Sm proteins on the SMN complex and to test the influence of arginine methylation on Sm protein binding (32, 33). It was an open question why UsnRNP assembly is strictly dependent on the SMN complex *in vivo* even though this reaction is spontaneous *in vitro*. Our assembly studies with the *D. melanogaster* SMN complex show that precise assembly of the Sm core domain on UsnRNA was possible only when Sm proteins were prebound to the SMN complex, whereas misassembly of isolated Sm proteins occurred under the same conditions. In addition, we demonstrate that human SMN and Gemin2 are likewise sufficient to specifically transfer Sm proteins onto UsnRNA. Hence, these data and similar studies performed in vertebrates argue for a dual role of the SMN complex as an RNP assembler and chaperone (31).

From an evolutionary point of view, our findings raise the question why dipterans can afford a minimized assembly system, whereas apparently other branches in the animal kingdom require a multicomponent SMN complex. The most plausible explanation for this paradox is that Gemins 3–8 are not primarily involved in the assembly reaction *per se* but rather in other steps during the UsnRNP biogenesis. Thus, it is known that the human SMN complex integrates several steps in biogenesis, such as cap hypermethylation (10) and nuclear import (9). We speculate that these steps will occur in dipterans independent of the SMN complex and

EVOLUTION

may hence allow for the omission of individual Gemins. Further studies will be needed to test whether this is indeed the case.

In conclusion, our studies have shown that the integration of bioinformatics and biochemistry can be used to analyze cellular pathways functionally and evolutionarily. Similar strategies may prove to be powerful tools in the analysis of even more complex systems such as the spliceosome.

Experimental Procedures

Bioinformatic Strategies. ***Ortholog identification.*** To identify orthologs of SMN complex members, a PSI BLAST profile was generated for each protein by searching for five iterations with the human protein against the National Center for Biotechnology Information nonredundant peptide database (nr). The resulting profile was searched against the proteomes of the selected organisms. For validation, each candidate ortholog was manually searched backwards against nr and accepted only if known members of the SMN complex were retrieved significantly as best hits. In cases of unclear orthologous relationships caused for example by widespread domains in the query protein (Gemin3 and Gemin5) a phylogenetic tree was calculated. This allowed the separation of the orthologous subgroup from paralogs. Additionally, each protein was searched manually against nr to identify orthologs missing in the genome-specific gene prediction as well as against the National Center for Biotechnology Information's whole genome shotgun reads to identify orthologs missing in the genome assembly.

Evolutionary rate analysis for dipteran Gemin3 and 5. Gemin3 and Gemin5 protein sequences from all organisms under investigation were aligned, and a phylogenetic tree was calculated by using proml from the PHYLIP package (34). In both cases of Gemin3 and Gemin5, the tree followed the species tree. To test whether the evolutionary rate of the dipteran proteins significantly differs from the other Gemin3 and Gemin5 orthologs, respectively, we used the codeml program of the PAML package (35). Using the topology of the calculated tree, two models were assumed and the fit of the data to the models was tested. First, a constant evolutionary rate was assumed all over the tree. Second, the dipteran branch (i.e., the one leading from the branching of *A. mellifera* to the speciation of *Drosophila* and *Anopheles*) was allowed a different rate. For both Gemin3 and Gemin5, the second model generated a faster evolutionary rate for the dipteran branch. To test whether this result was significant, a log-likelihood ratio test was performed. Here, the duplicated difference between the log-likelihoods of both models has to exceed a given cutoff drawn from the χ^2 distribution. Because the only difference between the models was the rate on one branch, one degree of freedom was assumed. In both cases, the log-likelihood ratio exceeded the χ^2 critical value for $P = 0.001$. A list of cDNAs and oligonucleotides used in this study is provided in Tables S2 and S3.

Generation of Stable Cell Lines and Affinity Purification of Complexes. *Drosophila* Schneider2 cells (37) were cultured in Schneider's *Drosophila* medium (Biowest) containing 10% vol/vol FCS (PAA Laboratories). Cells were transfected with the respective pM*TagIt* expression construct using calcium phosphate. For the generation of stable cell lines, the pCoBlast vector (Invitrogen) was cotransfected and resistant cells were selected with blasticidin at 25 μg/ml (Invivogen). Induction of protein expression was performed with copper sulfate in a concentration determined for each construct. For extract preparation, cell pellets were resuspended in 3 volumes of PBS supplemented with 0.01% Igepal CA630 (PBS-I; Sigma) and disrupted by sonication. The lysate was cleared by centrifugation and subsequent filtration through a 0.25-μm syringe filter. For cytosolic extract preparation, Schneider2 cells were resuspended in 20 mM Hepes/KOH (pH 7.4), 80 mM potassium acetate, 4 mM magnesium acetate, and 50 μg/ml digitonin with protease inhibitors. After lysis for 10 min at room temperature, extract was cleared by centrifugation at 25,000 $\times$ *g*.

For *TagIt* purifications, extracts were incubated monoclonal antibody 7B10 (6) coupled to protein G Sepharose (GE Healthcare). After removal of unbound proteins by extensive washing, complexes were eluted with a 5-fold molar excess of synthetic peptide comprising the epitope recognized by 7B10.

For gradient centrifugation, 200 μl of extract prepared from S2 cells was layered on top of a 5–30% vol/vol glycerol gradient and centrifuged at 4°C at 38,500 rpm for 5 h in a Sw60Ti rotor. Fractions were separated by SDS/PAGE, and proteins were detected by Western blot analysis as described (30).

Additional Details. For additional experimental procedures see *SI Text*.

ACKNOWLEDGMENTS. We thank members of our labs for help and reagents. M.K. is a fellow of the M.D./Ph.D. program of the University of Würzburg. U.F. was supported by Deutsche Forschungsgemeinschaft Grants SFB581 (TP18) and FZT 82.

1. Will CL, Luhrmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol* 13:290–301.
2. Raker VA, Plessel G, Luhrmann R (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.
3. Raker VA, Hartmuth K, Kastner B, Luhrmann R (1999) Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol Cell Biol* 19:6554–6565.
4. Fischer U, Liu Q, Dreyfuss G (1997) The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* 90:1023–1029.
5. Liu Q, Fischer U, Wang F, Dreyfuss G (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.
6. Meister G, Buhler D, Pillai R, Lottspeich F, Fischer U (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat Cell Biol* 3:945–949.
7. Otter S, *et al.* (2007) A comprehensive interaction map of the human SMN complex. *J Biol Chem* 282:5825–5833.
8. Battle DJ, Kasim M, Wang J, Dreyfuss G (2007) SMN-independent subunits of the SMN complex: Identification of a SNRNP assembly intermediate. *J Biol Chem* 282:27953–27959.
9. Narayanan U, Achsel T, Luhrmann R, Matera AG (2004) Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol Cell* 16:223–234.
10. Mouaikel J, *et al.* (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep* 4:616–622.
11. Feng W, *et al.* (2005) Gemins modulate the expression and activity of the SMN complex. *Hum Mol Genet* 14:1605–1611.
12. Battle DJ, *et al.* (2006) The Gemin5 protein of the SMN complex identifies snRNAs. *Mol Cell* 23:273–279.
13. Shpargel KB, Matera AG (2005) Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc Natl Acad Sci USA* 102:17372–17377.
14. Hunt SL, Hsuan JJ, Totty N, Jackson RJ (1999) unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes Dev* 13:437–448.
15. Grimmler M, *et al.* (2005) Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. *Hum Mol Genet* 14:3099–3111.
16. Hannus S, Buhler D, Romano M, Seraphin B, Fischer U (2000) The Schizosaccharomyces pombe protein Yab8p and a novel factor, Yip1p, share structural and functional similarity with the spinal muscular atrophy-associated proteins SMN and SIP1. *Hum Mol Genet* 9:663–674.
17. Miguel-Aliaga I, Chan YB, Davies KE, van den Heuvel M (2000) Disruption of SMN function by ectopic expression of the human SMN gene in Drosophila. *FEBS Lett* 486:99–102.
18. Rajendra TK, *et al.* (2007) A Drosophila melanogaster model of spinal muscular atrophy reveals a function for SMN in striated muscle. *J Cell Biol* 176:831–841.
19. Meister G, *et al.* (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.
20. Salz HK, *et al.* (2004) The Drosophila U1–70K protein is required for viability, but its arginine-rich domain is dispensable. *Genetics* 168:2059–2065.
21. Nagengast AA, Salz HK (2001) The Drosophila U2 snRNP protein U2A′ has an essential function that is SNF/U2B″ independent. *Nucleic Acids Res* 29:3841–3847.
22. Stitzinger SM, Conrad TR, Zachlin AM, Salz HK (1999) Functional analysis of SNF, the Drosophila U1A/U2B″ homolog: Identification of dispensable and indispensable motifs for both snRNP assembly and function in vivo. *RNA* 5:1440–1450.
23. Fabrizio P, Laggerbauer B, Lauber J, Lane WS, Luhrmann R (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J* 16:4092–4106.
24. Jady BE, Kiss T (2001) A small nucleolar guide RNA functions both in 2′-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* 20:541–551.
25. Kiss T, Fayet E, Jady BE, Richard P, Weber M (2006) Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harbor Symp Quant Biol* 71:407–417.
26. Eggert C, Chari A, Laggerbauer B, Fischer U (2006) Spinal muscular atrophy: The RNP connection. *Trends Mol Med* 12:113–121.
27. Maki JA, Schnobrich DJ, Culver GM (2002) The DnaK chaperone system facilitates 30S ribosomal subunit assembly. *Mol Cell* 10:129–138.
28. Beltrao P, Serrano L (2007) Specificity and evolvability in eukaryotic protein interaction networks. *PLoS Comput Biol* 3:e25.
29. Gates J, Lam G, Ortiz JA, Losson R, Thummel CS (2004) rigor mortis encodes a novel nuclear receptor interacting protein required for ecdysone signaling during Drosophila larval development. *Development* 131:25–36.
30. Meister G, Fischer U (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J* 21:5853–5863.
31. Pellizzoni L, Yong J, Dreyfuss G (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science* 298:1775–1779.
32. Gonsalvez GB, Rajendra TK, Tian L, Matera AG (2006) The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. *Curr Biol* 16:1077–1089.
33. Gonsalvez GB, *et al.* (2007) Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. *J Cell Biol* 178:733–740.
34. Felsenstein, J (2005) PHYLIP (Phylogeny Inference Package). (Department of Genome Sciences, Univ of Washington, Seattle), Version 3.6.
35. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
36. Kambach C, *et al.* (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* 96:375–387.
37. Schneider I (1972) Cell lines derived from late embryonic stages of Drosophila melanogaster. *J Embryol Exp Morphol* 27:353–365.
38. Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440:623–630.

Supporting Information

Kroiss *et al*. 10.1073/pnas.0802287105

SI Text

Databases. To analyze the phylogenetic distribution of the members of the SMN complex, proteomes of selected marker organisms were downloaded. These included from Ensembl the proteome of *Homo sapiens*, *Danio rerio*, *Ciona intestinalis*, *Anopheles gambiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (http://www.ensembl.org). The proteome from *Nematostella vectensis*, *Branchiostoma floridae*, and *Ostreococcus tauri* were downloaded from JGI (http://www.jgi.doe.gov/). The proteome of *Strongylocentrotus purpuratus* was downloaded from NCBI (http://ncbi.nlm.nih.gov). Additionally, the following proteomes were extracted from organism specific databases: *Saccharomyces cerevisiae*, SGD (http://www.yeastgenome.org); *Schizosaccaromyces pombe*, Sanger Centre (http://sanger.ac.uk); *Apis mellifera*, BeeBase (http://beebase.csl.gov.uk); *Dictyostelium discoideum*, DictyBase (http://dictybase.org/); *Arabidopsis thaliana*, TAIR (http://arabidopsis.org/); *Oryza sativa* ssp. *japonica*, MIPS (http://mips.gsf.de/).

Protein MS. Sample preparation for MS was performed according to Shevchenko *et al.* (1). Nano-LC-MS/MS analyses were accomplished on a LTQ XL (Thermo Scientific), a QStarElite, and a Qtrap 4000 (both Applied Biosystems), respectively coupled to Ultimate or Ultimate 3000 nano-HPLC systems (both Dionex). For this, peptides were preconcentrated in 0.1% TFA on a 150-μm ID self-packed RP trapping column (HydroRP, 2-cm length, 4-μm particle size) (2) and afterward separated on a self-packed 75-μm ID RP main column (HydroRP, 15-cm length, 2-μm particle size) by applying a 40-min binary gradient (solvent A, 0.1% FA; solvent B, 0.1% FA, 84% ACN) ranging from 5% to 50% of solvent B at a flow rate of 250 nl/min. After elution the column was rinsed with 95% solvent B for 10 min and subsequently equilibrated with 5% B for 20 min. Peptides were directly eluted into an ESI MS. MS acquisition duty cycle was set up to a 1-s survey scan followed by five dependent scans, each $\approx$1 s. Mass spectra were transformed into peak lists in mgf format using an in-house software solution, LCQ_dta and wiff2dta. Default values for generating mgf files were applied and data were processed using the search algorithm MASCOT (version 2.2, Matrixscience) (3). Database searches were accomplished by using the Flybase database (version 4.3 March 2006, http://chervil.bio.indiana.edu:7092/). As fixed modification, carbamidomethylation of cysteines was selected and oxidation of methionine as variable modification. Trypsin with a maximum of one missing cleavage site was chosen as enzyme, peptide and MS/MS tolerances in daltons were selected depending on the respective mass analyzer (0.8/0.8 for the LTQ XL, 0.2/0.5 for the QStarElite, 0.4/0.4 for the Qtrap 4000). Only peptide spectra with a MASCOT score above 40 were considered for further data interpretation. Additionally, only proteins with at least three manually validated spectra of different tryptic peptides were accepted as true positive identifications.

Northern Blotting. RNA was separated by denaturing 10% RNA/PAGE and blotted at 0.8 mA/cm^2 in TBE buffer (89 mM Tris, pH 8.3/89 mM boric acid/2 mM EDTA) for 2 h onto a Hybond N membrane (GE Healthcare). The membrane was blocked with RapidHyb hybridization buffer (GE Healthcare). A total of 10 pmol oligonucleotide probe each was radioactively labeled with 30 μCi of [^{32}P]ATP (PerkinElmer) using polynucleotide kinase (Fermentas) according to the manufacturer's description and PAGE-purified. Hybridization was carried out at 42°C for 2 h. After washing with SSC buffer (150 mM NaCl/15 mM sodium citrate), the membrane was exposed on Hyperfilm MP (GE Healthcare).

Immunofluorescence Microscopy. For immunofluorescence microscopy, Schneider2 cells were grown on coverslips and processed as described earlier (4). Imaging was carried out by using a Zeiss Axiovert 200 M equipped with a $\times$63 Plan NeoFluar objective, standard filter sets, and Axiocam MRm.

Cloning and Reconstitution of the dSMN Complex. cDNAs (see Tables S2 and S3) were PCR-amplified and cloned into expression vectors pETM-30, pM*TagIt*, and pBADM-11, respectively. For expression of the dSMN-dGemin2 dimer, a dicistronic plasmid was generated applying a linker ligation approach (36) and transformed into *E. coli* RosettaII (Novagen). Expression was induced for 4 h with 0.02% arabinose, and proteins were purified on a Ni-NTA matrix. Purified complex was then bound to glutathione-Sepharose beads in the presence of 1 mg/ml BSA and incubated for 1 h with Sm protein heterodimers (36). Bound proteins were eluted with SDS sample buffer and analyzed by SDS/PAGE. For assembly assays, complexes were eluted with 30 mM glutathione, incubated with *in vitro* transcribed RNA, and analyzed as described (30). The human SMN-Gemin2 heterodimer was reconstituted by the cotransformation of GST-Gemin2 and H_6-SMN as described (7). After reconstitution with Sm proteins on glutathione-Sepharose, the human complexes were eluted by incubation with 20 μg of PreScission protease overnight at 4°C.

1. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68:850–858.
2. Mitulovic G, *et al.* (2003) An improved method for tracking and reducing the void volume in nano HPLC-MS with micro trapping columns. *Anal Bioanal Chem* 376:946–951.
3. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567.
4. Kroiss M, *et al.* (2006) Transporter regulator RS1 (RSC1A1) coats the trans-Golgi network and migrates into the nucleus. *Am J Physiol* 291:F1201–F1212.
5. Hamm J, Kazmaier M, Mattaj IW (1987) In vitro assembly of U1 snRNPs. *EMBO J* 6:3479–3485.
6. Meister G, Buhler D, Pillai R, Lottspeich F, Fischer U (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat Cell Biol* 3:945–949.

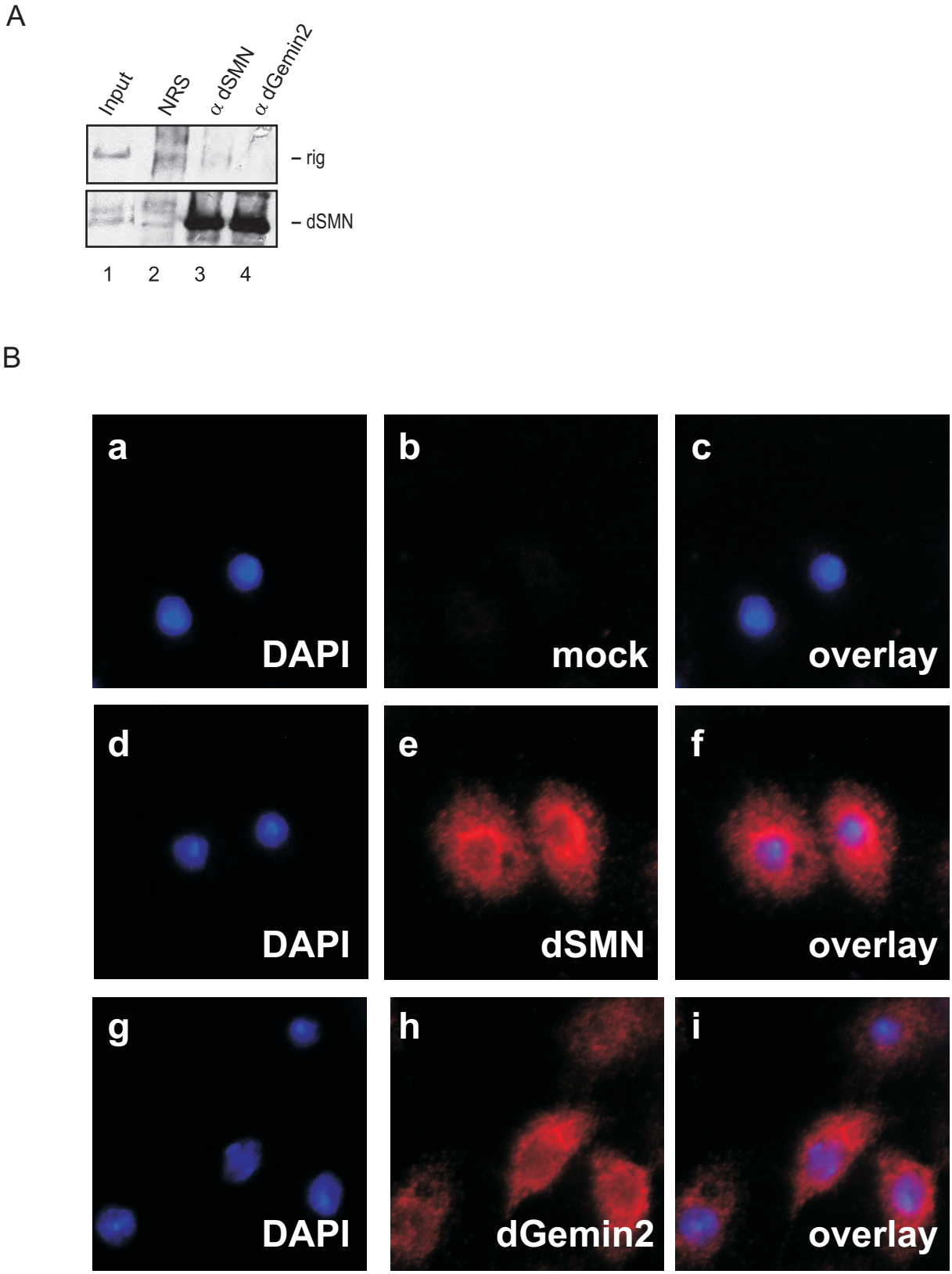


Fig. S1. (*A*) Further evidence that Rigor mortis is not a component of the *Drosophila* SMN complex. Extracts from Schneider2 cells were used for immunoprecipitations with anti-dSMN (lane 3) or anti-dGemin2 (lane 4) antibodies or normal rabbit serum (NRS) as control. The immunoprecipitates were resolved by SDS/PAGE and tested in Western blots for dSMN (*Lower*) and Rigor mortis (*Upper*). Lane 1 shows the Western blot analysis of the extract used for immunoprecipitations. (*B*) dSMN and dGemin2 localization in Schneider2 cells. Adherent Schneider2 cells were probed with antibodies against dSMN (d–f) and dGemin2 (g–i). Primary antibody was omitted in the mock control (a–c). Nuclei were stained with DAPI. Indirect immunofluorescence microscopy reveals a prominent cytoplasmic localization for both dSMN and dGemin2.

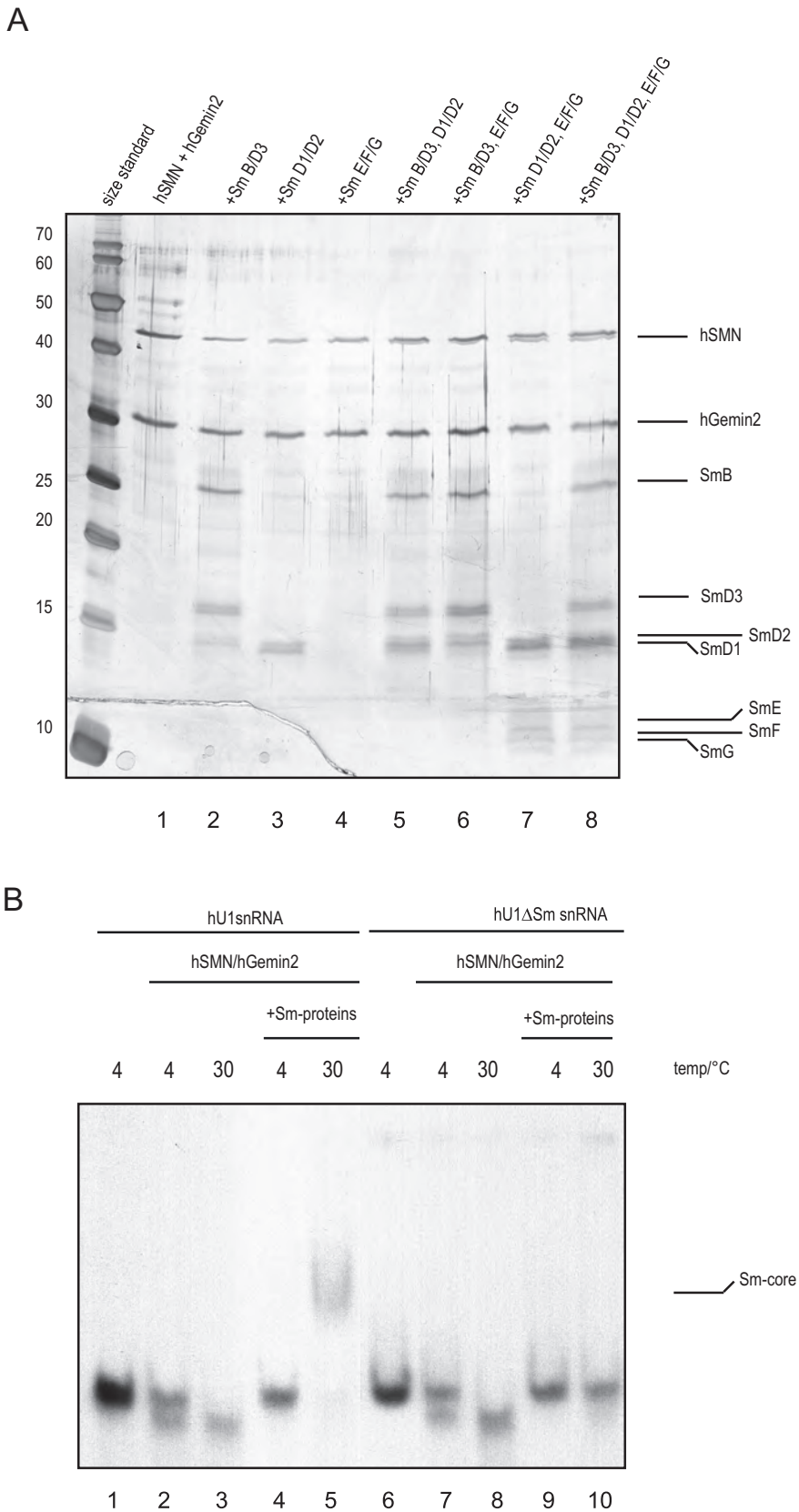


Fig. S2. (*A*) Recombinant human SMN and Gemin2 proteins were immobilized on glutathione Sepharose beads and loaded with recombinant Sm protein heterooligomers as indicated (lanes 1–7). (Note that this experiment is analogous to that shown in Fig. 4*A*.) After removal of unbound proteins, the complexes were eluted by cleavage with PreScission protease, separated by SDS/PAGE, and visualized by silver staining. (*B*) Human SMN/Gemin2 dimer alone (lanes 2, 3, 7, and 8) or loaded with Sm proteins (lanes 4, 5, 9, and 10) were incubated with hU1snRNA (lanes 1–5) or hU1ΔSm snRNA (5). Complex formation was analyzed by native gel electrophoresis as described (6), and complexes were visualized by autoradiography.

Table S1. Accession numbers of SMN complex components from various organisms

Species	SMN	Gemin2	Gemin5	Gemin3	Gemin4	Gemin8	Gemin7	Gemin6
Homo sapiens	2498924	6094289	22001417	12643886	122939157	34531032	13376001	41393577
Danio rerio	7674338	62955193	125848666	125824010	125837883	120538703	125882644	125830092
Branchiostoma floridae	Brafl1|121736	Brafl1|205562	Brafl1|93160	Brafl1|123431	Brafl1|99221	Brafl1|117877	Brafl1|270123	Brafl1|253351
Ciona intestinalis	ENSCINP00000026142	ENSCINP00000022944	ENSCINP00000012233	26554509	26554517	26554705	26554715	26554705
Strongylocentrotus purpuratus	115698848	115663092	115772524	72021085	115619072	115666277	72009598	111799074
Drosophila melanogaster	10444398	27819903	45550470	17647335	—	—	—	—
Anopheles gambiae	119115770	158294153	158299195	158294110	—	—	—	—
Apis mellifera	Amel|GB13986-PA	66514142	110760819	110758233	—	110756316	110760436	110771578
Nasonia vitripennis	156554158	156553626	156552235	156543626		156541308	Scaffold13	156543489
Caenorhabditis elegans	71981679	71993428	—	154147106	—	71984344	154147316	17506317
Nematostella vectensis	156380554	156359607	156372854	156398602	156401356 (?)	156375116	156389177	156355941
Schizosaccharo-myces pombe	1175479	74624906	—	—	—	—	—	—
Saccharomyces cerevisiae	—	6325314	—	—	—	—	—	—
Lacaria bicolor	scaffold_1|351837|352609	Lacbi1|311626	—	—	—	—	—	—
Dictyostelium discoideum	66819065	66807205	66806047	66817852	—	—	—	—
Arabidopsis thaliana	30678010	4585999	—	—	—	—	—	—
Oryza sativa	115474435	115485137	—	—	—	—	—	—
Ostreococcus tauri	116054748	116059894	116056752	—	—	—	—	116000861

Table S2. cDNA identifiers and oligonucleotide sequences used for cloning and RT-PCR

			Primer	
Gene product	DGRC clone	Gene	Forward	Reverse
dSMN	LD23602	Smn	CATGCCATGGCAATGTCCGACGAGACGAACGC	ATTTGCGGCCGCTTACTTCTTGGGTGTCTTTTTCTTTCC
dGemin2	LD47479	CG10419	CATGCCATGGCAATGCAGCATGAGCCCGAAG	ATTTGCGGCCGCTTATATGTAATCCTTGAAATCATTCTGTGC
Dhh1	LD05563	Dhh1	CGGGATCCATGGAGCGCGAAATAGCG	ATTTGCGGCCGCTTAGTAATAGTTCTGTAGATTCTTGGC
Rigor mortis	SD03652	rig	CATGCCATGGCAATGAACTCGCAGGTGATGTACC	AAATGCGGCCGCTTAATGCTCGGCAGTAGATCCATTG
dU1	Gift from Steve Mount		GGAATTCTAATACGACTCACTATAGGGATACTTACCTGGCGTAGAGGTTAACCG	CCCAAGCTTAGATCTCGGGACGGCGCGAACG
dU85	RT-PCR from Schneider2 cell total RNA		GGAATTCTAATACGACTCACTATAGGGTGCCCATGATGAAATATTCGACATCGG	CCCAAGCTTTGCGCTCAGATTACTAAAGACGTCG

Table S3. Probe sequences for Northern blot analysis

Transcript	Oligonucleotide	Gene product	Oligonucleotide
dU1	gccttcgtgatcacggttaaC	dU2	Gttgtcctccaatggaggaac
dU4	Gcacctcaggaggacttcattg	dU5	Gactcattagagtgttcctctcc
Met-tRNA$_i$	GCAGAGCAAGGTTTCGATCCTC		

6. Discussion

The experimental findings of the results section and other findings of our laboratory have been discussed in a total of 5 reviews within the time frame of this thesis. Two of these reviews focus on medical consequences of impaired function of the assembly pathway.

As mentioned briefly in the introduction section, heterozygous compound mutations or deletions in the SMN gene lead to the neuromuscular disorder spinal muscular atrophy. Phenotypically, this disease is characterized by the selective degeneration of α-motor neurons in the anterior horns of the spinal cord. Clinical signs of SMA are symmetrical muscular weakness, which spreads from the proximal to distal extremities. This frequently also impairs the patients´ ability to breath, which is one of the major cause of death in this disease.

This exclusively neuronal phenotype strongly contrasts with the finding that the SMN-complex is responsible for the assembly of U snRNPs (an activity required by all tissues of the body). In Chapter 6.1 of this discussion section, two reviews draw a model how the impairment of the formation of U snRNPs can be causative for the devolopment of the SMA phenotype. The remaining three reviews in Chapter 6.2 of this discussion section, discuss the mechanism of the cellular U snRNP assembly reaction.

6.1 The Etiology of Spinal Muscular Atrophy

Spinal muscular atrophy: the RNP connection

Eggert C, Chari A, Laggerbauer B, Fischer U: **Trends Mol Med 2006**, 12:113-121.

Wenn Muskeln die Nerven die Nerven verlieren

Fischer U, Kroiss M, Chari A: **Biospektrum 2007**, 13. Jahrgang, Ausgabe 06.07

Opinion TRENDS in Molecular Medicine Vol.12 No.3 March 2006

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Spinal muscular atrophy: the RNP connection

Christian Eggert, Ashwin Chari, Bernhard Laggerbauer and Utz Fischer

Theodor Boveri Institute, Biocenter at the University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

Degenerated motor neurons in the spinal cord are the pathological hallmark of spinal muscular atrophy (SMA). SMA is caused by mutations in the ubiquitously expressed survival motor neuron 1 (*SMN1*) gene, which lead to reduced levels of functional SMN protein. Many different functions have been assigned to SMN, including assembly of ribonucleoproteins (RNPs), splicing, transcription and axonal mRNA transport. Recently, tissue from SMA patients and animal models has been used to determine which function of SMN is affected in SMA patients. A surprising picture has emerged: the impaired assembly of RNP subunits of the spliceosome seems to be responsible for SMA pathogenesis. Here, we present a model of how this defect might cause motor-neuron degeneration and consider potential therapies.

A motor-neuron disease caused by a mutated housekeeping gene

Among the inherited disorders characterized by tissue-specific phenotypes, those that are linked to ubiquitously expressed genes are the most puzzling, for example, forms of retinitis pigmentosa (RP) [1] and spinal muscular atrophy (SMA) [2]. In both diseases, the affected genes have 'housekeeping' functions (i.e. their encoded proteins are required in every tissue). How then can a tissue-specific phenotype be explained?

A simple interpretation of this paradox might be that many housekeeping genes, although they are essential in every cell type, are not uniformly expressed among tissues. Strong constitutive expression of a gene often occurs in the tissue that most depends on the encoded protein; therefore, it is conceivable that this tissue is more likely to be affected by mutations that cause reduction or inactivation of the encoded protein. However, an equally valid idea is that low, rather than high, gene expression makes a tissue more sensitive to such mutations because the encoded protein levels might quickly drop below a critical threshold. Another way to explain this tissue-specificity paradox is that the disease gene not only has housekeeping activity, but also cell-specific functions. Disease-linked genes often encode proteins with multiple domains, and it might be speculated that mutation selectively impairs the activity of one domain without compromising other regions. If this is the case, pathogenic mutations would interfere with the cell-specific activity of the protein, whereas its general function would be unaffected. Also, the housekeeping protein might be required for, or linked to, a catalytic reaction that transforms various substrates. Faustino and Cooper [3] have hypothesized that, in these instances, when the protein is mutated, 'weak' substrates, containing sub-optimal reaction sites, are more affected than 'strong' substrates. This hypothesis is becoming increasingly attractive given that genes linked to RP and SMA encode factors that are implicated in such a catalytic reaction - namely, the processing of precursor mRNA (pre-mRNA).

Here, the genetic and biochemical advances that help identify the cellular pathways that are impaired in SMA patients are discussed. Recent data showing that SMN is a multifunctional protein with both housekeeping and neuron-specific functions are summarized. Unexpectedly, defects in the housekeeping function of SMN as a facilitator of the assembly of uridine-rich small nuclear ribonucleoproteins (U snRNPs) seem to generate the motor-neuron phenotype of SMA.

Clinical features and genetic background of SMA

Proximal SMA is a common hereditary disease characterized by the selective degeneration of α-motor neurons in the anterior horns of the spinal cord. With a prevalence of 1 in 6000 live births, SMA is one of the most prominent autosomal recessive disorders and a frequent cause of early infant mortality [2,4]. Typical clinical signs are symmetrical muscular weakness, which spreads from proximal to distal extremities. Because muscular atrophy can also impair the patients' ability to breathe, opportunistic infections such as pneumonia are a frequent cause of death.

Three clinically distinct forms of SMA have been formally classified on the basis of disease onset and severity. SMA type I (also termed Werdnig-Hoffmann disease) is characterized by muscular weakness during the first months of life: patients are unable to sit or stand and usually die within the first two years of their life. Patients suffering from SMA type II can sit but are unable to walk without help; clinical manifestation of SMA type II appears 6-18 months after birth and life expectancy of these patients is between two and 30 years of age. SMA type III (also termed Kugelberg-Welander disease) is the mildest form of the disease and is characterized by a late onset (typically after the second year of life); patients can sit, stand and walk with some restrictions, and might

Corresponding author: Fischer, U. (utz.fischer@biozentrum.uni-wuerzburg.de).
Available online 13 February 2006

 1471-4914/$ - see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.molmed.2006.01.005

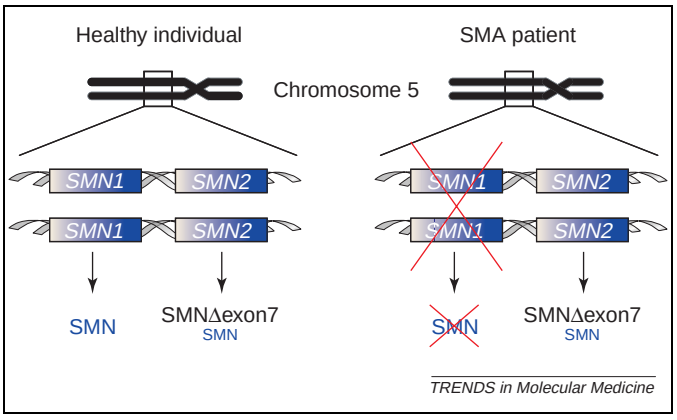


Figure 1. The *SMN* gene locus in healthy individuals and in SMA patients. Two nearly identical copies of the *SMN* gene are located on the long arm of chromosome 5 (5q13). Functional SMN protein is predominantly produced from *SMN1*, whereas the major product of *SMN2* is a truncated and non-functional protein (SMNΔexon7). Mutations that cause disease inactivate *SMN1*, leaving *SMN2* as the only source of functional protein.

reach almost average life expectancy. In the past decade, two additional types of SMA were categorized: the embryonic lethal form of SMA (type 0) and the rare adult-onset form (type IV) [5].

Despite the clinical heterogeneity of SMA, most SMA cases ($>96\%$) are due to mutations in the gene encoding the survival motor neuron protein, *SMN1* [6] (Figure 1); consequently, no functional SMN protein is produced from this mutated gene. Humans have a second copy of this gene, termed *SMN2*, which is unaffected in SMA patients. Curiously, this copy cannot compensate for the mutations in *SMN1* because $\leq 80\%$ of spliced *SMN2* transcripts lack exon 7 (SMNΔexon7) and the resulting protein rapidly degrades in cells; moreover, additional studies suggested that the SMNΔexon7 protein isoform is not functional [7]. Interestingly, some SMA patients have additional copies of the *SMN2* gene and they display less severe symptoms. This finding indicates that the severity of the disease directly correlates with the abundance of functional SMN protein [8]. No patient with complete SMN deficiency has so far been identified, which is consistent with the observation that the *Smn* knockout in mice leads to early embryonic lethality [9]. Therefore, *SMN* is considered an essential gene.

The unfavorable balance towards the SMNΔexon7 protein has attracted much attention, particularly because a shift towards the full-length (functional) SMN2 protein has therapeutic potential to compensate for the pathogenic *SMN1* mutations. A comparison between *SMN1* and *SMN2* pinpointed the reason for the increased level of exon-7 exclusion to a single-nucleotide difference in this exon. Several laboratories have provided convincing evidence that this 'cytidine→thymidine transition' in *SMN2* dramatically increases exon-7 exclusion [10,11]; this position in conjunction with its surrounding nucleotides forms a recognition site for alternative-splicing factors. However, debate is ongoing as to whether the C→T transition in *SMN2* leads to loss or gain of a regulatory signal. Kashima and Manley [12] showed that the transition results in the gain of a negative splicing element, termed exonic splicing silencer (ESS). It was suggested that ESS provides a binding platform for the negative splice regulators heterogeneous nuclear ribonucleoprotein (hnRNP) A1 and hnRNP A2, thereby preventing efficient inclusion of exon 7 into the *SMN2* mRNA [12]. Others, by contrast, have provided evidence that the C→T transition destroys a positive splicing element, exonic splicing enhancer (ESE); consequently, positive splicing factors [e.g. spicing factor 2 (SF2), also known as alternative splicing factor (ASF)] fail to bind to this ESE and, hence, inclusion of exon 7 is disfavored [7,13–15]. A recent systematic analysis by Cartegni *et al.* [16] has identified an ESS in both *SMN1* and *SMN2* genes; the authors showed that the ESS functions irrespective of the cytidine substitution. Therefore, it is unlikely that the aforementioned cytidine substitution significantly contributes to the splicing unbalance. Notably, Cartegni *et al.* [16] also identified an ESE element in *SMN1* that becomes inactivated in *SMN2* upon the C→T transition, implying that an unbalanced interplay occurs among antagonistic splicing factors acting on *SMN2* pre-mRNA, which prevents sufficient compensation of pathogenic mutations in *SMN1*.

Animal models of SMA

Much of the knowledge about SMA has been obtained from studies on mice (*Mus musculus*), fruit flies (*Drosophila melanogaster*), worms (*Caenorhabditis elegans*) and the fission yeast (*Schizosaccharomyces pombe*). Unlike humans, these organisms contain only one gene encoding SMN (equivalent to the human *SMN1* gene) and thus it is not surprising that its homozygous deletion is lethal [17–20].

To create a model of SMA that could recapitulate important aspects of the disease, two groups of researchers have generated transgenic $Smn^{-/-}$ mice that express the human *SMN2* gene [21–24]. The idea behind this strategy was that exon-7 inclusion into human *SMN2* mRNA would be similarly inefficient in mice, leading to the production of low levels of full-length SMN2 protein. Because full-length human SMN2 protein shares 98% homology with its murine ortholog, it was hypothesized that this protein would improve the condition of these mice. Indeed, this strategy circumvented embryonic lethality of the $Smn^{-/-}$ mice and, consistent with the human SMA pathology, pathological changes in the spinal cord and skeletal muscles became apparent. The severity of symptoms strictly correlated with the dosage of transgenic *SMN2*: $Smn^{-/-}$ mice containing few *SMN2* transgenes experienced a short presymptomatic period before developing severe muscular atrophy and dying within days, whereas littermates with more *SMN2* copies survived longer and exhibited a milder phenotype. These findings confirmed the observation in humans that the severity of the SMA-disease phenotype strictly depends on the level of functional SMN protein [8]. The availability of a murine model for SMA also provided insight into disease progression. Surprisingly, even mice that suffer from severe SMA display moderate motor-neuron loss (as determined by histology of these mice a few days after birth) during their presymptomatic period [24]. Two years later, with the improvement of electrophysical methods,

Box 1. Pre-mRNA splicing

Almost every protein-coding gene in humans is interspersed by non-coding sequences, known as introns. Splicing is the post-transcriptional process in which these introns are excised from precursor mRNA molecules (pre-mRNA) and the coding exons are ligated. Introns contain conserved 5′ and 3′ splice sites and a branch site, which need to be positioned and activated for the two consecutive trans-esterification reactions to proceed. In the first of these reactions, a nucleophilic attack by the 2′ hydroxyl group of a branch-site adenosine cleaves the 5′ exon-intron boundary, generating a 3′ hydroxyl group on the upstream exon and an intermediate product that contains the lariat-shaped intron and the downstream exon. In the second reaction, the 3′ hydroxyl group of the 5′ exon attacks the 3′ intron-exon junction, resulting in the ligation of exons and displacement of the lariat-shaped intron. Notably, the efficiency of this basic splicing reaction is often influenced by proteins that bind to regulatory elements within pre-mRNAs [89].

The spliceosome is the macromolecular entity that promotes and controls the splicing reaction. Its functional subunits are uridine-rich small nuclear ribonucleoproteins particles (U snRNPs), which are composed of one (U1, U2, U5) or two (U4/U6) uridine-rich RNAs and characteristic sets of proteins [39]. Together with other non-snRNP proteins [90], these U snRNPs constitute the spliceosome. U snRNPs are key players in splicing because they recognize and activate intronic sites for splicing [89]. The formation of a spliceosome is initiated by the U1 and U2 snRNP, the RNA of which base-pair with the 5′ splice site and the branch site, respectively. Thereafter, the U4/U6 and U5 snRNPs join as a pre-formed U4/U6.U5 tri-snRNP complex to complete the spliceosome. Within this structure, a series of RNA-RNA rearrangements occur, ultimately forming a catalytic RNA network in the spliceosome [91].

Spliceosomes and their U snRNP subunits are formed in a highly ordered process. Seven so-called Sm proteins, which are common to all spliceosomal U snRNPs, are assembled on the U snRNA to form the core structure of these particles [92,93]. These 'Sm core domains' of U snRNPs additionally receive distinct sets of proteins, which confer specific activities to each snRNP [39]. During the past decade, it became evident that the assembly of U snRNP core structures requires assistance by the SMN protein *in vivo*. The reduction of SMN levels, as it occurs in SMA patients, impairs the assembly of U snRNPs [61–64] and inefficient supply of these splicing factors presumably compromises the outcome of splicing.

high motor-neuron counts were also observed in SMA patients before the clinical manifestation of the disease [25]. These studies indicated that motor-neuron degeneration in the spinal cord is a late event in the progression of the disease.

In the past five years, several groups have studied the events preceding the loss of motor neurons in the spinal cord. Frugier *et al.* [26] used transgenic mice to express selectively SMNΔexon7 in motor neurons. These mice displayed only a moderate loss of cell bodies of α-motor neurons in the spinal cord [26], but exhibited misorganized neuromuscular junctions and severe defects in their projecting axons [27]. Studies on motor-neuron cultures derived from the SMA mouse model are in agreement with this phenotype: changes in the survival rate of these cells were minor (compared with wild type), but a strong decrease in axon length was observed [28,29]. The results from these studies have been supported by McWhorter *et al.* [30], who established a morpholino-based model (i.e. an antisense strategy that selectively impairs translation of the target mRNA) for SMA in zebrafish (*Danio rerio*). This approach is unique in that it enables fine adjustment of gene silencing and monitoring of neurons *in situ*. Upon suppression of SMN expression in zebrafish, motor-axon-specific pathfinding defects were clearly present. Similar to motor-neuron cultures from the SMA mouse model, survival of motor neurons in the zebrafish was not dramatically compromised. Together, these data indicate that a high level of SMN protein is crucial (either directly or indirectly) to motor-axon outgrowth. Therefore, a pathfinding defect caused by impaired axonal outgrowth might be an early consequence of SMN deficiency and trigger motor-neuron degeneration in SMA patients and animal models.

SMN has both common and cell-specific functions

The discovery of SMN generated speculation that this gene might have an exclusively neuronal function and that loss of its tissue-specific function elicits the disease. If this were the case, a correlation between function and tissue specificity should be mirrored by the expression profile of SMN. Unexpectedly, however, SMN is ubiquitously expressed, with only few tissue and developmental differences [31,32]. Within a single cell, the protein is present in both the cytoplasm and the nucleus, where it is concentrated in two sub-structures that often overlap, Cajal bodies and Gems [33–36].

When isolated from cultivated cells, a large proportion of cellular SMN is present as part of a multi-subunit macromolecular entity. This SMN complex contains a set of common 'core' proteins, termed Gemins 2-7, and several transient or sub-stoichiometric factors [37,38]. The best characterized sub-stoichiometric factors are the Sm proteins that form part of the U snRNPs, which are involved in pre-mRNA splicing (Box 1). These abundant RNA-protein complexes, termed U1, U2, U4/6 and U5 snRNPs, are present in the nuclei of all eukaryotic cells. They bind, together with many additional proteins, to introns of pre-mRNA to form the spliceosome (the macromolecular structure in which the splicing reaction is catalyzed). Each U snRNP consists of one (U1, U2 or U5) or two (U4 and U6) small nuclear RNAs (snRNAs) associated with a set of seven Sm proteins (known as B/B′, D1, D2, D3, E, F and G) and several other specific factors. A key step in the assembly of these particles is the nuclear export of uridine-rich snRNAs (U snRNAs) and their association with Sm proteins (and other proteins), before the assembled U snRNP is imported back into the nucleus [39]. Several studies have demonstrated that the SMN complex is necessary and sufficient by itself to mediate transfer of Sm proteins into the U snRNP [40–42]. However, several groups have shown that this reaction is greatly enhanced by a second macromolecule, the protein arginine methyltransferase 5 (PRMT5) complex (also known as the 'methylosome') [43,44]. Thus, the SMN complex and the PRMT5 complex cooperate in the formation of spliceosomal U snRNPs [37,38,41,42,45] (Figure 2a). Intriguingly, the SMN complex is also able

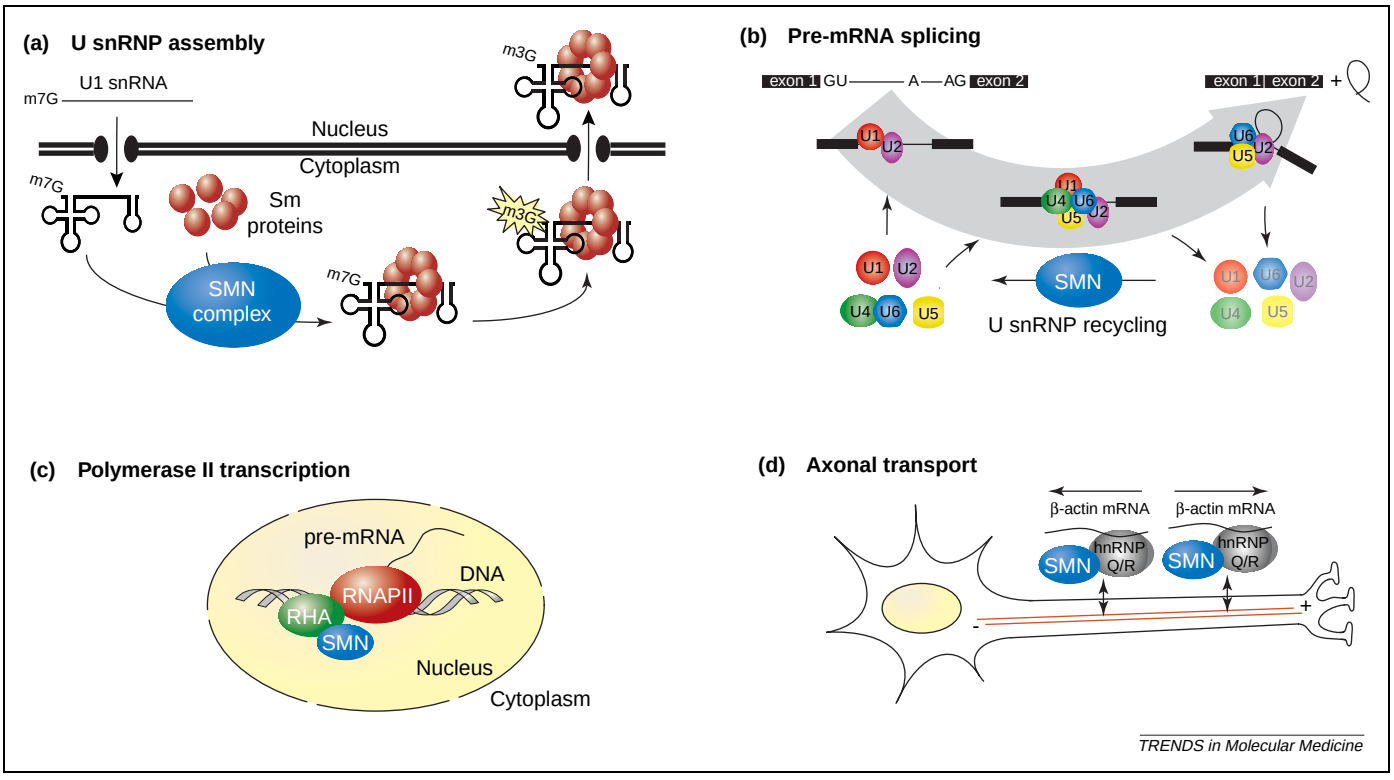


Figure 2. Proposed cellular functions of SMN. **(a)** Assembly pathway for U1 snRNP; U2, U4 and U5 snRNPs use a similar pathway. SMN (as part of the SMN complex) assists the formation of spliceosomal U snRNPs. Initially, the U snRNA that is transcribed in the nucleus is transported to the cytoplasm where it assembles with Sm proteins to form the U snRNP particle. After modification of the m^7G-cap into m_3G-cap, the mature particle is transported back to the nucleus. **(b)** SMN participates in the splicing reaction, probably during the process of spliceosome regeneration. **(c)** SMN interacts with RNA helicase A (RHA) and stimulates RNA polymerase II (RNAPII) transcription. **(d)** SMN binds via hnRNP Q/R to the β-actin mRNA, which is actively transported within motor axons. Along with the visualization of SMN movement in motor axons, a role of SMN in axonal transport has been proposed.

to promote the maturation of other RNPs such as the U7 snRNP, the function of which relates to histone-mRNA processing rather than splicing. These findings assign the SMN complex with the potential role of 'master RNP assembler' [46,47].

Based on biochemical evidence, at least two additional housekeeping functions of SMN have been postulated. One is a link between SMN and the catalytic steps of pre-mRNA splicing, in which SMN is proposed to function as a recycling factor that regenerates U snRNPs after splicing catalysis [48–50] (Figure 2b). Other studies have revealed an association of SMN with RNA helicase A (RHA) and RNA polymerase II (RNAPII), raising the possibility that SMN has a role in gene expression [51–53] (Figure 2c). In support of this hypothesis, the overexpression of a dominant SMN mutant leads to the inhibition of transcription *in vivo*, whereas wild-type SMN leads to stimulation of transcription [51]. Furthermore, Gemin3, a known component of the SMN complex, has also been linked to regulation of transcription of certain reporter genes by RNAPII [53–55]. However, the mechanism by which SMN and its interacting partners act in these processes remains unknown.

In addition to the reported housekeeping functions of SMN, a neuron-specific role in axonal mRNA transport has been proposed (Figure 2d). The first suggestion of this activity came from the finding that SMN interacts with hnRNP protein R, a protein that binds to the 3′ untranslated region (UTR) of β-actin mRNA [50,56]. Rossoll *et al.* [29] suggested that the expression of either *Smn* or *hnRNP R* transgenes in the rat pheochromocytoma cell line PC12 enhances the differentiation of these cells and their axonal outgrowth [29]. In conjunction with the finding that motor neurons that are isolated from the SMA mouse model show reduced levels of β-actin mRNA at distal axons and growth cones, the authors suggested that Smn and hnRNP R modulate axonal mRNA transport [29]. One prerequisite for this model is the movement of SMN along axons, which has been observed by Bassell *et al.* [57]. Another study has revealed an interaction of SMN with profilin II [58], which is a neuron-specific factor required for actin polymerization [59]. Therefore, it might be speculated that SMN, hnRNP R and profilin II are all involved in axonal trafficking. Clearly, further biochemical and functional characterization of SMN interactions within motor neurons will be needed to understand the neuronal function(s) of SMN fully.

Molecular defects in SMN-deficient cells

The variety of general and neuron-specific functions assigned to the SMN protein raises the question of which of these activities is responsible for the motor-neuron defects in SMA. An obvious prediction is that SMA results from impairment of a neuron-specific function of SMN, such as the proposed SMN-mediated axonal transport of β-actin mRNA [29,60]. If defects in axonal outgrowth are indeed the consequence of inefficient β-actin mRNA transport, a functional correlation with SMN might explain the axonopathy (i.e. the degeneration of motor axons into short and abnormally branched forms,

impairing proper innervation of muscles) that is a typical feature of SMA. Similarly, the interaction between SMN and profilin II seems crucial for actin polymerization in axons [59]. Although both scenarios are attractive and deserve careful consideration, further experimental evidence is needed to prove that mRNA transport or actin polymerization are linked directly to reduced levels of SMN.

In contrast to the aforementioned model, the idea that impairment of a general function of SMN might cause SMA has been supported experimentally. Using RNA-interference in non-neuronal cell cultures, several groups have established a direct correlation between the level of SMN and the ability of these cells to promote RNP assembly [61–64]. Because this phenomenon was also observed in fibroblast cultures from SMA patients [61], it is likely that impaired assembly of U snRNP is a general biochemical feature of SMA. Using zebrafish as a model for SMA [30], Winkler *et al.* [62] supported this hypothesis with two lines of evidence. First, the silencing of Gemin2 or pICln (which are both components of the SMN complex or the PRMT5 complex, respectively) conferred the same phenotype as that obtained by knockdown of SMN expression. Notably, the sensitivity of motor neurons to low levels of SMN or Gemin2 was also demonstrated in $Smn^{+/-}$ mice in which Gemin2 levels were reduced [65]. Second, the motor-axon phenotype of zebrafish that lack either SMN or Gemin2 can be prevented by addition of purified spliceosomal U snRNPs (i.e. the end product of the SMN-assembly pathway). These studies provided first experimental support for the idea that reduced assembly of U snRNPs contributes directly to SMA pathogenesis.

Defects in basic cellular processes as a cause for disease: a model

The causative relationship of disrupted assembly of U snRNPs and SMA echoes the initial question of this review: how do the basic functions of every cell relate to tissue-specific pathology? Although this question cannot be answered with certainty at present, it is reasonable to believe that reduced assembly of U snRNPs decreases the quantity of active spliceosomes (the entities that excise introns from pre-mRNA). If this is the case, what causes the difference between motor neurons and all the other cell types that are not, or at least less severely, affected? In pursuit of an answer, the metabolism in motor neurons should be considered: motor neurons exhibit a high turnover of mRNAs that encode proteins with tissue-specific activities (e.g. in axonal pathfinding) [66,67]. Therefore, it is possible that motor neurons cannot meet the demand for such proteins if splicing of their pre-mRNAs is inefficient (Figure 3). The hypothesis proposed by Faustino and Cooper [3] assumes that some pre-mRNAs of tissue-specific genes might be spliced less efficiently than pre-mRNAs of constitutive genes. Thus, a group of tissue-specific mRNAs (or even a single mRNA) that contains sub-optimal splice sites might be more sensitive to reduced levels of U snRNPs than other mRNAs. Reduced levels of the proteins involved in the biogenesis of U snRNPs - as observed in SMA - might thus evoke a specific phenotype, even though it serves a general cellular function. This scenario is supported by studies on the molecular basis of RP, which has many independent genetic causes (http://www.retnet.org). Although most pathogenic mutations occur in genes that are linked to photoreception, during the past four years, RP has also been linked to mutations in three essential components of the U4/U6.U5 tri-snRNP (i.e. the superstructure within which the U4/U6 and the U5 snRNP enter the spliceosome) [1]. The autosomal dominant mode of inheritance supports the possibility that, at least in some cases, the outcome of these mutations is haplo-insufficiency (i.e. functional inactivation of one allele). Assuming that these factors have no other function (for which no evidence exists), the most likely consequence of mutations in these loci is a weakening of the splicing machinery. Clearly, future studies aimed at the identification of pre-mRNAs with splicing patterns that are offset will substantially contribute to the understanding of both SMA and RP.

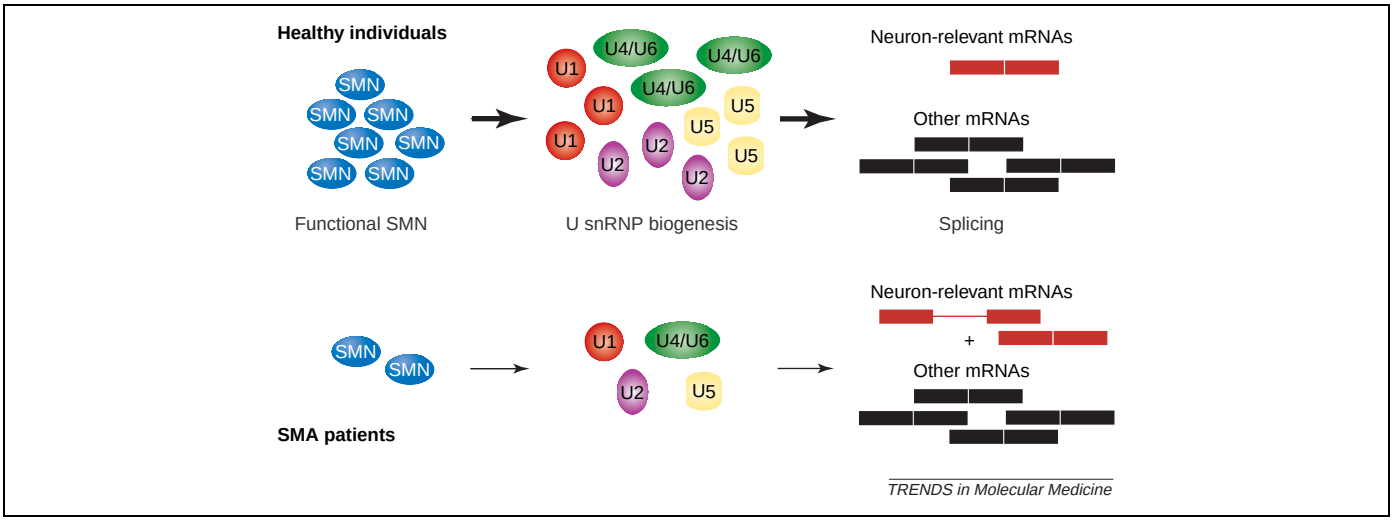


Figure 3. A possible scenario of the events that lead to SMA. SMN deficiency results in reduced production of spliceosomal U snRNPs. As a consequence, processing of mRNAs with sub-optimal splice sites (e.g. tissue-specific transcripts) would be compromised. Therefore, inefficient splicing of neuron-relevant mRNAs might be the basis for the tissue-specific phenotype in SMA.

Towards a treatment of SMA

The knowledge gained about the genetic and biochemical background of SMA has revealed many new leads for potential treatments of this disease (Figure 4). Because motor neurons are primarily affected, a possibility is to protect these cells from degeneration by administration of either neuroprotective or neurotrophic agents [68,69] (Figure 4f). These so-called 'non-targeted' approaches seem promising; indeed, Riluzole, a neurotrophic agent, can attenuate disease progression in a mouse model [68] and a phase I clinical trial of this agent has recently been concluded [70].

Equally promising are 'targeted' approaches that are designed to enhance the production of functional SMN protein in motor neurons (Figure 4a-d). For example, the lentiviral-gene-transfer approach developed by Azzouz *et al.* [71], which successfully amended SMN levels in fibroblasts in SMA type I patients (Figure 4a): when it is injected into different muscles of SMA mice, the lenti-vector restores SMN levels in motor neurons, thus reducing motor-neuron death and increasing the life expectancy of these animals by 3-5 days [71].

In addition to gene-transfer strategies, the existence of a second copy of the disease gene (*SMN2*) offers an excellent target for potential therapeutics [72]. Transcriptional upregulation of the *SMN*2 gene (Figure 4b) seems to be the most obvious strategy, and high-throughput screens designed to identify small molecules with this activity have been performed. Histone-deacetylase (HDAC) inhibitors were detected in these screens [73–79], which is not surprising because they are known to enhance transcription of ~2% of all cellular genes. Two HDAC inhibitors, valproic acid and 4-phenylbutyrate, have already been approved for the treatment of other diseases, such as epilepsy, solid tumors and hematological malignancies, and are currently in clinical trials for the treatment of SMA patients [74,80,81]. Interestingly, although initially identified as transcriptional activators, some HDAC inhibitors also affect splicing by enhancing exon-7 inclusion [73–76] (Figure 4c). An alternative way to increase SMN levels might be related to the presence of a candidate interferon-stimulated response element (ISRE) in the promoter region of both *SMN1* and *SMN2*, implicating interferon as a regulator of their transcription [82]. Indeed, the authors showed that the administration of interferons leads to upregulated *SMN* transcription in various cell types. It will be interesting to see whether this promising therapeutic strategy finds proof in animal studies and, ultimately, in patients.

A potential caveat of all the strategies discussed is that they aim solely at enhancing the transcription of the inherited genes, which means that the majority of *SMN* transcripts are either mutant *SMN1* transcripts or the predominant form of *SMN2* transcripts (SMNΔexon7). Therefore, an important task will be careful evaluation of potential adverse effects induced by overproduction of these mRNAs. Encouragingly, excess SMNΔexon7 protein is not toxic when is associated with full-length SMN in mice [83].

An alternative approach to transcriptional upregulation might be to shift *SMN2* splicing towards a higher rate

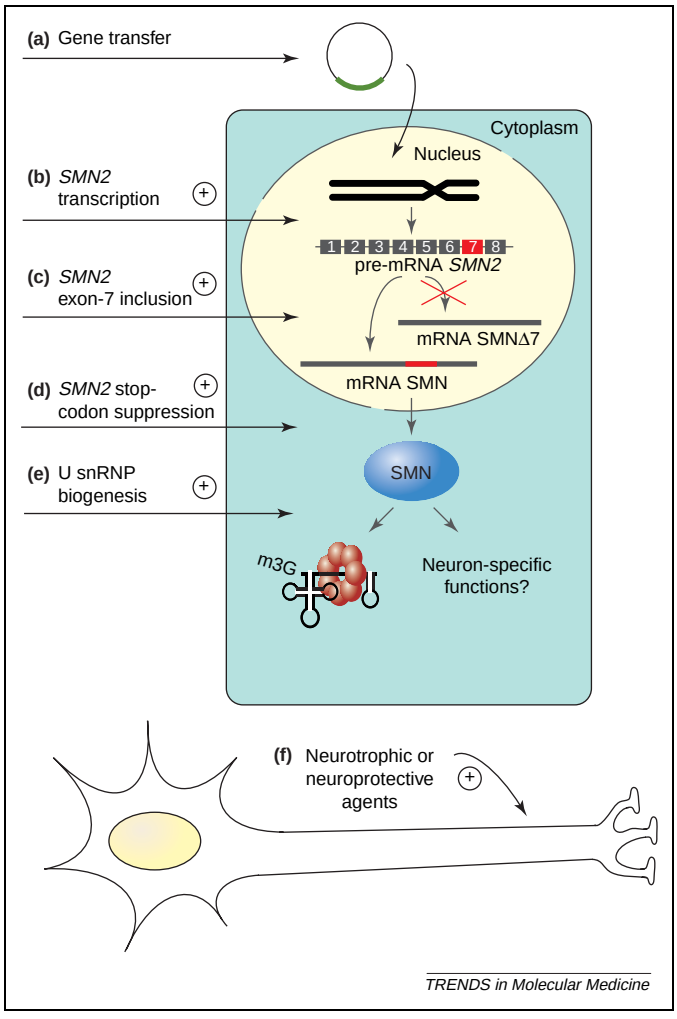


Figure 4. Potential therapeutic strategies for a treatment of SMA. **(a)** Gene therapy: transfer of genes that prevent motor-neuron degeneration (e.g. *SMN1*). **(b)** Enhancement of *SMN2* transcription. **(c)** Modulation of *SMN2* splicing. **(d)** Suppression of the use of SMN2 stop codon. **(e)** Stimulation of SMN function in U snRNP biogenesis. **(f)** Use of neurotrophic or neuroprotective agents.

of exon-7 inclusion (Figure 4c). Several groups have employed oligonucleotides that hybridize with exon 7 and that also contain either ESE sequences or short peptides that are designed to recruit positively acting splicing factors [84–86]. Indeed, these modified nucleic acids were potent activators of exon-7 inclusion in splicing assays *in vitro* and in cell culture. However, it remains to be established how these molecules can be targeted to the tissue affected in SMA patients. In an attempt to find a non-canonical therapeutic approach, Wolstencroft *et al.* [87] employed aminoglycosides (i.e. sugar derivatives containing amino groups) to modulate translational fidelity (Figure 4d). The aminoglycosides suppressed the use of the stop codon in *SMN2*, leading to an extended product with increased physiological half life. This remarkable effect now awaits validation in animal models of SMA.

Improved understanding of the molecular basis of SMA also extends the scope for therapy: finding ways to improve the process of U snRNP assembly would be one possibility (Figure 4e), for example, by speeding up the activity of the remaining functional SMN protein in

Box 2. Future goals in the study of SMA

- Understanding SMN function in motor neurons.
- Identification of mRNAs for which splicing is compromised in SMA.
- Analysis of signal-transduction pathways that affect the activity of SMN.
- Delineation of a feasible therapeutic strategy (agent or combination of agents and appropriate time point of administration).

patients. The activity of SMN and the SMN complex are regulated by phosphorylation [88] and are temporally controlled with regard to the developmental and differentiation state of tissues [31]. Therefore, it might be useful to search for the elusive kinases and phosphatases that are involved and to screen for compounds that act on them. Furthermore, the identity of mRNAs with splicing patterns that are primarily offset because of the reduction of SMN levels will give important insight into how the weakening of the splicing machinery translates into specific degeneration of motor neurons.

Concluding remarks

At present, SMA is one of the best-understood neurodegenerative diseases, at both the genetic and the biochemical level. The molecular basis of SMA, caused by reduced synthesis of the SMN protein, currently remains a matter of debate particularly because of its many proposed cellular functions. The identification of a biochemical pathway affected in SMA - the assembly of U snRNPs - and its correlation with motor-axon defects provides first insight into the molecular consequences of SMN deficiency. The next step is analysis of whether other known functions of SMN are also affected in SMA patients and whether they contribute to the disease phenotype. The advent of therapeutic strategies awaits a detailed analysis of the optimal time point for intervention. In particular, the question of whether treatment is possible in a postsymptomatic phase of the disease requires careful consideration (Box 2). With increasing knowledge about the molecular basis of SMA and approaches from various disciplines, new ideas for therapy can be expected.

Acknowledgements

We thank M. Sendtner and A.R. Krainer for helpful discussions on the manuscript. We apologize to colleagues whose work could not be discussed or cited owing to space limitations.

References

1 Kennan, A. *et al*. (2005) Light in retinitis pigmentosa. *Trends Genet.* 21, 103–110
2 Sendtner, M. (2001) Molecular mechanisms in spinal muscular atrophy: models and perspectives. *Curr. Opin. Neurol.* 14, 629–634
3 Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437
4 Schmalbruch, H. and Haase, G. (2001) Spinal muscular atrophy: present state. *Brain Pathol.* 11, 231–247
5 Brahe, C. and Bertini, E. (1996) Spinal muscular atrophies: recent insights and impact on molecular diagnosis. *J. Mol. Med.* 74, 555–562
6 Lefebvre, S. *et al*. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155–165
7 Lorson, C.L. and Androphy, E.J. (2000) An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene *SMN*. *Hum. Mol. Genet.* 9, 259–265
8 Lefebvre, S. *et al*. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nat. Genet.* 16, 265–269
9 Schrank, B. *et al*. (1997) Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9920–9925
10 Lorson, C.L. *et al*. (1999) A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6307–6311
11 Monani, U.R. *et al*. (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene *SMN1* from the copy gene *SMN2*. *Hum. Mol. Genet.* 8, 1177–1183
12 Kashima, T. and Manley, J.L. (2003) A negative element in *SMN2* exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* 34, 460–463
13 Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in *SMN2* causes spinal muscular atrophy in the absence of *SMN1*. *Nat. Genet.* 30, 377–384
14 Hofmann, Y. *et al*. (2000) Htra2-β1 stimulates an exonic splicing enhancer and can restore full-length *SMN* expression to survival motor neuron 2 (*SMN2*). *Proc. Natl. Acad. Sci. U. S. A.* 97, 9618–9623
15 Young, P.J. *et al*. (2002) SRp30c-dependent stimulation of survival motor neuron (*SMN*) exon 7 inclusion is facilitated by a direct interaction with hTra2β1. *Hum. Mol. Genet.* 11, 577–587
16 Cartegni, L. *et al*. (2006) Determinants of exon 7 identity in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* 78, 63–77
17 Hannus, S. *et al*. (2000) The *Schizosaccharomyces pombe* protein Yab8p and a novel factor, Yip1p, share structural and functional similarity with the spinal muscular atrophy-associated proteins SMN and SIP1. *Hum. Mol. Genet.* 9, 663–674
18 Paushkin, S. *et al*. (2000) The survival motor neuron protein of *Schizosacharomyces pombe*. Conservation of survival motor neuron interaction domains in divergent organisms. *J. Biol. Chem.* 275, 23841–23846
19 Miguel-Aliaga, I. *et al*. (2000) Disruption of SMN function by ectopic expression of the human *SMN* gene in *Drosophila*. *FEBS Lett.* 486, 99–102
20 Miguel-Aliaga, I. *et al*. (1999) The *Caenorhabditis elegans* orthologue of the human gene responsible for spinal muscular atrophy is a maternal product critical for germline maturation and embryonic viability. *Hum. Mol. Genet.* 8, 2133–2143
21 Hsieh-Li, H.M. *et al*. (2000) A mouse model for spinal muscular atrophy. *Nat. Genet.* 24, 66–70
22 Jablonka, S. *et al*. (2000) Reduced survival motor neuron (*Smn*) gene dose in mice leads to motor neuron degeneration: an animal model for spinal muscular atrophy type III. *Hum. Mol. Genet.* 9, 341–346
23 Monani, U.R. *et al*. (2000) Animal models of spinal muscular atrophy. *Hum. Mol. Genet.* 9, 2451–2457
24 Monani, U.R. *et al*. (2000) The human centromeric survival motor neuron gene (*SMN2*) rescues embryonic lethality in $Smn^{-/-}$ mice and results in a mouse with spinal muscular atrophy. *Hum. Mol. Genet.* 9, 333–339
25 Bromberg, M.B. and Swoboda, K.J. (2002) Motor unit number estimation in infants and children with spinal muscular atrophy. *Muscle Nerve* 25, 445–447
26 Frugier, T. *et al*. (2000) Nuclear targeting defect of SMN lacking the C-terminus in a mouse model of spinal muscular atrophy. *Hum. Mol. Genet.* 9, 849–858
27 Cifuentes-Diaz, C. *et al*. (2001) Deletion of murine *SMN* exon 7 directed to skeletal muscle leads to severe muscular dystrophy. *J. Cell Biol.* 152, 1107–1114
28 Jablonka, S. *et al*. (2004) Axonal defects in mouse models of motoneuron disease. *J. Neurobiol.* 58, 272–286
29 Rossoll, W. *et al*. (2003) Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of β-actin mRNA in growth cones of motoneurons. *J. Cell Biol.* 163, 801–812
30 McWhorter, M.L. *et al*. (2003) Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J. Cell Biol.* 162, 919–931

31 Gabanella, F. *et al*. (2005) The activity of the spinal muscular atrophy protein is regulated during development and cellular differentiation. *Hum. Mol. Genet.* 14, 3629–3642
32 Jablonka, S. *et al*. (2001) Co-regulation of survival of motor neuron (SMN) protein and its interactor SIP1 during development and in spinal muscular atrophy. *Hum. Mol. Genet.* 10, 497–505
33 Liu, Q. and Dreyfuss, G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.* 15, 3555–3565
34 Carvalho, T. *et al*. (1999) The spinal muscular atrophy disease gene product, SMN: a link between snRNP biogenesis and the Cajal (coiled) body. *J. Cell Biol.* 147, 715–728
35 Matera, A.G. (2003) Cajal bodies. *Curr. Biol.* 13, R503
36 Malatesta, M. *et al*. (2004) Ultrastructural characterisation of a nuclear domain highly enriched in survival of motor neuron (SMN) protein. *Exp. Cell Res.* 292, 312–321
37 Meister, G. *et al*. (2002) SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol.* 12, 472–478
38 Gubitz, A.K. *et al*. (2004) The SMN complex. *Exp. Cell Res.* 296, 51–56
39 Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* 13, 290–301
40 Meister, G. *et al*. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol.* 3, 945–949
41 Pellizzoni, L. *et al*. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science* 298, 1775–1779
42 Meister, G. and Fischer, U. (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J.* 21, 5853–5863
43 Meister, G. *et al*. (2001) Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* 11, 1990–1994
44 Friesen, W.J. *et al*. (2001) The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* 21, 8289–8300
45 Narayanan, U. *et al*. (2004) Coupled *in vitro* import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell* 16, 223–234
46 Pillai, R.S. *et al*. (2003) Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev.* 17, 2321–2333
47 Terns, M.P. and Terns, R.M. (2001) Macromolecular complexes: SMN – the master assembler. *Curr. Biol.* 11, R862–R864
48 Meister, G. *et al*. (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
49 Pellizzoni, L. *et al*. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* 95, 615–624
50 Mourelatos, Z. *et al*. (2001) SMN interacts with a novel family of hnRNP and spliceosomal proteins. *EMBO J.* 20, 5443–5452
51 Pellizzoni, L. *et al*. (2001) A functional interaction between the survival motor neuron complex and RNA polymerase II. *J. Cell Biol.* 152, 75–85
52 Strasswimmer, J. *et al*. (1999) Identification of survival motor neuron as a transcriptional activator-binding protein. *Hum. Mol. Genet.* 8, 1219–1226
53 Voss, M.D. *et al*. (2001) Functional cooperation of Epstein–Barr virus nuclear antigen 2 and the survival motor neuron protein in transactivation of the viral LMP1 promoter. *J. Virol.* 75, 11781–11790
54 Campbell, L. *et al*. (2000) Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? *Hum. Mol. Genet.* 9, 1093–1100
55 Yan, X. *et al*. (2003) A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. *Mol. Cell. Biol.* 23, 414–423
56 Rossoll, W. *et al*. (2002) Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? *Hum. Mol. Genet.* 11, 93–105
57 Zhang, H.L. *et al*. (2003) Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. *J. Neurosci.* 23, 6627–6637
58 Giesemann, T. *et al*. (1999) A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with SMN in nuclear gems. *J. Biol. Chem.* 274, 37908–37914
59 Sharma, A. *et al*. (2005) A role for complexes of survival of motor neurons (SMN) protein with gemins and profilin in neurite-like cytoplasmic extensions of cultured nerve cells. *Exp. Cell Res.* 309, 185–197
60 Briese, M. *et al*. (2005) Is spinal muscular atrophy the result of defects in motor neuron processes? *Bioessays* 27, 946–957
61 Wan, L. *et al*. (2005) The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy. *Mol. Cell. Biol.* 25, 5543–5551
62 Winkler, C. *et al*. (2005) Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev.* 19, 2320–2330
63 Shpargel, K.B. and Matera, A.G. (2005) Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17372–17377
64 Feng, W. *et al*. (2005) Gemins modulate the expression and activity of the SMN complex. *Hum. Mol. Genet.* 14, 1605–1611
65 Jablonka, S. *et al*. (2002) Gene targeting of Gemin2 in mice reveals a correlation between defects in the biogenesis of U snRNPs and motoneuron cell death. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10126–10131
66 Chisholm, A. and Tessier-Lavigne, M. (1999) Conservation and divergence of axon guidance mechanisms. *Curr. Opin. Neurobiol.* 9, 603–615
67 Tear, G. (1999) Axon guidance at the central nervous system midline. *Cell. Mol. Life Sci.* 55, 1365–1376
68 Haddad, H. *et al*. (2003) Riluzole attenuates spinal muscular atrophy disease progression in a mouse model. *Muscle Nerve* 28, 432–437
69 Robertson, G.S. *et al*. (2000) Neuroprotection by the inhibition of apoptosis. *Brain Pathol.* 10, 283–292
70 Russman, B.S. *et al*. (2003) A phase 1 trial of riluzole in spinal muscular atrophy. *Arch. Neurol.* 60, 1601–1603
71 Azzouz, M. *et al*. (2004) Lentivector-mediated SMN replacement in a mouse model of spinal muscular atrophy. *J. Clin. Invest.* 114, 1726–1731
72 Lunn, M.R. and Stockwell, B.R. (2005) Chemical genetics and orphan genetic diseases. *Chem. Biol.* 12, 1063–1073
73 Chang, J.G. *et al*. (2001) Treatment of spinal muscular atrophy by sodium butyrate. *Proc. Natl. Acad. Sci. U. S. A.* 98, 9808–9813
74 Sumner, C.J. *et al*. (2003) Valproic acid increases SMN levels in spinal muscular atrophy patient cells. *Ann. Neurol.* 54, 647–654
75 Brichta, L. *et al*. (2003) Valproic acid increases the SMN2 protein level: a well-known drug as a potential therapy for spinal muscular atrophy. *Hum. Mol. Genet.* 12, 2481–2489
76 Andreassi, C. *et al*. (2004) Phenylbutyrate increases SMN expression *in vitro*: relevance for treatment of spinal muscular atrophy. *Eur. J. Hum. Genet.* 12, 59–65
77 Andreassi, C. *et al*. (2001) Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. *Hum. Mol. Genet.* 10, 2841–2849
78 Jarecki, J. *et al*. (2005) Diverse small-molecule modulators of SMN expression found by high-throughput compound screening: early leads towards a therapeutic for spinal muscular atrophy. *Hum. Mol. Genet.* 14, 2003–2018
79 Lunn, M.R. *et al*. (2004) Indoprofen upregulates the survival motor neuron protein through a cyclooxygenase-independent mechanism. *Chem. Biol.* 11, 1489–1493
80 Mercuri, E. *et al*. (2004) Pilot trial of phenylbutyrate in spinal muscular atrophy. *Neuromuscul. Disord.* 14, 130–135
81 Brahe, C. *et al*. (2005) Phenylbutyrate increases *SMN* gene expression in spinal muscular atrophy patients. *Eur. J. Hum. Genet.* 13, 256–259
82 Baron-Delage, S. *et al*. (2000) Interferons and IRF-1 induce expression of the survival motor neuron (*SMN*) genes. *Mol. Med.* 6, 957–968
83 Le, T.T. *et al*. (2005) SMNΔ7, the major product of the centromeric survival motor neuron (*SMN2*) gene, extends survival in mice with spinal muscular atrophy and associates with full-length SMN. *Hum. Mol. Genet.* 14, 845–857
84 Skordis, L.A. *et al*. (2003) Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates *SMN2* gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4114–4119

85 Cartegni, L. and Krainer, A.R. (2003) Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat. Struct. Biol.* 10, 120–125
86 Madocsai, C. *et al.* (2005) Correction of *SMN2* Pre-mRNA splicing by antisense U7 small nuclear RNAs. *Mol. Ther.* 12, 1013–1-22
87 Wolstencroft, E.C. *et al.* (2005) A non-sequence-specific requirement for SMN protein activity: the role of aminoglycosides in inducing elevated SMN protein levels. *Hum. Mol. Genet.* 14, 1199–1210
88 Grimmler, M. *et al.* (2005) Phosphorylation regulates the activity of the SMN complex during assembly of spliceosomal U snRNPs. *EMBO Rep.* 6, 70–76
89 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106–110
90 Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12, 5–14
91 Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* 36, 333–360
92 Kambach, C. *et al.* (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* 96, 375–387
93 Kiss, T. (2004) Biogenesis of small nuclear RNPs. *J. Cell Sci.* 117, 5949–5951

Elsevier celebrates two anniversaries with gift to university libraries in the developing world

In 1580, the Elzevir family began their printing and bookselling business in the Netherlands, publishing works by scholars such as John Locke, Galileo Galilei and Hugo Grotius. On 4 March 1880, Jacobus George Robbers founded the modern Elsevier company intending, just like the original Elzevir family, to reproduce fine editions of literary classics for the edification of others who shared his passion, other 'Elzevirians'. Robbers co-opted the Elzevir family's old printer's mark, visually stamping the new Elsevier products with a classic old symbol of the symbiotic relationship between publisher and scholar. Elsevier has since become a leader in the dissemination of scientific, technical and medical (STM) information, building a reputation for excellence in publishing, new product innovation and commitment to its STM communities.

In celebration of the House of Elzevir's 425th anniversary and the 125th anniversary of the modern Elsevier company, Elsevier will donate books to 10 university libraries in the developing world. Entitled 'A Book in Your Name', each of the 6 700 Elsevier employees worldwide has been invited to select one of the chosen libraries to receive a book donated by Elsevier. The core gift collection contains the company's most important and widely used STM publications including *Gray's Anatomy, Dorland's Illustrated Medical Dictionary, Essential Medical Physiology, Cecil Essentials of Medicine, Mosby's Medical, Nursing and Allied Health Dictionary, The Vaccine Book, Fundamentals of Neuroscience, and Myles Textbook for Midwives.*

The 10 beneficiary libraries are located in Africa, South America and Asia. They include the Library of the Sciences of the University of Sierra Leone; the library of the Muhimbili University College of Health Sciences of the University of Dar es Salaam, Tanzania; the library of the College of Medicine of the University of Malawi; and the libraries of the University of Zambia, Universite du Mali, Universidade Eduardo Mondlane, Mozambique; Makerere University, Uganda; Universidad San Francisco de Quito, Ecuador; Universidad Francisco Marroquin, Guatemala; and the National Centre for Scientific and Technological Information (NACESTI), Vietnam.

Through 'A Book in Your Name', the 10 libraries will receive approximately 700 books at a retail value of approximately 1 million US dollars.

For more information, visit www.elsevier.com

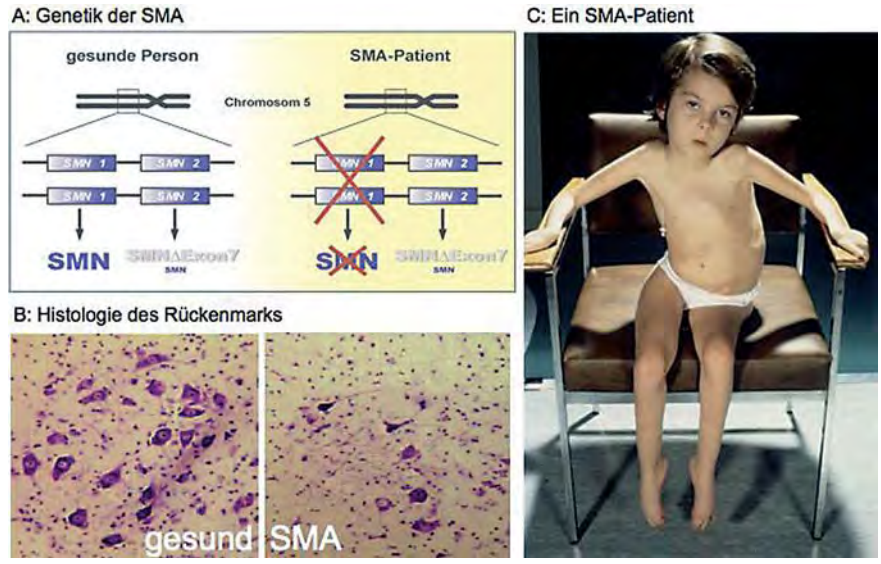
Molekulare Medizin

Wenn Muskeln die Nerven verlieren

UTZ FISCHER, MATTHIAS KROISS, ASHWIN CHARI
THEODOR-BOVERI-INSTITUT AM BIOZENTRUM, UNIVERSITÄT WÜRZBURG

In Deutschland werden jährlich ca. 300 Kinder mit einer oft fatalen neuronalen Krankheit, der Spinalen Muskelatrophie (SMA), geboren. Klinisch fallen die Patienten mit einer schlaffen Lähmung der Extremitätenmuskulatur auf, in schweren Fällen lernen sie weder sitzen noch laufen. Lebensbedrohlich wird die autosomal rezessive Erkrankung durch die begleitende Lähmung der Atemmuskulatur, die schwere Atemwegsinfektionen zur Folge hat. Daran versterben die Patienten häufig bereits im Kleinkindalter.

Pathologisches Korrelat der Krankheit ist die Degeneration der Motoneuronen, die für die Innervation der Muskulatur zuständig sind. Obwohl es durchaus unterschiedliche Schweregrade der Krankheit gibt, sind nahezu alle SMA-Fälle durch homozygote Deletionen oder Mutationen in einem einzigen Gen, dem Survival Motor Neuron (SMN)-Gen I, verursacht[1] (**Abb. 1**). Im Gegensatz zu anderen Organismen wie z. B. Mäusen besitzt der Mensch eine nahezu identische zweite Kopie des SMN-Gens (SMNII-Gen). Die Transkripte dieses Gens werden jedoch aufgrund einer stillen Punktmutation anders prozessiert als jene des SMNI-Gens: Ein Protein-kodierender Abschnitt der prä-mRNA, das Exon 7, wird übersprungen (Exon-Skipping)[2]. Die Folge ist fatal: Von SMNII wird nur eine geringe Menge funktionsfähiges SMN-Protein hergestellt. Daher kann ein Defekt des SMNI-Gens nicht durch das SMNII-Gen kompensiert werden. Die Spinale Muskelatrophie kann man daher auch als SMN-Mangelkrankheit bezeichnen.

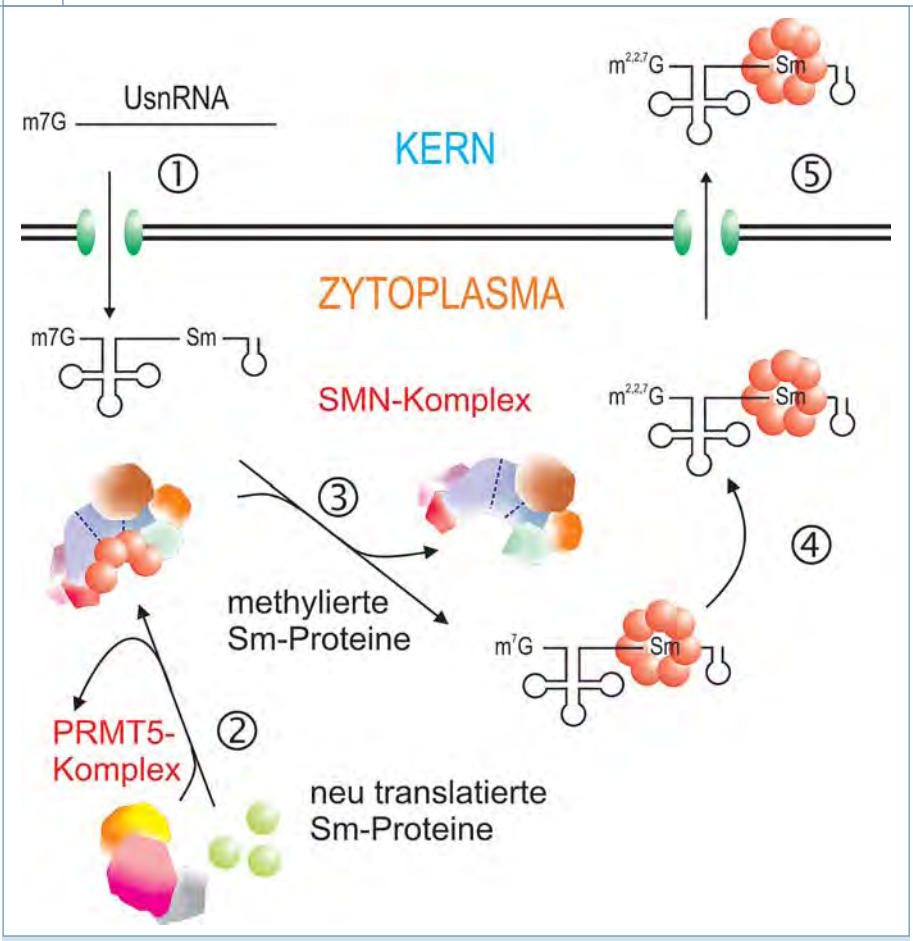


▲ **Abb. 1:** Die Spinale Muskelatrophie ist eine neuromuskuläre genetische Krankheit. **A,** Die homozygote Deletion oder Mutation der SMNII-Genkopie führt zur SMA. **B,** Die reduzierte Anzahl von Motoneuronen im Rückenmark (violette Färbung) ist ein pathologisches Merkmal der SMA. Gezeigt sind histologische Schnitte durch das Rückenmark des SMA-Mausmodells. **C,** Ein SMA-Patient. Die Fotos wurden freundlicherweise von Frau Prof. Rudnik-Schöneborn (C) und Herrn Prof. Sendtner (B) zur Verfügung gestellt.

Die Aufgaben von SMN

Aufgrund der oben skizzierten genetischen Situation war es frühzeitig klar, dass ein molekulares Verständnis der SMA nur über die Analyse der zellulären Funktion von SMN führen würde. Da bei SMA-Patienten ausschließlich die Motoneuronen betroffen sind, lag die Vermutung nahe, dass SMN eine hochspezifische Aufgabe in diesen Zellen erfüllt. Der Verlust dieser Funktion führt dann zur Krankheit. Überraschenderweise fand man aber, dass das SMN-Protein ubiquitär exprimiert wird. Es gilt daher heute als gesichert, dass SMNI ein typisches „Haushaltsgen" darstellt, das allgemeine zelluläre Funktionen ausübt. Obwohl auch Zelltyp-spezifische Funktionen beschrieben wurden, konnte man das SMN-Protein mit einer Reihe von ganz allgemeinen Vorgängen in Verbindung bringen. Die am besten auf molekularer Ebene verstandene Funktion besteht in der Zusammenlagerung von RNA-Proteinkomplexen des Spleißosoms (U snRNPs). In der Tat gibt es gute Gründe anzunehmen, dass diese Aktivität auch entscheidend zur Pathogenese der SMA beiträgt.

Das Spleißosom ist eine makromolekulare Maschine von unglaublicher Komplexität und Dynamik: Seine wichtigsten Bestandteile sind vier unterschiedliche U snRNPs, welche ihrerseits aus einer namensgebenden U snRNA-Komponente (U1, U2, U4/6 und U5) und mehreren Proteinen bestehen. Diese U snRNPs lagern sich an nicht-kodierende Bereiche (Introns) einer prä-mRNA an und wirken zusammen, um die kodierenden Bereiche (Exons) unter gleichzeitiger Entfernung der Introns zusammenzufügen. Dies ist einer der wichtigen Schritte bei der Realisierung des genetischen Kodes. Allerdings müssen die einzelnen U snRNPs ihrerseits in der Zelle in einem komplizierten und mehrstufigen Prozess hergestellt werden (**Abb. 2**). Die transkribierten U snRNAs werden dazu aus dem Zellkern ins Zytoplasma transportiert, wo sieben Sm-Proteine an einen bestimmten Teil der U snRNA binden. Es entsteht eine außerordentlich kompakte und stabile ringförmige Struktur, die das strukturelle Grundgerüst aller U snRNPs bildet. Man bezeichnet sie als „Sm-core"-Domäne. Nach weiteren Reifungsschritten gelangt das Partikel dann in den



▲ **Abb. 2:** Die Zusammenlagerung von spleißosomalen U snRNPs erfolgt durch die koordinierte Aktivität zweier Proteinkomplexe. Nach Transkription der U snRNA im Zellkern (1) erfolgt ihr Export in das Zytoplasma. Dort translatierte Sm-Proteine (grün) werden durch den PRMT5-Komplex symmetrisch an Arginin-Resten di-methyliert (2, rot). Anschließend werden sie auf den SMN-Komplex übertragen. Nur mit den am SMN-Komplex gebundenen Sm-Proteinen kann die U snRNA mit ihrer Sm-Stelle in Wechselwirkung treten (3). Es bildet sich ein stabiles Sm-core-Partikel aus, das aus der U snRNA und einem siebenteiligen Ring von Sm-Proteinen besteht. Nun wird die Cap-Struktur der U snRNA hypermethyliert (4) und das fertige Sm-core-Partikel in den Zellkern importiert (5).

Zellkern und erfüllt schließlich seine Funktion im Spleißosom.

Das Paradoxon

Obwohl die Zusammenlagerung von U snRNA und Sm-Proteinen im Reagenzglas spontan erfolgen kann, war für das Verständnis der SMA die Beobachtung entscheidend, dass dieser Vorgang in der lebenden Zelle die Mitwirkung des SMN-Proteins erfordert[3]. So wurde zunächst eine transiente Wechselwirkung zwischen SMN und U snRNAs im Zytoplasma beobachtet. In weitergehenden Studien ließ sich dann zeigen, dass das SMN-Protein Teil einer makromolekularen Einheit, des SMN-Komplexes ist, der die Sm-Proteine in einer aktiven Weise auf die U snRNA lädt[4]. Eine solche faktorvermittelte RNP-Zusammenlagerung war bis dahin nicht beschrieben worden. Interessanterweise bedürfen Sm-Proteine zuvor einer posttranslationalen Modifikation, um effizient an den SMN-Komplex zu binden: Einige Arginin-Reste der Sm-Proteine werden durch den PRMT5-Komplex symmetrisch di-methyliert[5, 6].

Seit diese Zusammenhänge erforscht wurden, fragen sich die Wissenschaftler, warum überhaupt so viele assistierende Faktoren an der Biogenese der U snRNPs beteiligt sind, wenn sich diese Partikel doch augenscheinlich *in vitro* spontan und ohne jede Hilfe bilden können. Eine mögliche Erklärung für dieses Paradoxon könnten die physikochemischen Eigenschaften der Sm-Proteine bieten. Die hoch abundanten Proteine dieser Familie besitzen nämlich ungewöhnlich viele basische Aminosäuren und stets zwei einander gegenüberliegende Protein-Protein-Interaktionsoberflächen. Man kann sich daher vorstellen, dass ein geeignetes „Kontrollsystem" die Sm-Proteine daran hindern muss, in der Zelle zu aggregieren oder mit anderen, „falschen" RNAs in Wechselwirkung zu treten. Das SMN/PRMT5-System hätte demnach die Funktion eines Chaperons: Es ermöglicht einerseits die Zusammenlagerung eines funktionellen RNPs, andererseits schützt es aber auch die noch nicht zusammengelagerten Komponenten vor fehlerhaften Interaktionen.

Auswirkungen für SMA-Patienten

Jedoch: Was hat dies mit der bei SMA-Patienten beobachteten Degeneration von Motoneuronen zu tun? Gibt es tatsächlich eine Verbindung zwischen der Funktion von SMN in der Zusammenlagerung von U snRNPs und der SMA? Man müsste dem Gesagten zufolge postulieren, dass in Zellen von Patienten (und hier besonders in den betroffenen Motoneuronen) die Zusammenlagerung der U snRNPs vermindert abläuft. Dies ließ sich in der Tat experimentell nachweisen. Aber ist dieser biochemische Defekt auch direkt für den Untergang der Motoneurone verantwortlich oder spielen hier noch andere Mechanismen eine Rolle? Obwohl diese Frage noch nicht abschließend beantwortet ist, deuten einige im Zebrafisch-Modell erhaltene experimentelle Befunde darauf hin. Durch Antisense-Strategien ließ sich in diesem Tiersystem die SMN-Expression herunterregulieren, wodurch man eine Situation wie bei SMA-Patienten simulieren konnte. Erwartungsgemäß zeigten so veränderte Fische einen Phänotyp, der sehr an die Situation bei der SMA erinnert: Die Beweglichkeit der Fische war stark eingeschränkt und Motoneurone wiesen massiv verkürzte und verzweigte Axone auf. Interessanterweise ließ sich dieser Phänotyp durch die Bereitstellung von vollständig zusammengebauten U snRNPs effizient unterdrücken. In der Tat scheint also die unzureichende Bereitstellung von U snRNPs die Motoneurone in den Untergang zu treiben[7].

Schlussfolgerung

Da nun U snRNPs essenzielle Bestandteile des Spleißosoms sind, könnte es – so eine gängige Arbeitshypothese – im Körper von SMA-Patienten zu einer generellen Schwächung des Spleißprozesses kommen[8]. Prä-mRNAs, die aufgrund ihrer Beschaffenheit auch unter normalen zellulären Bedingungen nur ineffizient gespleißt werden, ließen sich unter SMN-Mangelbedingungen möglicherweise gar nicht mehr prozessieren. Bislang wurden solche Prä-mRNAs noch nicht identifiziert. Jedoch könnte genau ein solches Szenario erklären, weshalb sich ein genereller zellulärer Defekt nur in einem Zelltyp auswirkt: in den Motoneuronen. ■

Literatur

[1] **Lefebvre, S., Bürglen, 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., Melki, J.** (1995): Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80: 155–165.

[2] **Lorson, C. L., Hahnen, E., Androphy, E. J., Wirth, B.** (1999): A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.* 96: 6307–6311.

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

[4] **Meister, G., Bühler, D., Pillai, R., Lottspeich, F., Fischer, U.** (2001): A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol.* 3: 945–949.

[5] **Paushkin, S., Gubitz, A. K., Massenet, S., Dreyfuss, G.** (2002): The SMN complex, an assemblyosome of ribonucleoproteins. *Curr. Opin. Cell Biol.* 14: 305–312.

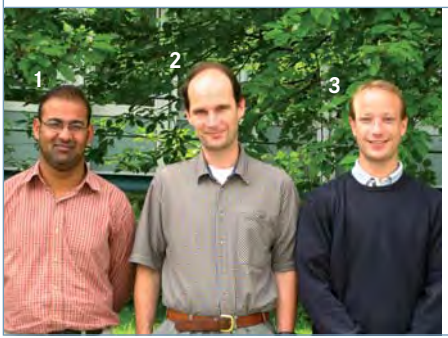
[6] **Meister, G., Eggert, C., Fischer, U.** (2002): SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol.* 12: 472–478.

[7] **Winkler, C., Eggert, C., Gradl, D., Meister, G., Giegerich, M., Wedlich, D., Laggerbauer, B., Fischer, U.** (2005): Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev.* 19: 2320–2330.

[8] **Eggert, C., Chari, A., Laggerbauer, B., Fischer, U.** (2006): Spinal muscular atrophy: the RNP connection. *Trends Mol. Med.* 12: 113–121.

Korrespondenzadresse:
Prof. Dr. Utz Fischer
Theodor-Boveri-Institut am Biozentrum
Universität Würzburg
Am Hubland
D-97074 Würzburg
Tel.: 0931-888 4029
Fax: 0931-888 4028
utz.fischer@biozentrum.uni-wuerzburg.de
www.biochem.biozentrum.uni-wuerzburg.de

AUTOREN



Utz Fischer[2]
1983–1988 Biochemiestudium an der Freien Universität Berlin, **1989–1992** Promotion an der Philipps-Universität Marburg, **1992–1997** Postdoctoral Fellow am IMT in Marburg und am Howard Hughes Medical Institute, PA, USA, **1997–2003** Leiter einer Unabhängigen Arbeitsgruppe am MPI für Biochemie, Martinsried. Seit **2003** Professor für Biochemie am Theodor-Boveri-Institut (Biozentrum) der Universität Würzburg

Ashwin Chari[1]
1999–2004: Studium der Biochemie, Molekularbiologie und Biophysik an der ETH Zürich. Seit **2004** Doktorand am Lehrstuhl für Biochemie, Theodor-Boveri-Institut (Biozentrum), Universität Würzburg.

Matthias Kroiß[3]
1998–2004 Medizinstudium an der Universität Würzburg, **2004–2006** Promotion am Institut für Anatomie und Zellbiologie der Universität Würzburg. Seit **2005** Doktorand am Lehrstuhl für Biochemie, Theodor-Boveri-Institut (Biozentrum), Universität Würzburg.

6.2 The Mechanistic Basis of Cellular U snRNP Formation

Deciphering the assembly pathway of Sm-class U snRNPs

Neuenkirchen N, Chari A, Fischer U: **FEBS Lett 2008**, 582:1997-2003.

The role of RNP biogenesis in spinal muscular atrophy

Chari A, Paknia E, Fischer U: **Curr Opin Cell Biol 2009**, 21:387-393.

Molekulares Bodybuilding: Wie Zellen RNA-Proteinkomplexe herstellen

Chari A, Fischer U: **Biospektrum 2009**, 15. Jahrgang, in press

FEBS Letters 582 (2008) 1997–2003

Minireview

Deciphering the assembly pathway of Sm-class U snRNPs

Nils Neuenkirchen, Ashwin Chari, Utz Fischer*

Chair of Biochemistry, Theodor-Boveri Institute at the Biocenter, Am Hubland, University of Wuerzburg, D-97074 Wuerzburg, Germany

Received 11 February 2008; revised 9 March 2008; accepted 10 March 2008

Available online 17 March 2008

Edited by Ulrike Kutay

Abstract The assembly of the Sm-class of uridine-rich small nuclear ribonucleoproteins (U snRNPs), albeit spontaneous in vitro, has recently been shown to be dependent on the aid of a large number of assisting factors in vivo. These factors are organized in two interacting units termed survival motor neuron (SMN)- and protein arginine methyltransferase 5 (PRMT5)-complexes, respectively. While the PRMT5-complex acts early in the assembly pathway by activating common proteins of U snRNPs, the SMN-complex functions to join proteins and RNA in a highly ordered, apparently regulated manner. Here, we summarize recent progress in the understanding of this process and discuss the influence exerted by the aforementioned *trans-acting* factors.

© 2008 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: U snRNP biogenesis; SMN-complex; PRMT5-complex; Sm proteins; Splicing; U snRNA

1. Assembly of RNA–protein complexes in the cellular environment

The ability to reconstitute macromolecular complexes from purified single components in vitro has fostered the concept of self-assembly, one of the central principles of molecular biology. From a theoretical, "in vitro" point of view, the formation of macromolecular complexes depends on diffusion-driven, random, and reversible encounters of the cognate subunits [1,2]. The stability of the resulting entities is determined by the ratio of the respective association and dissociation rate constants. While these considerations hold true for idealized, aqueous solutions, the situation in vivo is markedly different. In cells, the local concentration of individual components of macromolecular complexes and other proteins is usually relatively low, yet as a whole they occupy a significant fraction of the total volume. Hence, the possibility for unspecific interactions arises, which hinder the assembly pathway. Individual molecules therefore need to be directed to the site of complex assembly, in order to increase their local concentration and protect them against unfavorable interactions. Furthermore, a separation of both the site of assembly and the site of function should help to prevent assembly intermediates from adversely affecting the function of fully assembled macromolecular complexes. Taking these considerations into account, it is not surprising that cells have evolved strategies to ensure the faithful generation of macromolecular assemblies [1]. Among others, three features seem to predominate: (1) the segregation of biosynthesis of individual components and their assembly into higher-order structures into different subcellular compartments; (2) the evolution of molecular chaperones, which promote formation of intermediates, shielding these intermediates from adverse, premature interactions with substrate molecules of the finally assembled macromolecular complexes; and (3) *trans-acting* factors, working as scaffolds to coordinate several processes of the assembly reaction. A number of biological processes follow, at least in part, these principles and well known examples are the assembly of proteasomes and nucleosomes [3,4].

In this review, we summarize recent experimental advances in the understanding of the in vivo biogenesis pathway of macromolecular RNA–protein complexes found in the spliceosome uridine-rich small nuclear ribonucleoproteins (U snRNPs). We speculate that many aspects of this pathway have evolved as a consequence of the theoretical considerations outlined above.

2. Biogenesis of snRNPs

The spliceosome is a macromolecular machine comprising several RNP subunits that catalyzes the removal of intervening (non-coding) sequences from pre-mRNA (pre-messenger RNA). U snRNPs of the major spliceosome (responsible for splicing the majority of pre-mRNA introns) consist of either one (U1, U2, U5) or two (U4/U6) small nuclear RNAs (snRNAs), and a large number of proteins [5,6]. As individual classes of U snRNPs perform distinct functions in the spliceosome, it is not surprising that each is characterized by a specific set of proteins. However, all U snRNPs also contain a set of seven common, so-called Sm proteins B/B′, D1, D2, D3, E, F and G (B′ is a splicing variant of B). These evolutionarily related proteins form a heptameric ring around a conserved sequence

*Corresponding author. Fax: +49 931 888 4028.
E-mail address: utz.fischer@biozentrum.uni-wuerzburg.de (U. Fischer).

Abbreviations: snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein; SMN, survival motor neuron; PRMT5, protein arginine methyltransferase 5; WD45, WD repeat domain 45; pICln, chloride conductance regulatory protein; CBC, cap-binding complex; PHAX, phosphorylated adaptor for RNA export; CRM1, chromosome region maintenance 1; RanGTP, Ras-related nuclear protein bound to GTP; NPC, nuclear pore complex; Tgs1, trimethylguanosine synthetase1; NLS, nuclear localization signal; SPN1, snurportin-1; Lsm proteins, like Sm protein

0014-5793/$34.00 © 2008 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2008.03.009

motif ($PuAU_{4-6}GPu$) on the snRNA termed the "Sm-site" [7], which is a structural hallmark of these particles [8,9]. The biochemical composition of the minor spliceosome, specialized for processing of a small group of introns (so-called ATAC introns), requires a different set of U snRNPs (termed U11, U12, U4atac and U6atac; U5 appears to be identical in both the major and the minor spliceosome) [10]. However, the general architecture and hence the biogenesis of the U snRNPs of the major and minor spliceosome is believed to be very similar. Thus, although we will concentrate on the description of biogenesis of the components of the major spliceosome, the same principles will most likely also apply for the minor snRNPs.

A large number of studies performed mainly in *Xenopus laevis* oocytes but also in somatic cells have contributed to the understanding of the transport pathways enabling the biogenesis of spliceosomal U snRNPs [6]. Due to these studies the biogenesis of U snRNPs could be divided into several steps, some of which may actually be coupled. Initially, the RNA polymerase II (pol II) transcribes precursor snRNAs (pre-snRNAs) of U1, U2, U4 and U5 in the nucleus that are co-transcriptionally m^7G-capped (Fig. 1, step 1). Transcriptional termination was recently shown to be dependent on a megadalton complex termed Integrator [11]. In the mechanistic model, the Integrator complex is assumed to interact with RNA polymerase II at the promoter of the snRNA genes. Remaining attached to RNA polymerase during transcription, the Integrator complex endonucleolytically cleaves the pre-snRNA upon interaction with the 3′-box; a *cis-acting* element of 13–16 nucleotides that is required for efficient pre-snRNA formation. The m^7G-cap of the pre-snRNA is specifically recognized by the cap-binding complex (CBC), which itself associates with phosphorylated adaptor for RNA export (PHAX), chromosome region maintenance 1 (CRM1) and Ras-related nuclear protein bound to GTP (RanGTP) to form the nuclear export complex [12]. After transport through the nuclear pore complex (NPC) (step 2), GTP hydrolysis of Ran and dephosphorylation of PHAX cause the transport factors to dissociate from the snRNA [13,14] (step 3). U6 RNA (and also U6atac) differ from other snRNAs in that they are transcribed by RNA polymerase III, acquire a γ-monomethyl cap and appear not to leave the nucleus post-transcriptionally [15,16].

In the cytoplasm, the survival motor neuron complex (SMN-complex) facilitates the transfer of seven Sm proteins onto the snRNA's "Sm-site", which results in the formation of the Sm core domain (step 4; see next paragraph for a detailed description). Recent evidence suggests that the U snRNP remains bound to the SMN-complex even after assembly is completed [17]. The Sm core domain and the SMN-complex then cooperate to recruit the cap-hypermethylase trimethylguanosine synthetase1 (Tgs1) [18]. This leads to the conversion of the m^7G-cap of the snRNA to its hypermethylated form, the 2,2,7-trimethylguanosine-cap (m_3G, also termed TMG) (step 5). At an as yet to be defined step in the cytoplasm, the mature 3′-end of U snRNAs is generated by unknown (exo)ribonuclease(s) in a process referred to as 3′-trimming [19,20].

The assembled and processed particle is then imported into the nucleus by means of a bipartite nuclear localization signal (NLS) on the U snRNP and at least two specific transport factors. One part of the NLS is the m_3G-cap that is recognized by the protein snurportin-1 (SPN1) [21]. This interaction alone is not sufficient for import of U snRNPs but requires assistance of importin β [22]. It is believed that importin β has a docking site on the Sm core-bound SMN-complex [23] and hence may constitute the second part of the bipartite NLS, which has long-since been suspected to lie on the Sm core domain [24]. Once both transport factors have bound to their respective signals, nuclear import can be effected (step 6). Within the nucleus, the import complex dissociates and the transport factors are recycled to the cytoplasm (step 7). The U snRNPs, presumably still attached to the SMN-complex, transiently accumulate in subnuclear domains termed Cajal bodies (step 8). Within these domains, small Cajal body RNAs (scaRNAs) introduce site-specific pseudouridylation (Ψ) and 2′-*O*-methylation (m) in the snRNAs [25,26] and thereby complete processing of U snRNAs (step 9). For most snRNP-specific proteins it remains unknown whether they join the particle already in the cytoplasm, as is the case for Sm proteins, or only after re-import into the nucleus. Ultimately, the mature spliceosomal snRNPs accumulate in interchromatin regions in structures referred to as splicing speckles (step 10). The SMN-complex, which is dissociated from the U snRNP at a yet to be identified stage in the nucleus, is then believed to return into the cytoplasm (step 11), where it can re-enter the biogenesis cycle (step 12).

3. *Trans-acting* factors mediate the assembly of spliceosomal U snRNPs

When isolated Sm proteins are incubated with U snRNA under appropriate conditions in vitro, efficient Sm core formation can be observed. This process takes place in an apparently ordered and defined manner. Prior to RNA-binding, Sm proteins form heterooligomeric complexes composed of D1–D2, B/B′–D3, and E–F–G [27,28]. RNP-binding occurs in two steps, first, the so-called subcore particle is formed by interaction of D1–D2 and E–F–G, which is then transformed by the addition of B/B′–D3 into the core particle [28]. These observations have led to the conclusion that all structural information required for the formation of this core-RNP is encoded within these components (i.e. RNA and proteins). But the fact that these structures form in vitro does not necessarily prove that it is also the case in vivo, in particular if the deliberations of the first paragraph are taken into account. Indeed, a body of recent evidence indicates that formation of the Sm core domain of U snRNPs requires ATP and assisting factors, the number of which exceeds that of proteins actually assembled onto the U snRNA [29,30].

The first factor implicated in the biogenesis of U snRNPs was the survival motor neuron (SMN) protein. This factor, whose reduced expression elicits the neuromuscular disease spinal muscular atrophy (SMA), was found to transiently interact with U snRNAs in the cytoplasm but was not an integral part of mature U snRNPs [31]. As detailed biochemical studies further revealed, SMN is a constituent of a macromolecular complex consisting of at least eight key subunits (termed Gemins2–8 and unrip; Fig. 2) [29,30,32–34]. In addition to these integral components, spliceosomal Sm proteins and U snRNAs can be found transiently associated with the SMN-complex. Initial insights into the function of the SMN-complex in U snRNP biogenesis were gained by the biochemical reconstitution of the in vivo assembly reaction from isolated components [29]. These studies have revealed that Sm proteins must first bind to

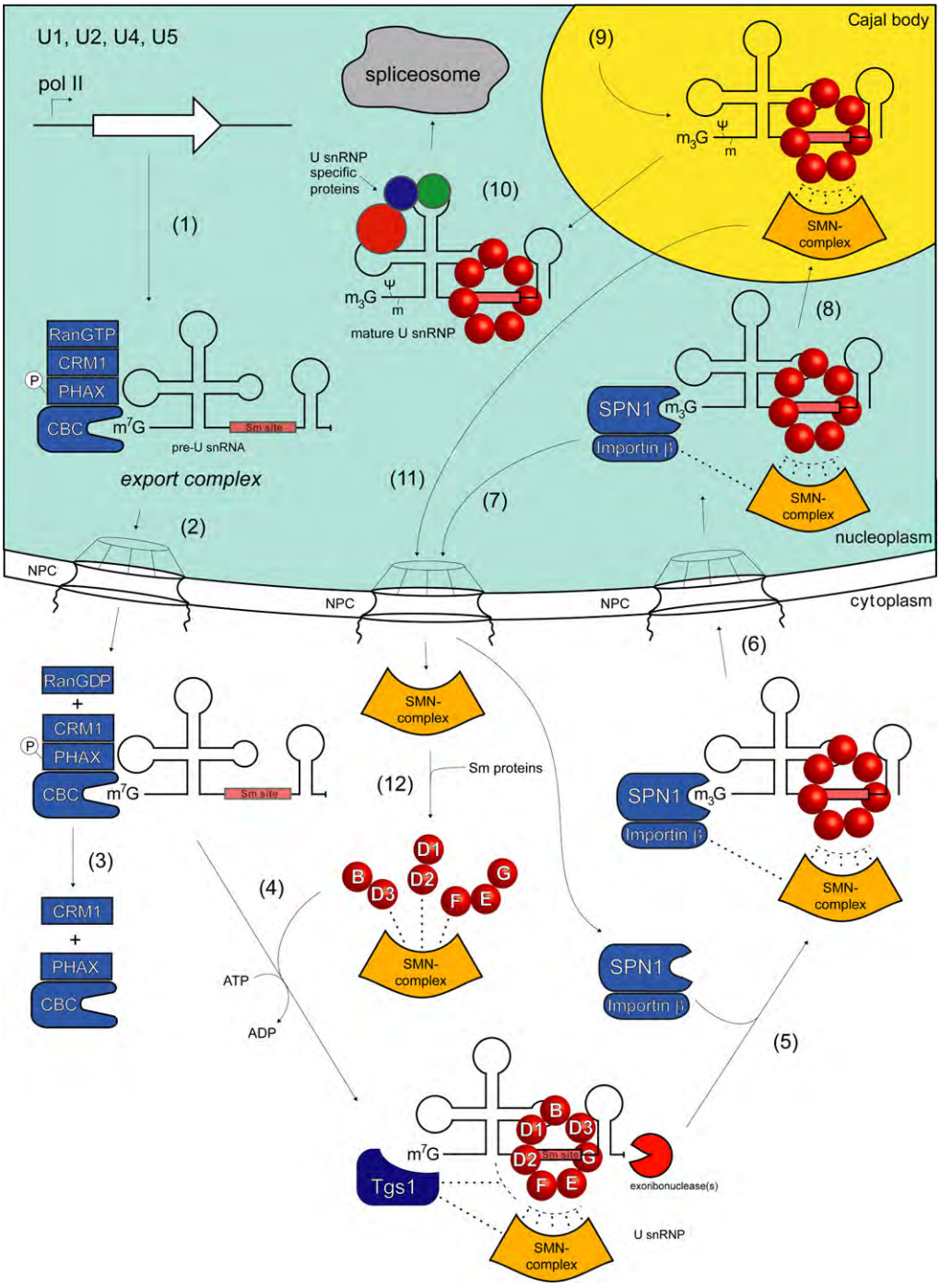


Fig. 1. Biogenesis pathway of spliceosomal U snRNPs. Pre-U snRNA is transcribed by RNA polymerase II (pol II) and m^7G-capped in the nucleus (step 1). After the export complex, consisting of pre-U snRNA, CBC, PHAX, CRM1 and RanGTP, has formed, it is actively transported into the cytoplasm via the NPC (step 2). There, export factors and pre-U snRNA dissociate from each other (step 3) and Sm proteins provided by the SMN-complex are assembled onto the "Sm-site" of pre-U snRNA (step 4). Following recruitment by the SMN-complex and Sm core domain, the hypermethylase Tgs1 modifies the m^7G-cap to m_3G (step 5), before the import factors SPN1 and importin β mediate translocation into the nucleus (step 6). There, both factors dissociate and are recycled into the cytoplasm (step 7), and U snRNPs associated with the SMN-complex enrich in Cajal bodies (step 8). After scaRNA guided pseudouridylation (Ψ) and 2′-*O*-methylation (m; step 9), the mature U snRNP is directed to the spliceosome, (step 10), whereas the SMN-complex is believed to be exported into the cytoplasm (step 11), where it can re-enter the biogenesis cycle (step 12).

the SMN-complex before they can be transferred onto U snRNA. Thus, unlike in vitro, Sm proteins cannot be directly delivered onto U snRNA within the context of a living cell and hence the assembly reaction does not strictly follow a "self-assembly route" (see Fig. 1, step 4).

Although the SMN-complex loaded with all Sm proteins is alone necessary and sufficient for U snRNP assembly, this process is influenced by another complex, whose name-giving constituent is the methyltransferase protein arginine methyltransferase 5 (PRMT5) [35–37]. This enzyme, in conjunction

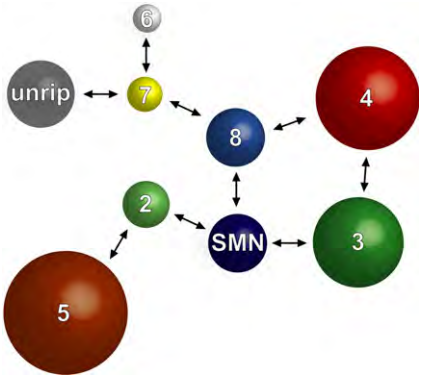


Fig. 2. Interaction map of the human SMN-complex. Schematic of protein interactions within the human SMN-complex as described in [34]. The SMN protein together with Gemin8 and Gemin7 form a core scaffold of the SMN-complex by which the remaining components are recruited. SMN is directly connected to Gemin2, which itself is associated with Gemin5. Furthermore, both Gemin6 and unrip are recruited by Gemin7, whereas Gemin3 and Gemin4 are cooperatively bound by SMN and Gemin8, respectively.

with its interaction partners WD repeat domain 45 (WD45), also termed Mep50, and chloride conductance regulatory protein (pICln), catalyzes the formation of symmetrical dimethylarginines within the C-terminal tails of the Sm proteins B/B′, D1 and D3. Since these modifications increase the affinity of the Sm proteins for the SMN protein in vitro, it is assumed that one function of the PRMT5-complex is the stimulation of the assembly process [35,38]. Recently, evidence for yet another arginine methyltransferase protein arginine methyltransferase 7 involved in Sm protein activation has been reported in HeLa cells which acts in a non-redundant fashion on Sm protein modification [39]. Interestingly, however, genetic inactivation of PRMT5 in *Drosophila melanogaster* has no recognizable influence on U snRNP biogenesis [40]. Therefore, the question arises whether the symmetric dimethylation of Sm proteins is an absolute prerequisite or whether it is only an accessory function in snRNP biogenesis. SMN- and PRMT5-complexes directly interact with each other in a higher-order structure [41]. The cooperating SMN- and PRMT5-complexes can be hence envisaged as the functional unit that promotes and regulates the assembly of spliceosomal U snRNPs.

4. A model for the assisted assembly of spliceosomal U snRNPs

Based on reported data from several laboratories, a model for the assisted U snRNP assembly process can be proposed [42,43]. Sm proteins, synthesized in the cytoplasm, are initially sequestered by the PRMT5-complex (Fig. 3). PICln is likely to play an important role in complex formation as it directly binds to Sm proteins and PRMT5 [35–37]. These initial steps commit Sm proteins to the SMN-mediated assembly pathway. Once recruited onto the PRMT5-complex, Sm proteins B/B′, D1 and D3 are symmetrically dimethylated on arginine residues and may hence become "activated" for subsequent steps (Fig. 3, step 1). We speculate that the PRMT5-complex (or parts thereof) facilitate(s) additional events in the assembly pathway, such as organization of specific Sm protein sub-complexes. In this context, it is noteworthy that each Sm protein occupies a specific spatial position within the Sm core domain of the assembled U snRNP [8]. PRMT5- and SMN-complexes then join to form the SMN-/PRMT5-complex, in which the complete set of Sm proteins is first transferred onto the SMN-complex (step 2), and then passed onto U snRNA (step 3). Ultimately, the assembled U snRNP is transferred along with the SMN-complex to the nucleus, where the U snRNP is further processed and targeted to its site of function. The SMN-complex is then exported into the cytoplasm to engage in another assisted U snRNP biogenesis cycle (step 4, see also Fig. 1, step 12).

5. Open questions and future directions

Some questions need to be addressed to understand the mechanism of this unique assembly system. One of them is how the flow of Sm proteins through the assembly machinery onto U snRNA is facilitated and regulated. We favor a model, in which Sm proteins are pre-assembled on the PRMT5-complex to form the heterooligomers B/B′–D3, D1–D2 and E–F–G, as a prerequisite for the transfer onto the SMN-complex. This model implies that Sm proteins on the PRMT5-complex do not have the propensity of binding to U snRNA, a situation that obviously changes upon transition to the SMN-complex associated state. What could be the switch to turn it from an assembly incompetent into an assembly active state? The answer may lie within the architecture of the components that make up the Sm core domain of U snRNPs, with the RNA being tightly surrounded by seven Sm proteins. Two obvious scenarios could explain formation of such a structure: In one scenario, the Sm protein ring is formed on the SMN-complex and subsequently the RNA threaded through the central hole. Considering the spatial organization of the U snRNP, this mechanism appears to be rather unlikely. A more probable scenario may be a clamp-loading like mechanism. In this process, the Sm proteins are kept on the SMN-complex in an "open ring" configuration. Upon binding of the U snRNA, the SMN-complex may undergo structural rearrangements leading to the closure of the Sm protein ring around the "Sm-site". Therefore, such an "open ring" conformation should be induced by the topology of the SMN-complex. In preceding steps of the assembly line, however, it should be disallowed. This model implies a conformational switch of the SMN-complex and a step in which the RNA is identified and bound onto the open Sm ring. Gemin3 and Gemin5 may play crucial roles in these postulated events. The Gemin3 protein belongs to the DEAD-box family of RNA helicases and may explain why assembly is dependent on the hydrolysis of ATP. Gemin5, in contrast, has been shown to specifically recognize "Sm-site" containing RNAs and may hence guide the U snRNA to the "Sm-site" [44]. Further studies are required to clarify, whether this scenario holds true, or another yet to be discovered mechanism, accounts for the formation of the Sm core domain.

As outlined above, U snRNPs (like other RNPs) can assemble spontaneously in vitro. Therefore, one may ask why *trans-acting* assembly factors are required in vivo. Life without this system is impossible as illustrated by the fact that inactivation of SMN, Gemin2 and pICln is lethal in several organisms,

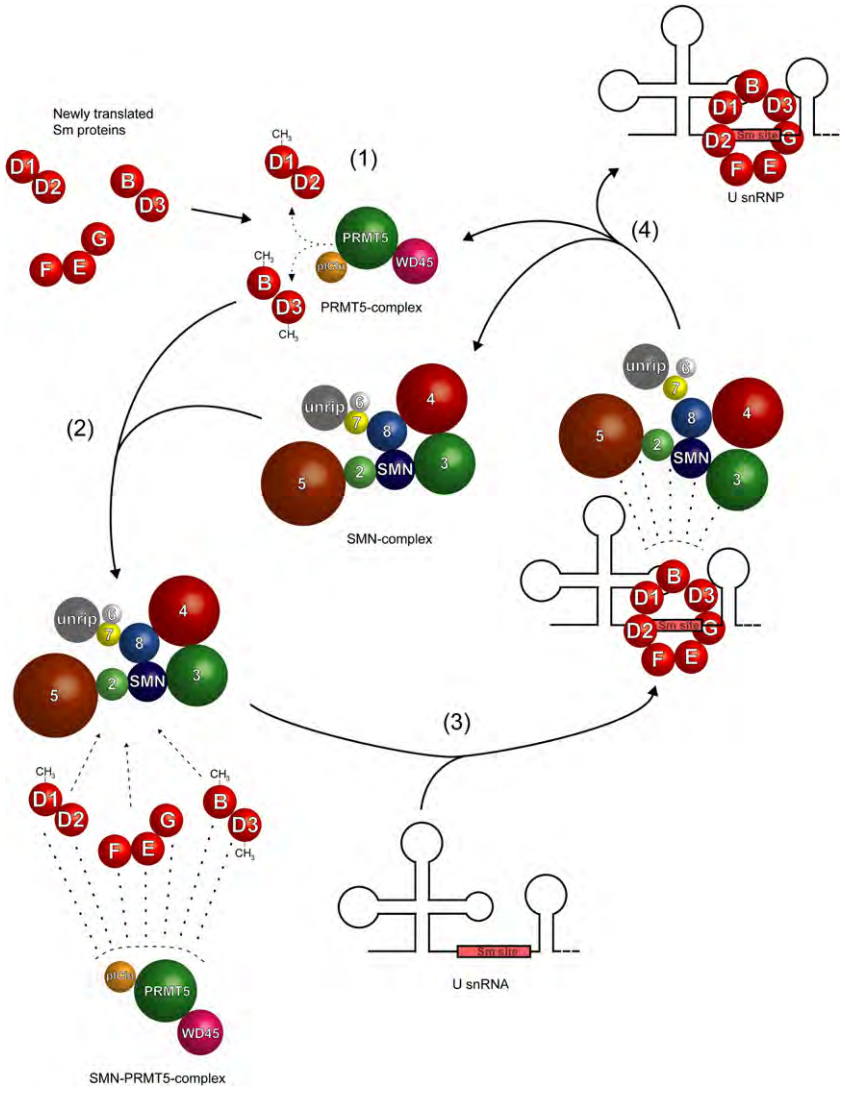


Fig. 3. Model of assisted assembly of U snRNPs. Sm proteins are initially translated in the cytoplasm and sequestered by the PRMT5-complex, consisting of the Type II methyltransferase PRMT5, WD45 (also termed Mep50) and pICln, which promotes symmetric dimethylation of arginines on Sm proteins B/B′, D1 and D3 (step 1). Next, the SMN-complex interacts with the PRMT5-complex to form an SMN-PRMT5-complex in which the Sm proteins are transferred onto the SMN-complex (step 2). These Sm proteins are assembled onto the "Sm-site" of U snRNAs to form U snRNPs (step 3). Finally, the U snRNP, the SMN-complex and PRMT5-complex dissociate and the latter two engage in a new round of U snRNP biogenesis (step 4).

including mice (see [32] for a review). This excludes the possibility that spontaneous assembly is a default pathway in vivo, which is redundant with assisted RNP formation. We speculate that one function of the SMN–PRMT5-system is to serve as a chaperone system that prevents mis-assembly of Sm proteins to non-target RNA and Sm protein aggregation. Indeed, first experimental evidence for this activity has been provided recently [30].

Spliceosomal U snRNPs are certainly the most abundant class of particles, assembly of which is mediated by the SMN-PRMT5-system. But are there also other classes of particles depending on this system? This is true for U snRNPs of the minor spliceosome (i.e. U11, U12, U4atac), which contain an Sm core domain indistinguishable from their counterparts of the major spliceosome. Thus, we are confident to postulate a common assembly for most, probably all, particles with "canonical" Sm core domains. However, a large number of RNPs contain core structures composed, at least in part, of different components. One such case is the U7 snRNP, a particle involved in 3′-end processing of histone mRNA. The core domain of this particle consists of the canonical Sm proteins B, D3, E, F and G, whereas D1 and D2 are replaced by the "like Sm proteins" 10 and 11 (termed Lsm10 and 11) [45]. Interestingly, assembly of the Lsm/Sm core of the U7 snRNP has been shown to be dependent on the SMN-complex charged with the U7-specific set of core proteins [46]. Thus, also a particle with a mixed Lsm/Sm core depends on this assembly machinery.

A group of related like Sm proteins (Lsm proteins) has been identified recently (termed Lsm1–8) which can form heptameric ring-like structures very similar in shape and size to the "canonical" Sm core domain [47]. Depending on their composition, they mediate either RNA degradation (Lsm1–7) [48,49], or function as core components of the spliceosomal snRNAs U6 and U6atac (Lsm2–8) [50–52]. Interestingly, these

Lsm-exclusive rings can form in the absence of RNA and hence behave in this respect markedly different than Sm (and Lsm10/11) proteins [50]. If the SMN-complex acts exclusively as a clamp loader onto the respective cognate RNAs, as outlined above, assembly of the Lsm rings may occur independently of this system. However, SMN (as a single protein) has been shown to bind to Lsm proteins in vitro, providing the possibility that at least some proteins of the SMN-complex also play a role in the biogenesis of Lsm rings [53].

Finally, some nuclear and nucleolar RNAs such as box C/D small nucleolar RNA, box H/ACA and telomerase RNA have been shown to associate with distinct subsets of Sm proteins or other classes of proteins, which are able to interact with SMN [54–58]. We regard the development of in vitro assembly assays, which recapitulate the in vivo situation, an obligate prerequisite to address the question whether the SMN–PRMT5-system is indeed a master assembler for a large variety of different RNPs or whether this system is restricted to a smaller class harboring only specific sets of Sm and Lsm proteins.

Acknowledgements: We thank all members of the Fischer lab, in particular Elham Paknia, Matthias Kroiß and Clemens Grimm, for helpful comments on the manuscript. We apologize to our colleagues whose work could not be cited due to size constraints.

References

[1] Ellis, R.J. and Minton, A.P. (2006) Protein aggregation in crowded environments. Biol. Chem. 387, 485–497.
[2] Zimmerman, S.B. and Minton, A.P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu. Rev. Biophys. Biomol. Struct. 22, 27–65.
[3] Dutta, S., Akey, I.V., Dingwall, C., Hartman, K.L., Laue, T., Nolte, R.T., Head, J.F. and Akey, C.W. (2001) The crystal structure of nucleoplasmin-core: implications for histone binding and nucleosome assembly. Mol. Cell 8, 841–853.
[4] Kruger, E., Kloetzel, P.M. and Enenkel, C. (2001) 20S proteasome biogenesis. Biochimie 83, 289–293.
[5] Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell 12, 5–14.
[6] Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr. Opin. Cell Biol. 13, 290–301.
[7] Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. and Jacob, M. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J. 1, 1259–1265.
[8] Kambach, C. et al. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96, 375–387.
[9] Stark, H., Dube, P., Luhrmann, R. and Kastner, B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature 409, 539–542.
[10] Will, C.L. and Luhrmann, R. (2005) Splicing of a rare class of introns by the U12-dependent spliceosome. Biol. Chem. 386, 713–724.
[11] Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell 123, 265–276.
[12] Askjaer, P. et al. (1999) RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. Mol. Cell Biol. 19, 6276–6285.
[13] Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I.W. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101, 187–198.
[14] Segref, A., Mattaj, I.W. and Ohno, M. (2001) The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. RNA 7, 351–360.
[15] Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. Proc. Natl. Acad. Sci. USA 83, 8575–8579.
[16] Reddy, R., Henning, D., Das, G., Harless, M. and Wright, D. (1987) The capped U6 small nuclear RNA is transcribed by RNA polymerase III. J. Biol. Chem. 262, 75–81.
[17] Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I.W. and Dreyfuss, G. (2002) The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. Mol. Cell Biol. 22, 6533–6541.
[18] Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. and Bordonne, R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. Mol. Cell 9, 891–901.
[19] Neuman de Vegvar, H.E. and Dahlberg, J.E. (1990) Nucleocytoplasmic transport and processing of small nuclear RNA precursors. Mol. Cell Biol. 10, 3365–3375.
[20] van Hoof, A., Lennertz, P. and Parker, R. (2000) Yeast exosome mutants accumulate 3′-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol. Cell Biol. 20, 441–452.
[21] Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. and Luhrmann, R. (1998) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. EMBO J. 17, 4114–4126.
[22] Palacios, I., Hetzer, M., Adam, S.A. and Mattaj, I.W. (1997) Nuclear import of U snRNPs requires importin beta. EMBO J. 16, 6783–6792.
[23] Narayanan, U., Ospina, J.K., Frey, M.R., Hebert, M.D. and Matera, A.G. (2002) SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. Hum. Mol. Genet. 11, 1785–1795.
[24] Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Luhrmann, R. (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.
[25] Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2′-*O*-methylation and pseudouridylation guide RNAs. EMBO J. 21, 2746–2756.
[26] Jady, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E. and Kiss, T. (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. EMBO J. 22, 1878–1888.
[27] Hermann, H., Fabrizio, P., Raker, V.A., Foulaki, K., Hornig, H., Brahms, H. and Luhrmann, R. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. EMBO J. 14, 2076–2088.
[28] Raker, V.A., Plessel, G. and Luhrmann, R. (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.
[29] Meister, G., Buhler, D., Pillai, R., Lottspeich, F. and Fischer, U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat. Cell Biol. 3, 945–949.
[30] Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. Science 298, 1775–1779.
[31] Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell 90, 1023–1029.
[32] Eggert, C., Chari, A., Laggerbauer, B. and Fischer, U. (2006) Spinal muscular atrophy: the RNP connection. Trends Mol. Med. 12, 113–121.
[33] Gubitz, A.K., Feng, W. and Dreyfuss, G. (2004) The SMN complex. Exp. Cell Res. 296, 51–56.
[34] Otter, S., Grimmler, M., Neuenkirchen, N., Chari, A., Sickmann, A. and Fischer, U. (2007) A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J. Biol. Chem. 282, 5825–5833.
[35] Friesen, W.J. et al. (2001) The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol. Cell Biol. 21, 8289–8300.

[36] Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C. and Fischer, U. (2001) Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr. Biol. 11, 1990–1994.
[37] Pu, W.T., Krapivinsky, G.B., Krapivinsky, L. and Clapham, D.E. (1999) pICln inhibits snRNP biogenesis by binding core spliceosomal proteins. Mol. Cell Biol. 19, 4113–4120.
[38] Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. and Luhrmann, R. (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.
[39] Gonsalvez, G.B., Tian, L., Ospina, J.K., Boisvert, F.M., Lamond, A.I. and Matera, A.G. (2007) Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. J. Cell Biol. 178, 733–740.
[40] Gonsalvez, G.B., Rajendra, T.K., Tian, L. and Matera, A.G. (2006) The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. Curr. Biol. 16, 1077–1089.
[41] Meister, G. and Fischer, U. (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. EMBO J. 21, 5853–5863.
[42] Meister, G., Eggert, C. and Fischer, U. (2002) SMN-mediated assembly of RNPs: a complex story. Trends Cell Biol. 12, 472–478.
[43] Yong, J., Wan, L. and Dreyfuss, G. (2004) Why do cells need an assembly machine for RNA–protein complexes? Trends Cell Biol. 14, 226–232.
[44] Battle, D.J., Lau, C.K., Wan, L., Deng, H., Lotti, F. and Dreyfuss, G. (2006) The Gemin5 protein of the SMN complex identifies snRNAs. Mol. Cell 23, 273–279.
[45] Pillai, R.S., Will, C.L., Luhrmann, R., Schumperli, D. and Muller, B. (2001) Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. EMBO J. 20, 5470–5479.
[46] Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Luhrmann, R., Fischer, U. and Schumperli, D. (2003) Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. 17, 2321–2333.
[47] Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. and Seraphin, B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J. 18, 3451–3462.
[48] Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. and Seraphin, B. (2000) A Sm-like protein complex that participates in mRNA degradation. EMBO J. 19, 1661–1671.
[49] Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D. and Parker, R. (2000) Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404, 515–518.
[50] Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Luhrmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. EMBO J. 18, 5789–5802.
[51] Mayes, A.E., Verdone, L., Legrain, P. and Beggs, J.D. (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J. 18, 4321–4331.
[52] Pannone, B.K. and Wolin, S.L. (2000) Sm-like proteins wRING the neck of mRNA. Curr. Biol. 10, R478–R481.
[53] Friesen, W.J. and Dreyfuss, G. (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.
[54] Fu, D. and Collins, K. (2006) Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association. Genes Dev. 20, 531–536.
[55] Jones, K.W., Gorzynski, K., Hales, C.M., Fischer, U., Badbanchi, F., Terns, R.M. and Terns, M.P. (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.
[56] Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R. and Branlant, C. (2003) A structural, phylogenetic, and functional study of 15.5-kDa/Snu13 protein binding on U3 small nucleolar RNA. RNA 9, 821–838.
[57] Peculis, B.A. and Steitz, J.A. (1994) Sequence and structural elements critical for U8 snRNP function in *Xenopus oocytes* are evolutionarily conserved. Genes Dev. 8, 2241–2255.
[58] Pellizzoni, L., Baccon, J., Charroux, B. and Dreyfuss, G. (2001) The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Curr. Biol. 11, 1079–1088.

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Cell Biology

The role of RNP biogenesis in spinal muscular atrophy

Ashwin Chari, Elham Paknia and Utz Fischer

Mutations that affect pre-mRNA processing are the cause for many genetic diseases. Most such mutations target *cis*-acting regulatory sequences in a given transcript, thus preventing its proper maturation. Only recently however, mutations in *trans*-acting factors involved in pre-mRNA processing have likewise been linked to disease. One prominent example is spinal muscular atrophy (SMA), a monogenic, neuromuscular disorder caused by reduced levels of functional survival motor neuron (SMN) protein. This ubiquitous factor is part of a complex that mediates the formation of spliceosomal snRNPs. The detailed biochemical investigation of SMN under normal conditions and in SMA has provided clues how mutations in factors with general functions elicit tissue-specific phenotypes.

Address
Department of Biochemistry, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

Corresponding authors: Chari, Ashwin (ashwin.chari@biozentrum.uni-wuerzburg.de), Paknia, Elham (paknia.elham@biozentrum.uni-wuerzburg.de) and Fischer, Utz (utz.fischer@biozentrum.uni-wuerzburg.de)

Current Opinion in Cell Biology 2009, **21**:387–393

This review comes from a themed issue on
Nucleus and gene expression
Edited by Elisa Izaurralde and Phillip Zamore

Available online 13th March 2009

0955-0674/$ – see front matter
© 2009 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.ceb.2009.02.004

Introduction

The faithful realization of the genetic code relies on the accurate processing of primary transcripts (pre-mRNAs) to mature messenger RNAs (mRNAs) (Figure 1, A2) and their translation into protein (Figure 1, A4). One key step to this end in eukaryotes is splicing, which comprises the removal of non-coding segments of the message (introns) and the joining of coding sequences (exons).

Pre-mRNA splicing occurs at consensus splice sites, that is, the 5′-region, branch point region and 3′-region of the intron. These sequences are recognized by a dynamic macromolecular machine, the spliceosome, which catalyzes the splicing reaction [1]. The vast majority of eukaryotic introns are processed by the major spliceosome, while a small fraction is processed by the so-called minor spliceosome [2]. Both spliceosomes are composed of RNA–protein complexes, known as snRNPs and a multitude of non-snRNP proteins [3]. The major spliceosome consists of the snRNPs U1, U2, U4/U6, and U5. The minor spliceosome, by contrast, comprises snRNPs U11, U12, U4atac/U6atac, and U5.

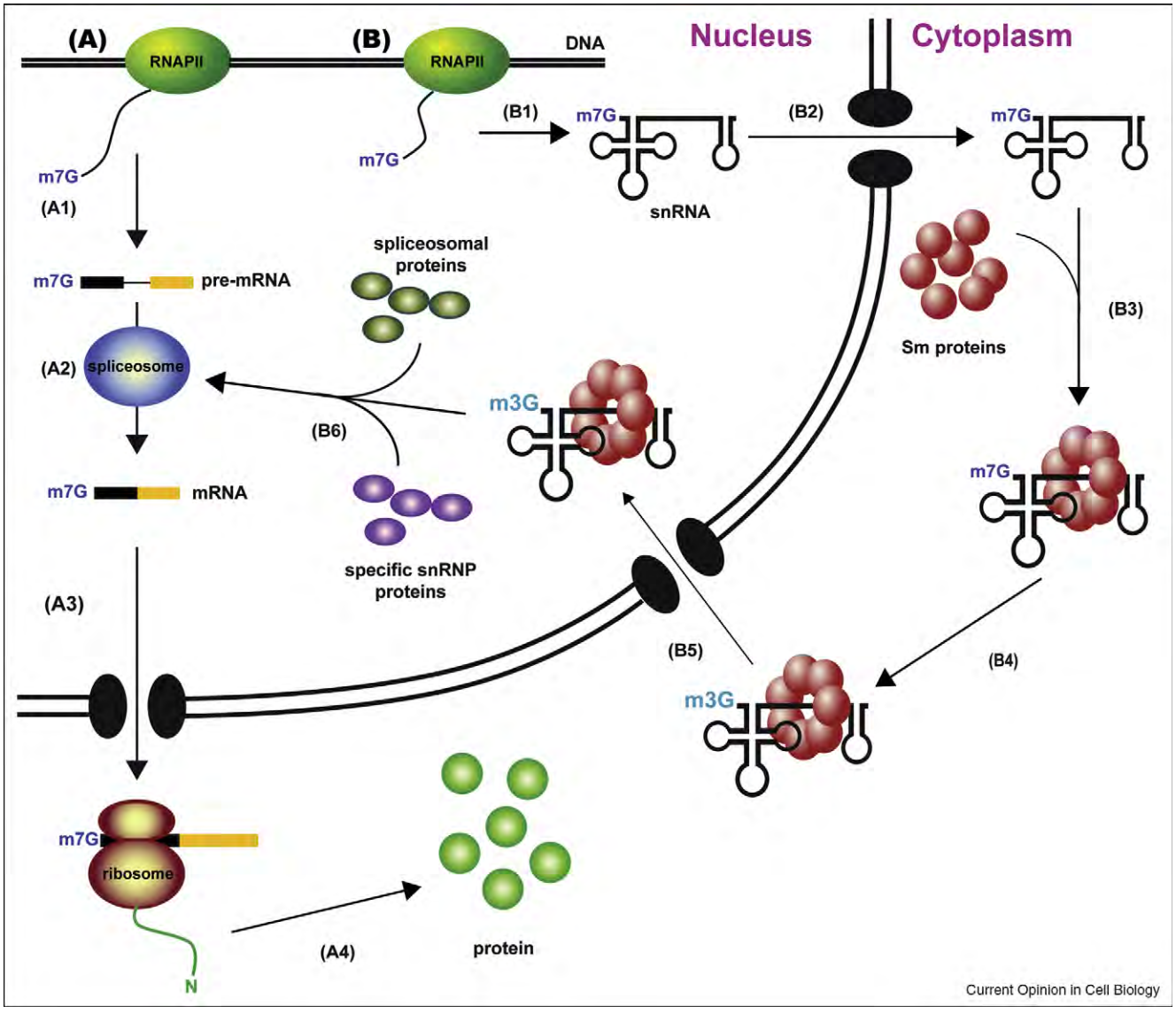
The activity of both spliceosomes is subject to a number of regulatory events. For instance, the processing of many pre-mRNAs is modulated by enhancer and silencer *cis*-elements located in either the intron itself or in the flanking exons. These elements are mostly binding sites for *trans*-acting factors, which alter the splicing pattern of a given pre-mRNA [4,5].

It has been appreciated in the past that mutations in splice sites, regulatory sequences, or *trans*-acting factors are a major cause of genetic diseases [6]. These disease mutations in *cis*-elements of pre-mRNAs often interfere with the appropriate processing of the affected primary transcript and hence with the generation of translation-competent mRNAs. However, the situation is more complicated when either the binding factors of splice-regulatory elements or components of the general splicing machinery are affected. In the former case, the splicing of a group of pre-mRNAs is expected to be impaired, while in the latter a more general defect is anticipated. The neuromuscular disorder spinal muscular atrophy (SMA) is an example for an ailment where a general splice factor is affected. The study of its disease gene product, the survival motor neuron (SMN) protein, has given a first glimpse at the etiology of this disease. These findings help to put forward a general model, how defects in a housekeeping factor elicit tissue-specific phenotypes.

SMA is a monogenic disorder linked to RNA metabolism

SMA is a common autosomal recessive disorder characterized by the degeneration of α-motoneurons in the spinal cord [7]. Compound heterozygous mutations or deletions of the SMN1 gene copy (*SMN1*), which prevent the expression of functional protein from this locus cause this monogenic disease [8]. Humans contain one or several copies of a second SMN-gene (termed *SMN2*). Even though both genes potentially encode for the same protein, a single nucleotide C-T transversion in exon 7 results in mis-splicing of >80% of the pre-mRNA transcribed from *SMN2* [9]. As a consequence, the majority of protein generated from this gene copy is truncated and unstable. Thus, SMA can be considered an SMN-deficiency syndrome, in which the deletion of the *SMN1* gene cannot be fully complemented by *SMN2*. SMN knockout animals are not viable and no SMA

Figure 1



Biogenesis of spliceosomal snRNPs converges with cellular gene expression. **(A)** Generation of mRNA and conversion into protein. **(A1)** Transcription of protein coding genes by RNA polymerase II (RNAPII) generates pre-mRNA. **(A2)** Removal of non-coding sequences (introns) and joining of the flanking, coding sequences (exons) by the spliceosome gives rise to mRNA. **(A3)** Export of the mature mRNA to the cytoplasm. **(A4)** Translation of mRNA to protein by the ribosome. **(B)** Biogenesis of spliceosomal snRNPs. **(B1)** Transcription of snRNA by RNAPII. **(B2)** Export of snRNA to the cytoplasm. **(B3)** Assembly of Sm proteins onto the snRNA. **(B4)** Hypermethylation of the monomethyl-guanosine (m^7G) cap to trimethyl-guanosine (m_3G) cap. **(B5)** Import of the assembled snRNP into the nucleus. **(B6)** Addition of specific snRNP proteins to the snRNP and incorporation of the mature snRNP along with non-snRNP spliceosomal proteins into the spliceosome (required for step A2).

patients have been identified that completely lack SMN. This suggests that SMN fulfills an essential function [10].

Biochemical analysis of SMN reveals an unexpected link to snRNP biogenesis

The finding that SMN is a housekeeping gene rather than a neuron-specific factor has raised the question about its function and its relation to the disease phenotype [11]. This question has been addressed and has revealed an unexpected link to the formation of major and minor spliceosomal snRNPs [12]. These splicing factors consist of an eponymous snRNA, common Sm proteins B, D1, D2, D3, E, F, and G and specific proteins. The intricate, segmented assembly pathway of these particles starts with the transport of newly transcribed snRNA to the cytoplasm (Figure 1, B2) [3]. In this compartment, Sm proteins are loaded onto a single stranded motif of the snRNA, called the Sm site, leading to the formation of a

ring-shaped Sm core domain (Figure 1, B3). In a next step, the monomethyl-guanosine (m^7G) cap of the snRNA is converted to a trimethyl-guanosine (m_3G) cap (Figure 1, B4). The assembled snRNP particle is then re-imported into the nucleus, its site of function (Figure 1, B5).

In vitro studies on the assembly of snRNPs using purified Sm proteins and RNA had delineated several principles by which these structures form [13]. (1) Sm proteins do not exist as individual polypeptides; instead they are organized into three heterooligomers B-D3, D1-D2, and E-F-G. (2) None of the Sm protein heterooligomers are individually capable of stable association with snRNA. (3) The cooperative joining of D1-D2 and E-F-G with snRNA generates the earliest RNP assembly intermediate, the Sm subcore. (4) Subsequently, the addition of the B-D3 heterodimer leads to the conversion of the Sm subcore into the mature Sm core structure. (5) The structural information required for Sm core assembly is present in both the Sm protein and snRNA counterparts of this domain; hence they are capable of self-assembly. While these notions are no doubt true for the dilute, aqueous solutions in which the above experiments were performed, they called into question whether *in vivo* snRNPs are likewise generated along these aforementioned lines.

Insights into the assembly of snRNPs under *in vivo* conditions came from the functional characterization of the SMN protein [14]. This protein engages with eight other proteins (Gemins 2-8 and unrip) to form the SMN-complex [15–17]. The function of this complex is to recruit Sm proteins and snRNAs and mediate formation of the Sm core domain [18–20,21•]. After assembly, the SMN-complex has to be replenished with Sm proteins, which is assisted by the PRMT5-complex [22,23,24•]. This entity is composed of its name-giving constituent, the protein arginine methyltransferase 5 (PRMT5), WD45/MEP50, and pICln. The latter factor recruits Sm proteins and enables the symmetric methylation of a subset on designated arginine residues by PRMT5 (Figure 2a, 1 and 2).

The ordered interplay of the PRMT5-complex and the SMN-complex for snRNP assembly *in vivo* is puzzling, considering that Sm core formation occurs spontaneously *in vitro* using purified Sm proteins and snRNA. Consequently, it was believed that these assembly factors fundamentally modulate the reaction pathway of Sm core formation. A recent study has now elucidated the mechanism of the cellular snRNP assembly pathway [25••]. An intermediate with a sedimentation coefficient of 6S was identified and shown to consist of pICln and the Sm proteins D1-D2 and E-F-G (Figure 2a, 3). This complex is most likely identical to a 6S RNA-free precursor of snRNPs defined decades ago [26]. The Sm proteins B-D3 likewise associate with pICln (Figure 2a, 2). In contrast to the 6S complex, the pICln-B-D3 trimer appears to be either short-lived or remain stably bound to the PRMT5-complex.

The 6S complex neither alone nor along with the pICln-B-D3 complex could form RNPs [25••]. Electron microscopy of the 6S complex suggests that pICln occupies an identical position as the B-D3 heterodimer in the mature Sm core (Figure 2b). Hence, RNP formation appears to be inhibited by the steric hindrance of the RNA binding site on the inner cavity of the Sm protein ring. Based on its capacity to induce the formation of higher order Sm protein structures and prevent their premature association with RNA, pICln has the characteristics of an assembly chaperone. This class of factors has previously been described for proteinacous macromolecular complexes [27,28]. pICln, however, appears to be the first member of this class implicated in the formation of RNPs.

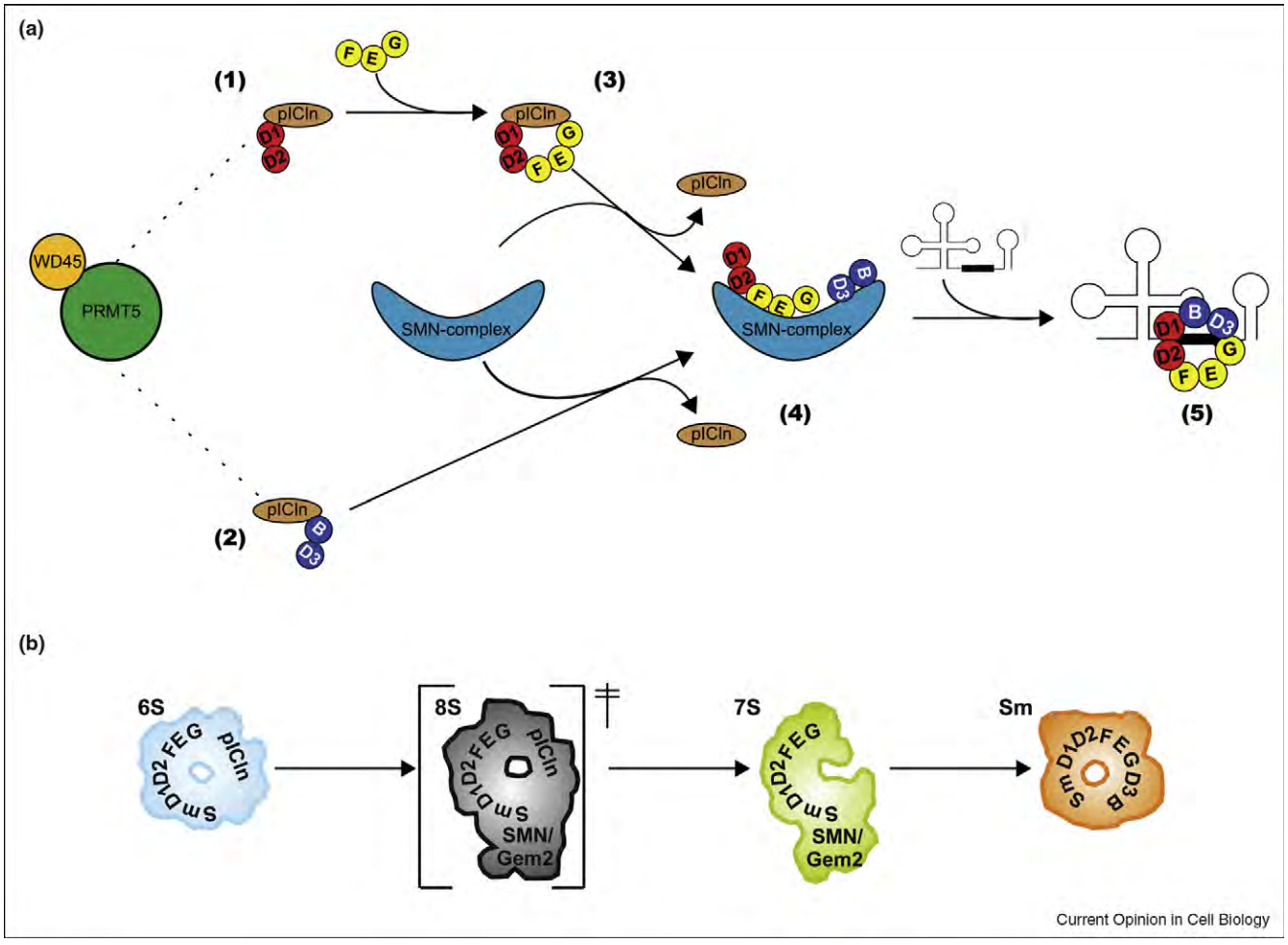
The formation of pICln-Sm protein complexes, therefore, introduces a roadblock in snRNP biogenesis by preventing the transfer of Sm proteins onto snRNA. The SMN-complex is required to release the Sm proteins from this pICln-imposed block and allow the assembly process to proceed. This is achieved by the transfer of Sm proteins from the 6S-complex and the pICln-B-D3-complex onto the SMN-complex and the simultaneous dissociation of pICln (Figure 2a, 4). Surprisingly, most subunits of the SMN-complex, with the exception of SMN and Gemin2, were dispensable for this reaction *in vitro*. These findings are reinforced by a recent study, in which bioinformatic analyses define a SMN-Gemin2 dimer as the most ancestral form of the SMN-complex [29•].

Electron microscopy elucidated further details of the transition of Sm proteins from the 6S complex to the SMN-complex (Figure 2b). This reaction proceeds through the initial docking of the SMN and Gemin2 subunits onto the outer surface of the 6S complex. As a consequence, pICln dissociates and the Sm proteins are arranged on the SMN-complex as an open clamp. In this open state Sm proteins have regained their competence to form snRNPs. The transfer of Sm proteins from the 6S complex to snRNPs appears to be prevented by a high activation energy barrier between both states. The fact that Sm proteins are reactivated by the SMN-complex suggests that the latter lowers the activation energy barrier for snRNP assembly. This is consistent with a role of the SMN-complex as a catalyst for Sm core formation.

Defects in mRNA metabolism as a cause for SMA?

How can the reduced expression of an snRNP assembly factor, SMN, which is required in all tissues, elicit the neuronal phenotype in SMA? It is noteworthy that several common as well as neuron-specific functions have been attributed to SMN. The motoneuron-degeneration in SMA has been suggested to be a consequence of the

Figure 2



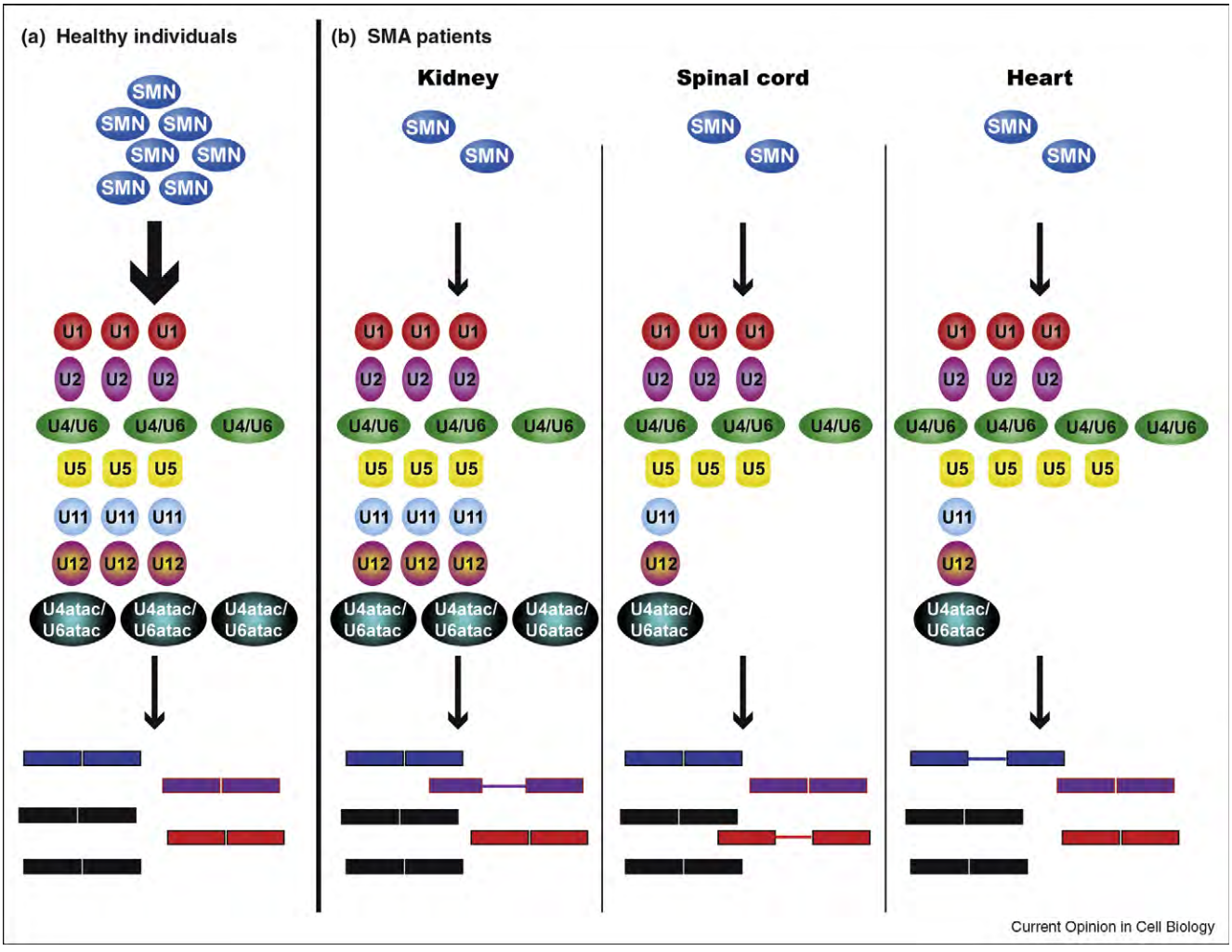
Model of the assisted assembly of snRNPs and structures of key intermediates. **(a)** The assembly chaperone pICln initially binds D1-D2 and D3-B (1 and 2). E-F-G is then recruited to pICln-D1-D2 to form the 6S complex (3). Transfer of the pICln-bound Sm proteins coincides with the displacement of pICln and forms the loaded SMN-complex (4). The SMN-complex recruits snRNA, catalyzes Sm ring closure, and releases the assembled RNP (5). **(b)** Structural EM models of intermediates of the assembly reaction. 6S (see also A3) forms a closed ring, 7S hold the Sm proteins in an open configuration (part of A4), and 8S represent a trapped transition complex from A3 to A4. The assembled core is derived from EM models of negatively stained snRNP cores.

reduced transport of ß-actin mRNA to growth cones of neurons or the impairment of actin metabolism [30,31]. These notions are attractive to explain the phenotype of SMA patients and certainly warrant further investigation. These findings have been extensively reviewed elsewhere and therefore will not be further discussed here [32,33]. Conversely, the finding that mature snRNPs rescue the neuronal phenotype in a zebrafish model of SMA [34•] raised the possibility that the defect in the general snRNP assembly pathway can likewise account for the clinical symptoms in SMA patients.

In accordance with this view, several studies reported markedly reduced snRNP assembly activity in SMN-deficient cells, tissues of SMA patients, and SMA mice [18,19,34•,35••,36••]. Rather unexpectedly, even under conditions where low levels of SMN nearly abolished the endogenous snRNP assembly activity, the total bulk amount of snRNPs in these cells was virtually unchanged [35••,36••]. This suggests that the capacity of cells to maintain the normal inventory of snRNPs far exceeds limiting amounts. By contrast, further reduction of SMN levels below a crucial threshold affected the steady state abundance of snRNPs. Under these conditions, however, not all snRNPs were equally affected; particularly components of the minor spliceosome appeared to be down-regulated. Intriguingly, the alteration of snRNP abundance was not uniform but differed in a tissue-specific manner (Figure 3b).

Expanding the results summarized above, Zhang *et al.* [35••] have studied the genome-wide impact of the altered repertoire of snRNPs on splicing with the help of exon specific microarrays in different tissues of SMA

Figure 3



Hypothetical scenario of the events leading to the SMA. **(a)** Situation in cells of healthy individuals: Normal levels of SMN ensure the generation of snRNPs in sufficient amounts to meet the demand of splicing in all cells of the body. **(b)** Situation in SMA patients: SMN-deficiency causes tissue-specific imbalance in the abundance of snRNPs, leading to varying splicing defects in different tissues. Motoneurons are particularly susceptible to this defect, giving rise to the SMA phenotype.

mice. This analysis led to the observation that profound alterations in the splicing pattern of many mRNAs occurred and affected genes with diverse functions.

Conclusions

The finding of splicing defects in SMN-deficient tissues may provide some initial clues why only motoneurons are affected in SMA. It is widely assumed that each cell type requires a balanced interplay of splicing factors to process all its transcripts properly. Because SMN-deficiency introduces an imbalance in spliceosomal components, the efficiency and fidelity of the splicing of certain introns might be influenced in a tissue-specific manner. Thinking along these lines, it is not far fetched to assume that the splicing of certain transcripts crucial to the survival of motoneurons is altered. The precise definition of this group of mRNAs is, hence, an important next step in understanding the molecular basis of this disease.

Acknowledgements

We thank Nils Neuenkirchen for comments on the manuscript. The authors apologize to colleagues, whose work could not be discussed owing to size constraints. This work was supported by Grant Fi-573-6/1.

References and recommended reading

Papers of particular interest published within the period of review have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Brow DA: **Allosteric cascade of spliceosome activation**. *Annu Rev Genet* 2002, **36**:333-360.

2. Patel AA, Steitz JA: **Splicing double: insights from the second spliceosome**. *Nat Rev Mol Cell Biol* 2003, **4**:960-970.

3. Will CL, Luhrmann R: **Spliceosomal UsnRNP biogenesis, structure and function**. *Curr Opin Cell Biol* 2001, **13**:290-301.

4. Matlin AJ, Clark F, Smith CW: **Understanding alternative splicing: towards a cellular code**. *Nat Rev Mol Cell Biol* 2005, **6**:386-398.

5. Black DL: **Mechanisms of alternative pre-messenger RNA splicing**. *Annu Rev Biochem* 2003, **72**:291-336.

6. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing**. *Nat Rev Genet* 2002, **3**:285-298.

7. Kostova FV, Williams VC, Heemskerk J, Iannaccone S, Didonato C, Swoboda K, Maria BL: **Spinal muscular atrophy: classification, diagnosis, management, pathogenesis, and future research directions**. *J Child Neurol* 2007, **22**:926-945.

8. Lefebvre S, Burglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M *et al.*: **Identification and characterization of a spinal muscular atrophy-determining gene**. *Cell* 1995, **80**:155-165.

9. Lorson CL, Hahnen E, Androphy EJ, Wirth B: **A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy**. *Proc Natl Acad Sci U S A* 1999, **96**:6307-6311.

10. Eggert C, Chari A, Laggerbauer B, Fischer U: **Spinal muscular atrophy: the RNP connection**. *Trends Mol Med* 2006, **12**:113-121.

11. Gabanella F, Carissimi C, Usiello A, Pellizzoni L: **The activity of the spinal muscular atrophy protein is regulated during development and cellular differentiation**. *Hum Mol Genet* 2005, **14**:3629-3642.

12. Neuenkirchen N, Chari A, Fischer U: **Deciphering the assembly pathway of Sm-class U snRNPs**. *FEBS Lett* 2008, **582**:1997-2003.

13. Raker VA, Plessel G, Luhrmann R: **The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro***. *EMBO J* 1996, **15**:2256-2269.

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

15. Carissimi C, Saieva L, Baccon J, Chiarella P, Maiolica A, Sawyer A, Rappsilber J, Pellizzoni L: **Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly**. *J Biol Chem* 2006, **281**:8126-8134.

16. Carissimi C, Saieva L, Gabanella F, Pellizzoni L: **Gemin8 is required for the architecture and function of the survival motor neuron complex**. *J Biol Chem* 2006, **281**:37009-37016.

17. Otter S, Grimmler M, Neuenkirchen N, Chari A, Sickmann A, Fischer U: **A comprehensive interaction map of the human survival of motor neuron (SMN) complex**. *J Biol Chem* 2007, **282**:5825-5833.

18. Wan L, Battle DJ, Yong J, Gubitz AK, Kolb SJ, Wang J, Dreyfuss G: **The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy**. *Mol Cell Biol* 2005, **25**:5543-5551.

19. Shpargel KB, Matera AG: **Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins**. *Proc Natl Acad Sci U S A* 2005, **102**:17372-17377.

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

21. • Pellizzoni L, Yong J, Dreyfuss G: **Essential role for the SMN complex in the specificity of snRNP assembly**. *Science* 2002, **298**:1775-1779.

This study provides evidence for a dual role of the SMN-complex in U snRNP biogenesis. One function of the SMN complex is the delivery of Sm proteins onto the U snRNA. In addition, the SMN-complex controls the specificity of the assembly reaction by preventing binding of Sm proteins onto non-cognate RNAs.

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

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

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

The authors reconstitute the assembly reaction of spliceosomal U snRNPs using purified components. The key finding of this study is that the PRMT5-complex and the SMN-complex form a functional entity that is sufficient for RNP formation. While the former acts in an early phase of the biogenesis pathway by committing Sm proteins to the assisted assembly pathway, the SMN complex acts as a molecular machine that facilitates RNP formation.

25. •• Chari A, Golas MM, Klingenhager M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H, Fischer U: **An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs**. *Cell* 2008, **135**:497-509.

In this study, the reaction mechanism of the assisted assembly of U snRNPs is elucidated by a combination of biochemical and structural approaches. The PRMT5-complex subunit pICln is identified as a novel type of assembly chaperone that organizes the Sm proteins in higher order complexes, while simultaneously preventing premature RNA association. The SMN-complex acts downstream of this chaperone system as a catalyst of RNP formation. Spatially, snRNP assembly appears to be similar to the way replication clamps are deposited onto DNA.

26. Fisher DE, Conner GE, Reeves WH, Wisniewolski R, Blobel G: **Small nuclear ribonucleoprotein particle assembly *in vivo*: demonstration of a 6S RNA-free core precursor and posttranslational modification**. *Cell* 1985, **42**:751-758.

27. Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M: **Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco**. *Cell* 2007, **129**:1189-1200.

28. Ramos PC, Dohmen RJ: **PACemakers of proteasome core particle assembly**. *Structure* 2008, **16**:1296-1304.

29. • Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U: **Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in *Drosophila melanogaster***. *PNAS* 2008, **105**:10045-10050.

Bioinformatic analyses using available whole-genome assemblies delineated the SMN-Gemin2 heterodimer as the most ancestral form of the SMN-complex. These *in silico* findings are corroborated by biochemical studies showing that *Drosophila melanogaster* contains a minimal SMN complex consisting of the same two subunits. Notably, this minimal unit is sufficient to perform and proof-read the RNP assembly reaction. These findings suggest that additional Gemin subunits present in higher eukaryotes fulfill other functions in the context of snRNP biogenesis.

30. Giesemann T, Rathke-Hartlieb S, Rothkegel M, Bartsch JW, Buchmeier S, Jockusch BM, Jockusch H: **A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with smn in nuclear gems**. *J Biol Chem* 1999, **274**:37908-37914.

31. Rossoll W, Jablonka S, Andreassi C, Kroning AK, Karle K, Monani UR, Sendtner M: **Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons**. *J Cell Biol* 2003, **163**:801-812.

32. Briese M, Esmaeili B, Sattelle DB: **Is spinal muscular atrophy the result of defects in motor neuron processes?** *Bioessays* 2005, **27**:946-957.

33. Monani UR: **Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease**. *Neuron* 2005, **48**:885-896.

34. • Winkler C, Eggert C, Gradl D, Meister G, Giegerich M, Wedlich D, Laggerbauer B, Fischer U: **Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy**. *Genes Dev* 2005, **19**:2320-2330.

Adopting the zebrafish model for spinal muscular atrophy (SMA), the authors study the underlying cause for the tissue-specific phenotype in the disease. Reduced expression of SMN elicits motoneuron degeneration, which can be rescued upon injection of exogenous snRNPs. These findings put forward the hypothesis that defects in the U snRNP assembly pathway affects the splicing machinery, which in turn, results in the pathological events leading to SMA.

35. Zhang Z, Lotti F, Dittmar K, Younis I, Wan L, Kasim M, Dreyfuss G:
•• **SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing**. *Cell* 2008, **133**:585-600.

This study illustrates that SMN deficiency causes tissue-specific and class-specific alterations in the abundance of individual snRNPs. The resulting consequence is a widespread splicing defect affecting predominantly pre-mRNAs containing a large number of introns. Notably, this defect occurs in a tissue-dependent manner. Along with the findings described in [34•] and [36••], these results suggest that SMA is a general splicing disease.

36. Gabanella F, Butchbach ME, Saieva L, Carissimi C, Burghes AH,
•• Pellizzoni L: **Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs**. *PLoS ONE* 2007, **2**:e921.

This study demonstrates for the first time that reduced expression of SMN causes a decrease in the levels of a subset of endogenous U snRNPs. Unexpectedly, components of the minor spliceosome appeared to be affected most. On the basis of these data the authors suggest that alterations in the cellular snRNP profile influence the splicing machinery in SMA patients.

Molekulare Maschinen

Molekulares Bodybuilding: Wie Zellen RNA-Proteinkomplexe herstellen

ASHWIN CHARI, UTZ FISCHER
THEODOR-BOVERI-INSTITUT, BIOZENTRUM, UNIVERSITÄT WÜRZBURG

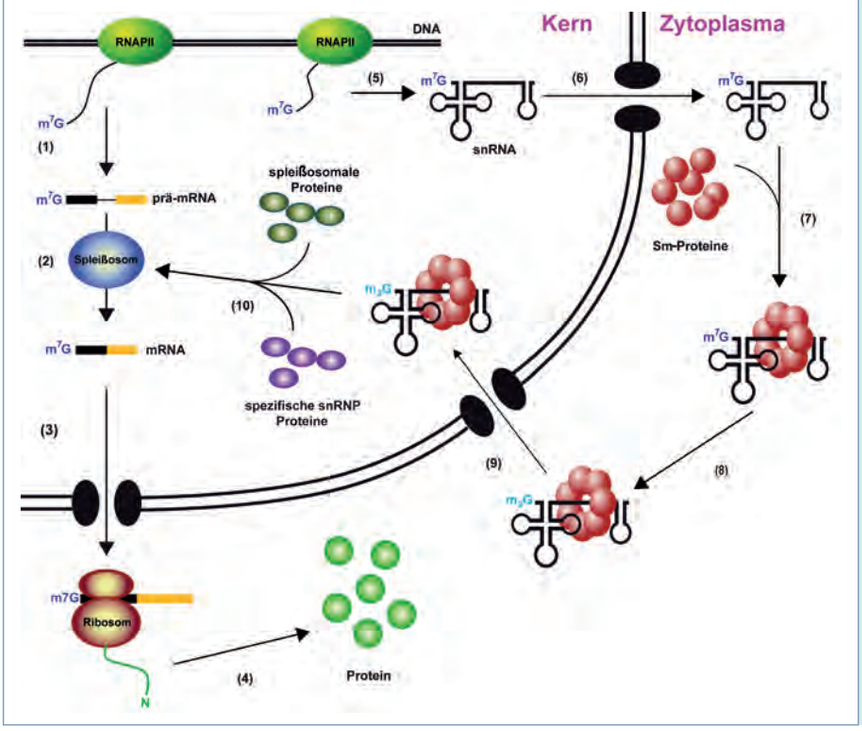
Viele RNAs lagern sich mit Proteinen zu komplexen funktionellen Einheiten zusammen. Entgegen der lange vorherrschenden Meinung, dass sich diese Zusammenschlüsse spontan bilden, mehren sich nun Hinweise, dass dieser Prozess häufig durch *assembly*-Chaperone gesteuert wird.

Many RNAs assemble with proteins to form functional units. Recent evidence suggests that these processes are mediated *in vivo* by a group of proteins termed assembly chaperones.

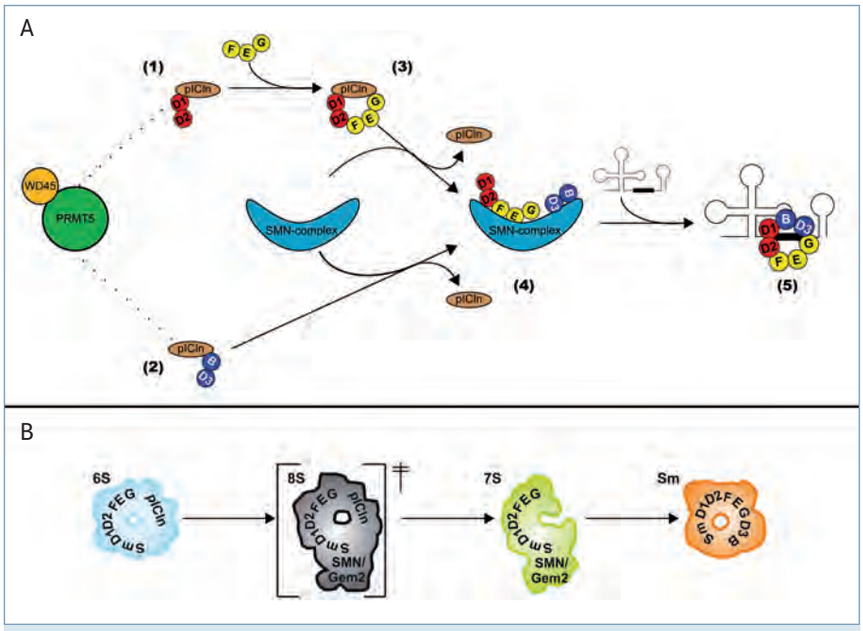
■ Eine große Anzahl von Prozessen, die in unseren Zellen ablaufen, wird von molekularen Maschinen ermöglicht. Diese Funktionseinheiten bestehen entweder ausschließlich aus Proteinen (z. B. das Protein-abbauende Proteasom) oder setzen sich aus Nukleinsäuren (häufig RNA) und Proteinen zusammen. Zur letzteren Klasse gehört das Spleißosom, eine große, dynamische Maschinerie, welche die Prozessierung von prä-mRNA-Molekülen in translatierbare mRNA katalysiert (**Abb. 1**, Schritte 1–4).

Weit über 100 Proteine und fünf kleine nicht-codierende RNAs sind Bestandteil menschlicher Spleißosomen. Diese Bestandteile sind in definierte Funktionseinheiten oder Module zusammengefasst, deren prominenteste Vertreter die snRNPs U1, U2, U4/6 und U5 sind. Hierbei handelt es sich um RNA-Proteinkomplexe, die eine bzw. im Falle von U4/6 zwei snRNAs und eine variable Anzahl von Proteinen enthalten [1].

Über den räumlichen Aufbau der U snRNPs ist heute einiges bekannt. So besitzen alle U snRNPs einen Satz von sieben gemeinsamen (Sm-)Proteinen, die man als B/B', D1, D2, D3, E, F und G bezeichnet. Diese evolutionär untereinander verwandten Proteine binden an eine definierte Region in den U snRNAs, bestehend aus einer einzelsträngigen Uridin-reichen Sequenz (Sm-Site), die von zwei Haarnadelschleifen flankiert wird. Die Sm-Proteine lagern sich so an die RNA an, dass sie einen Ring bilden, durch dessen Mitte die Sm-Site verläuft. Die daraus resultierende Struktur, das Sm-Core, bildet das Grundgerüst aller U snRNPs. Neben den gemeinsamen Proteinen binden an die einzelnen U snRNP-Spezies weitere, Partikel-spezifische Proteine. Diese sind häufig an definierten Funktionen des Spleißprozesses beteiligt, was unter ande-



◀ **Abb. 1:** Die Biogenese spleißosomaler snRNPs ist ein wichtiger Schritt in der eukaryotischen Genexpression. RNA-Polymerase II (RNAPII) transkribiert DNA-codierte Gene in prä-mRNA (1). Das Spleißosom katalysiert das Entfernen nicht-codierender Segemente (Introns) und das Verknüpfen codierender Segmente (Exons), um translatierbare mRNAs zu bilden (2). Die reife mRNA wird ins Zytoplasma exportiert (3). Die Translation von mRNAs am Ribosom führt zur Bildung von Proteinen (4). In der ersten Stufe der Biogenese von snRNPs, werden snRNA-Gene durch RNAPII transkribiert (5). Die snRNA wird danach ins Zytoplasma exportiert (6). Sm-Proteine assemblieren an die Sm-Site der snRNA (7). Die m^7G Cap-Struktur wird zur m_3G Cap-Struktur hypermethyliert (8). Das nun fertig assemblierte snRNP-Partikel wird nun in den Kern zurück transloziert (9). Nach der Zugabe Partikel-spezifischer und weiterer spleißosomaler Proteine kann das assemblierte snRNP-Partikel ins Spleißosom inkorporiert werden (10).



▲ **Abb. 2:** Modell für das zelluläre *assembly* von U snRNPs und Strukturen von Schlüsselintermediaten. **A,** Das *assembly*-Chaperon pICln bindet zunächst die Sm-Proteine D1-D2 und D3-B (1, 2). Die Rekrutierung der Sm-Proteine E-F-G bildet den sogenannten 6S-Komplex (3). Die Übertragung der Sm-Proteine auf den SMN-Komplex hat gleichzeitig die Dissoziation von pICln zur Folge (4). Der mit Sm-Proteinen beladene SMN-Komplex rekrutiert die U-snRNA, katalysiert die Schließung des Sm-Rings um die RNA und entlässt das assemblierte RNP (5). **B,** Elektronenmikroskopische Strukturmodelle von Intermediaten der *assembly*-Reaktion. Der 6S-Komplex ist ein geschlossener Ring (siehe A, 3), der 7S-Komplex (ein Teil des SMN-Komplexes; siehe A, 4) hält die Sm-Proteine als offenen Ring. Der 8S-Komplex stellt den Übergangszustand zwischen 6S und 7S dar. Das assemblierte Sm-Core ist auch dargestellt (Sm).

rem zu einer Arbeitsteilung der snRNPs im Spleißosom führt [2].

Die Biogenese der U snRNPs

Eine Frage, die Zellbiologen und Biochemiker gleichermaßen interessiert, ist, wie sich RNPs *in vivo* zu funktionsfähigen Einheiten zusammenlagern. Zur Beantwortung dieser grundsätzlichen Frage hat die Untersuchung der U snRNP-Biogenese wesentlich beigetragen.

Prinzipiell gilt es, bei der RNP-Zusammenlagerung ein logistisches Problem zu lösen: Die RNA-Komponente wird nämlich im Zellkern transkribiert, die Proteinbestandteile hingegen im Zytoplasma hergestellt. Aktive Transportprozesse müssen daher sicherstellen, dass RNA und Proteine zueinanderfinden. Im Falle der U snRNPs ist der Prozess der Zusammenlagerung primär ins Zytoplasma verlagert. Dementsprechend muss die snRNA zunächst vorübergehend aus dem Kern exportiert und dort mit den Sm-Proteinen beladen werden. Dieser Zusammenschluss ist Voraussetzung für diverse Modifikationsprozesse an der RNA, einschließlich der Methylierung der m^7G-Capstruktur zum hypermethylierten Trimethylguanosin($m^{2,2,7}$G)-Cap und dem anschließenden Import in den Kern zum Ort des Spleißens. Zu welchem Zeitpunkt in der Biogenese die spezifischen Proteine das Partikel vervollständigen ist in vielen Fällen noch nicht klar (**Abb. 1**, Schritte 5–9).

Der U snRNA-Export bzw. der U snRNP-Import wird durch jeweils spezialisierte Transportsysteme gewährleistet. Die Schlüsselfaktoren des Exports binden an den 5'-terminalen Bereich der U snRNA. Der Cap-Bindungskomplex (CBP20/80) bindet direkt an das m^7G-Cap der snRNA und ermöglicht so die Rekrutierung des Adapters PHAX und des Ran-GTP-gebundenen Exportfaktors CRM1. Dieser snRNA-Exportkomplex transloziert ins Zytoplasma, wo er durch Hydrolyse von GTP und Dephosphorylierung von PHAX zerfällt [3]. Der Kerntransport des zusammengelagerten U snRNPs benötigt hingegen die Importfaktoren Snurportin und Importin β, die das m_3G-Cap bzw. das Sm-Core erkennen [4]. Die Natur des Caps bestimmt daher maßgeblich, in welches Kompartiment die RNA transportiert werden kann: Hat sie ein m^7G-Cap, bedeutet das Export ins Zytoplasma; das m_3G-Cap erlaubt hingegen (zusammen mit dem Sm-Core) den Import des zusammengelagerten snRNPs in den Kern.

Die oben kurz beschriebenen Transportprozesse führen also die RNA und Proteinkomponenten an einem gemeinsamen Ort zusammen und schaffen damit die unmittelbare Voraussetzung für den Zusammenschluss zum RNP. Wie wird nun aber dieser *assembly*-Prozess gewährleistet? Aus früheren Studien am U snRNP wusste man, dass der Zusammenschluss *in vitro* spontan ablaufen kann. Dies bedeutet, dass die strukturellen Informationen für die Ausbildung des RNPs in den Komponenten selbst zu finden ist. Ähnliche *self-assembly*-Prozesse hatte man *in vitro* zuvor auch schon für Ribosomen beobachtet, weshalb man davon ausging, dass sich RNPs generell spontan ausbilden können. Nun muss man jedoch bedenken, dass der *assembly*-Prozess *in vivo* ganz andere Grundvoraussetzungen hat: Obwohl keine einzelne makromolekulare Spezies in hoher Konzentration vorliegt, ist dennoch die Gesamtkonzentration aller Proteine, RNAs und anderer Biomoleküle im zellulären Milieu extrem hoch (*molecular crowding*). Dies führt dazu, dass die Wahrscheinlichkeit einer produktiven Interaktion von Proteinen mit den kognaten RNAs beim RNP-*assembly* geringer, die für unspezifische Wechselwirkungen mit nicht-kognaten RNAs hingegen größer wird. Um dieser Tatsache Rechnung zu tragen, gibt es in der Zelle eine große Anzahl von assistierenden Faktoren (z. B. Chaperonen), die ungewollte Interaktionen verhindern und gewollte ermöglichen [5].

In der Tat hat man vor Kurzem eine Gruppe von trans-agierenden Faktoren entdeckt, welche die Zusammenlagerung der U snRNPs regulieren bzw. ermöglichen. Über deren Funktionsweise konnte in den letzten Jahren signifikante Fortschritte erzielt werden.

Die *assembly*-Maschinerie der U snRNPs

Zentral für diesen Prozess sind mindestens elf Proteine, die sich in zwei distinkten Komplexen organisieren. Diese als PRMT5- und SMN-Komplexe bezeichneten Einheiten wirken gezielt auf die Zusammenlagerung der Sm-Core-Domäne der U snRNPs [6–8].

Der ca. 1 MDa große SMN-Komplex besteht aus dem SMN(*survival motor neuron*)-Protein und sieben weiteren Faktoren, die als Gemine 2–8 bezeichnet werden. Seine Hauptfunktion ist es, die Sm-Proteine in einer definierten Konfiguration zu binden, und so den Transfer auf die snRNA zu ermöglichen. Die-

ser „*assembly*-Maschinerie“ ist ein Chaperonsystem vorgeschaltet, welches die Ausbildung von RNA-freien Sm-Proteinkomplexen induziert. Hierbei spielt das *assembly*-Chaperon pICln die wichtigste Rolle. pICln rekrutiert fünf der sieben Sm-Proteine und ordnet sie in den Positionen an, wie sie auch im Sm-Core zu finden sind. In diesem Zustand sind die Sm-Proteine aber nicht in der Lage, die snRNA zu binden; pICln verhindert dies durch eine sterische Blockade der RNA-Bindungsstelle in den Sm-Proteinen. Die Assoziation der Sm-Proteine mit dem *assembly*-Chaperon induziert daher eine kinetische Falle, aus welcher Sm-Proteine allein nicht in der Lage sind, das Sm-Core zu formieren. Um diese Blockade aufzulösen und die RNA-Bindung zu erlauben, bedarf es des SMN-Komplexes (**Abb. 2**). Dieser übernimmt die Sm-Proteine in der vom Chaperon angeordneten räumlichen Orientierung und vermittelt die Dissoziation des für die RNP-Bildung inhibitorischen Chaperons. Als Folge dieser Reaktion liegen die Sm-Proteine am SMN-Komplex als offener Ring vor. In diesem Zustand kann nun die RNA an das Innere des Rings binden. Anschließend versiegelt der SMN-Komplex die Sm-Proteine um die Sm-Site zu der für snRNPs charakteristischen Ringstruktur. Der Befund, dass der SMN-Komplex die durch pICn verursachte kinetische Falle aufzulösen vermag, ist konsistent mit einer Funktion des Ersteren als Katalysator für den *assembly*-Prozess (**Abb. 2**) [9].

Ausblick

Der humane SMN-Komplex besteht aus acht Untereinheiten. Für seine katalytische Aktivität im U snRNP-*assembly* sind allerdings nur zwei Untereinheiten, nämlich SMN und Gemin 2, notwendig. Dieser Befund wird dadurch untermauert, dass einige Spezies, wie z. B. das Insekt *Drosophila melanogaster*, einen SMN-Komplex mit nur diesen zwei Untereinheiten besitzen. Was also ist die Funktion der weiteren sechs Untereinheiten? Eine naheliegende und zu untersuchende Hypothese ist, dass sie andere Prozesse während der snRNP-Biogenese wie etwa die Hypermethylierung der Capstruktur der snRNA oder den Kernimport des snRNPs mit dem *assembly* koppeln.

Die wichtigste Frage, die es in naher Zukunft zu klären gilt, ist sicherlich, wie der SMN-Komplex den Transfer der Sm-Proteine auf die RNA katalysiert. Um diese Frage zu beantworten, kann man sich zunächst die Wirkungsweise verwandter makromolekularer Maschinen zu Hilfe ziehen. Wie oben erwähnt, bilden die Sm-Proteine im Sm-Core eine Ringstruktur. Eine topologisch ähnliche Struktur wird durch DNA-Klemmen (*clamps*) gebildet, um gering-prozessive DNA-Polymerasen auf ihrem DNA-Template zu halten. Analog zum Sm-Core, welches durch den SMN-Komplex assembliert wird, werden *clamps* aktiv durch sogenannte *clamp loader*-Enzymkomplexe auf DNA geladen. Der Ringschluss der Klammer um die DNA erfolgt durch eine ATP-getriebene Konformationsänderung des *clamp loader*. Die Ermittlung der Struktur des SMN-Komplexes wird in Zukunft die Frage klären, ob der SMN-Komplex ähnlich wie ein *clamp loader* wirkt, um Sm-Proteine auf snRNA zu laden. ■

Literatur

[1] Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718

[2] Will CL, Lührmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol 13:290–301

[3] Ohno M, Segref A, Bachi A et al. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101:187–198

[4] Neuenkirchen N, Chari A, Fischer U (2008) Deciphering the assembly pathway of Sm-class U snRNPs. FEBS Lett 582:1997–2003

[5] Ellis RJ, Minton AP (2006) Protein aggregation in crowded environments. Biol Chem 387:485–497

[6] Meister G, Bühler D, Pillai R et al. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat Cell Biol 3:945–949

[7] Meister G, Fischer U (2002) Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. Embo J 21:5853–5863

[8] Pellizzoni L, Yong J, Dreyfuss G (2002) Essential role for the SMN complex in the specificity of snRNP assembly. Science 298:1775–1779

[9] Chari A, Golas MM, Klingenhäger M et al. (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. Cell 135:497–509

Korrespondenzadresse:
Ashwin Chari
Prof. Dr. Utz Fischer
Lehrstuhl für Biochemie
Theodor-Boveri-Institut (Biozentrum)
Universität Würzburg
Am Hubland
D-97074 Würzburg
Tel.: 0931-888-4029
Fax: 0931-888-4028
ashwin.chari@biozentrum.uni-wuerzburg.de
utz.fischer@biozentrum.uni-wuerzburg.de

AUTOREN

Ashwin Chari (links), Utz Fischer (rechts)

Ashwin Chari
1999–2004 Studium der Biochemie, Molekularbiologie und Biophysik an der ETH Zürich. Seit **2004** Doktorand am Lehrstuhl für Biochemie, Theodor-Boveri-Institut (Biozentrum), Universität Würzburg.

Utz Fischer
1983–1988 Biochemiestudium an der Freien Universität Berlin. **1989–1992** Promotion an der Philipps-Universität Marburg. **1992–1997** Postdoctoral Fellow am IMT in Marburg und am Howard Hughes Medical Institute, Philadelphia. **1997–2003** Leiter einer unabhängigen Arbeitsgruppe am MPI für Biochemie, Martinsried. Seit **2003** Professor für Biochemie am Theodor-Boveri-Institut, Universität Würzburg.

7. Conclusions

The present thesis has elucidated the reaction mechanism of cellular U snRNP assembly. By this a first glimpse has been obtained how RNA-protein complexes are assembled in the crowded cellular environment. However, some aspects still remain open:

1) This work has illustrated that besides SMN and Gemin2 most subunits are sufficient for faithful U snRNP assembly. Nevertheless, they appear to be highly conserved in higher eukaryotes. Knockdown experiments have shown that these subunits are required for the viability of cells, which suggests that they fulfill essential functions. However what can these functions be? An attractive and testable hypothesis is that the remaining subunits of the SMN-complex couple preceeding (recognition of the newly exported U snRNA and dissociation of the U snRNA export complex) or follwing steps (hypermethylation of the cap and re-import into the nucleus) in U snRNP biogenesis with formation of the Sm core domain. The reconstitution schemes devised in this thesis will be instrumental to address this question in the near future.

2) Biochemical and structural characterization of the SMN-assisted U snRNP assembly reaction in this thesis has revealed that formation of the Sm core domain proceeds in a structurally similar manner as the clamp loading reaction. In this reaction macromolecular complexes (clamp-loaders) deposit protein clamps onto DNA to tether poorly processive DNA polymerases. The detailed understanding of the mechanism of clamp loading has strongly benefited from the elucidation of structures of the clamp-loader in a free and clamp bound state. Similarly, it is foreseeable that the elucidation of three-dimensional structures of the U snRNP assembly machinery will allow an unprecedented, detailed view on mechanistic aspects of this reaction. The prerequisite for these structural studies, the availability of homogenous complexes in sufficient amounts, has been accomplished in this thesis.

8. Appendix

Within the time frame of this thesis the author was engaged in further research, which will be presented in this appendix section.

Sm/LSm proteins form an evolutionary conserved family of proteins that bind to single-stranded RNA. The canonical Sm proteins bind to U snRNAs to form the Sm core domain. As discussed briefly in the introduction section the paralogous LSm2-8 proteins associate with the 3´end of U6 and U6atac snRNA and localizes to the nucleus. In contrast, the LSm1-7 complex, in which the LSm 8 subunit is exchanged for the LSm1 protein is located in the cytoplasm of cells, concentrated in foci termed P-bodies. The function of this heteroheptameric complex is to bind deadenylated mRNAs and promote their destruction by the recruitment of decapping factors, thus regulating mRNA half-life. The first part of this Appendix section deals with the unexpected role of the LSm 1-7 complex in the translation and replication of positive-strand RNA virus genomes (Chapter 8.1).

In the second part of the appendix section (Chapter 8.2) additional cellular roles of the PRMT5-complex in cells have been analyzed. The PRMT5-WD45 heterodimer is found to methylate components of the 3´ mRNA processing machinery. Specifically, the CFIm68 subunit of mammalian pre-mRNA cleavage factor I is methylated on designated arginine residues by PRMT5. For this to occur, the WD45 subunit appears to act as a specificity factor.

Distal Spinal Muscular Atrophy 1 (DSMA1) is a disease phenotypically related to Spinal Muscular Atrophy (SMA). However, in contrast to SMA the cause for DSMA1 are mutations in the IGHMBP2 gene. Analysis of the primary structure of the IGHMBP2 protein shows resemblance to RNA helicases. In Chapter 8.3, proof is provided that IGHMBP2 is a ribosome associated, bona fide RNA helicase. Furthermore, DSMA1 causing mutations are shown to impair the ATP-dependent helicase activity of this protein. These findings suggest that lack of translation-related helicase activity of IGHMBP2 is the underlying cause for disease in DSMA1 patients.

Finally, phosphorylation/ dephosphorylation events are important in signal transduction cascades. These reactions are catalyzed *in vivo* by kinases on either serine, threonine, or tyrosine residues. The main substrates for these kinases are often known. In a given signal

transduction cascade, however, kinases appear to have a plethora of different substrates, which are largely unknown. In chapter 8.4, a generic method is presented to find physiological substrates of any given protein kinase. This method is applied to a Drosophila p21-activated kinase and a bona fide substrate is found.

8.1 The Role of the LSm 1-7 Complex in the Translation and Replication of Positive-Strand RNA Virus Genomes

Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates

Scheller N, Mina LC, Galao RP, Chari A, Gimenez-Barcons M, Noueiry A, Fischer U, Meyerhans A, Diez J: **Proc Natl Acad Sci 2009**, Epub

Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formulation of results:	10 %

LSm1-7 complexes bind to specific sites in viral RNA genomes and regulate their translation and replication

Galao RP, Chari A, Alves-Rodriguez I, Lobao D, Mas A, Kambach C, Fischer U, Diez J: **RNA**, in revision

Thesis author´s contribution:

Conception:	40 %
Experimental contribution:	10 %
Formulation of results:	20 %

Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates

Nicoletta Scheller[a,b,1], Leonardo Bruno Mina[a,1], Rui Pedro Galão[a,1], Ashwin Chari[c], Mireia Giménez-Barcons[a], Amine Noueiry[d,2], Utz Fischer[c], Andreas Meyerhans[b], and Juana Díez[a,3]

[a]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain; [b]Institute of Virology, Saarland University, D-66421 Homburg, Germany; [c]Department of Biochemistry, Biocenter, University of Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany; and [d]Apath, LLC, St. Louis, MO 63110

Communicated by Paul Ahlquist, University of Wisconsin, Madison, WI, June 15, 2009 (received for review September 11, 2008)

Inevitably, viruses depend on host factors for their multiplication. Here, we show that hepatitis C virus (HCV) RNA translation and replication depends on Rck/p54, LSm1, and PatL1, which regulate the fate of cellular mRNAs from translation to degradation in the 5′-3′-deadenylation-dependent mRNA decay pathway. The requirement of these proteins for efficient HCV RNA translation was linked to the 5′ and 3′ untranslated regions (UTRs) of the viral genome. Furthermore, LSm1–7 complexes specifically interacted with essential *cis*-acting HCV RNA elements located in the UTRs. These results bridge HCV life cycle requirements and highly conserved host proteins of cellular mRNA decay. The previously described role of these proteins in the replication of 2 other positive-strand RNA viruses, the plant brome mosaic virus and the bacteriophage Qß, pinpoint a weak spot that may be exploited to generate broad-spectrum antiviral drugs.

deadenylation-dependent mRNA decay | HCV | host factors | LSm1–7 | Rck/p54

The astonishing diversity in viral life cycles, even inside the same viral group, raises intriguing questions about their origins and evolutionary relationships. Because viruses are obligatory intracellular parasites, they depend on host factors for their multiplication. The requirements for common host factors could provide essential clues about their evolutionary links and would also have important practical implications since these host factors might serve as targets for broad-spectrum antiviral strategies.

The group of positive strand RNA [(+)RNA] viruses encompass over one-third of all virus genera. It includes numerous and serious pathogens, a notable example being the hepatitis C virus (HCV), which is a major cause of chronic liver disease and has chronically infected ≈170 million individuals worldwide. At early times of infection, (+)RNA viral genomes perform 2 essential functions. They act as messengers for translation and as templates for viral replication. Because these 2 functions are mutually exclusive, a key step in the replication of all (+)RNA viruses is the regulated exit of their genomic RNA from the cellular translation machinery to the replication complex, which is always associated to intracellular membranes (1).

The replication of the plant Brome mosaic virus in the yeast *Saccharomyces cerevisiae* has proven to be a fruitful model system for studying common steps of (+)RNA virus biology in a relatively simple genetic background (2). By using this model system we have shown that the host factors LSm1, LSm6, and LSm7, which are subunits of the heptameric ring LSm1–7, as well as Pat1 and Dhh1 play an essential role in translation and in the translation-replication transit of the BMV genome (3–5). In noninfected cells, these proteins act as activators of decapping in the 5′-3′-deadenylation-dependent mRNA decay pathway (6). Although their precise way of functioning at the molecular level is not totally understood, they have been suggested to determine mRNA fate by facilitating the exit of cellular mRNAs from active translation to a translationally inactive state that allows the assembly of the decapping complex (6–8). These nontranslating mRNAs together with proteins involved in translation repression, mRNA decay, and RNA-mediated silencing accumulate in dynamic cytoplasmic foci referred to as P-bodies (review in refs. 9 and 10). Experiments in yeast indicate that mRNAs targeted to P-bodies can be either decapped and degraded or stored for return to translation (10).

Given the conservation of the 5′-3′-deadenylation-dependent mRNA decay pathway from yeast to humans and the common need of all (+)RNA viruses to regulate the transition of their genomes from active translation to a translationally inactive state to allow replication, an exciting possibility is that the function of Dhh1, LSm1–7, and Pat1 is used not only by BMV to replicate in yeast but also by human viruses to replicate in human cells. By measuring HCV replicon amplification and infectious virus production, we show here that indeed the respective human homologues namely Rck/p54, LSm1–7, and PatL1 (9, 11) are necessary for HCV replication. We also found that they are required for efficient translation of the viral genome and that these requirements are functionally linked to the 5′ and 3′ untranslated regions (UTRs). Furthermore, reconstituted LSm1–7 rings specifically bind to defined sequences in the 5′ and 3′UTRs that are known to play key roles in the regulation of HCV translation and replication. Together this not only demonstrates a conserved utilization of an ancient host cell machinery by the major human pathogen HCV but also opens up perspectives for the development of broadly reactive antiviral drugs.

Results

The Host Factors Rck/p54, LSm1, and PatL1 Promote HCV Replicon Amplification and Infectious HCV Production. To study whether Rck/p54, LSm1–7, and PatL1 affect HCV replication, we used a gene silencing strategy and used (i) HCV RNA replicons that allow efficient replication but do not result in virus production and (ii) infectious viruses that reproduce the entire virus life cycle (Fig. 1*A*). The HCV replicons HCVrep-Luc and HCVrep-Neo belong to the 1b genotype and are composed of the HCV

Author contributions: A.N., U.F., A.M., and J.D. designed research; N.S., L.B.M., R.P.G., A.C., and M.G.-B. performed research; and A.M. and J.D. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]N.S., L.B.M., and R.P.G. contributed equally to this work.

[2]Present address: Polsinelli, St. Louis, MO 63102.

[3]To whom correspondence should be addressed at: Universitat Pompeu Fabra, Department of Experimental and Health Sciences, Dr. Aiguader 88, 08003, Barcelona, Spain. E-mail: juana.diez@upf.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0906413106/DCSupplemental.

MICROBIOLOGY

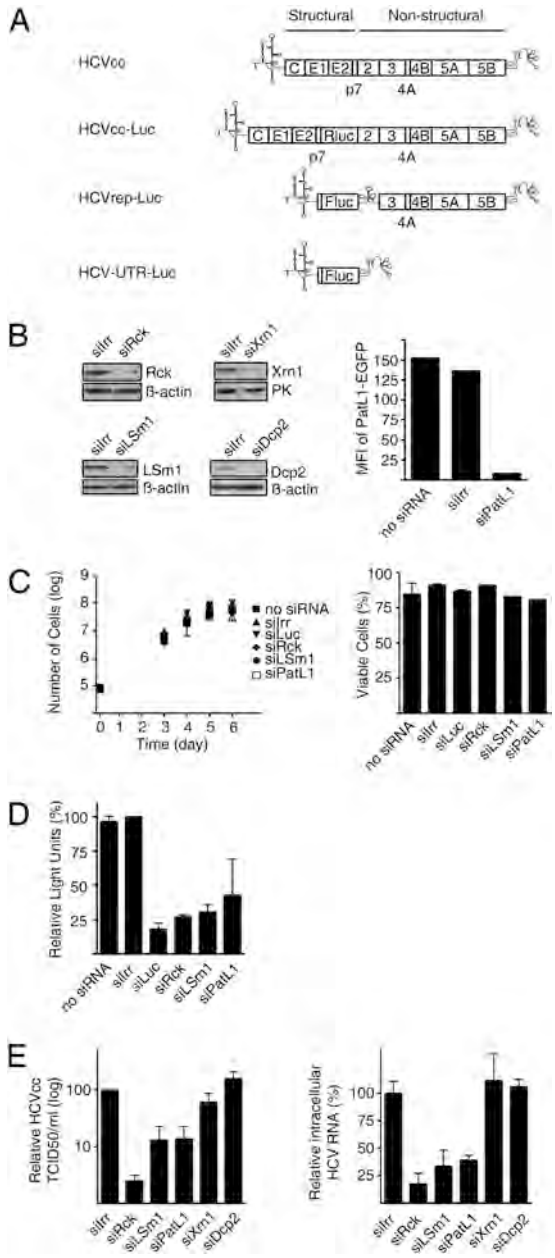


Fig. 1. Depletion of Rck/p54, LSm1 or PatL1 in hepatoma cell lines impairs HCV replication. (*A*) Schematic representations of the genomes of HCVcc, HCV Replicons and derivatives used in this study. (*B*) Huh7-Lunet cells were transfected with siRNA targeting Rck/p54, LSm1, PatL1, Xrn1, Dcp2, or a nontargeting siRNA (siIrr). Immunoblot analyses of Rck/p54, LSm1, Xrn1, Dcp2, β-actin, or pyruvate kinase levels are shown. Because no specific antibody is available for PatL1, to test PatL1 silencing, PatL1-EGFP expression plasmid and siRNAs were cotransfected and fluorescence was analyzed 1 day later by flow cytometry. Values are expressed in mean fluorescence intensity (MFI) (bar graph). Similar silencing results were obtained for Huh7.5 cells. (*C*) Cell growth of siRNA-transfected cells was followed for 6 days by counting the total number of cells (mean ± SEM; $n = 3$) (*Left*). The percentage of viable silenced cells at the day of maximum silencing was measured by propidium iodide staining (mean ± SEM; $n = 2$) (*Right*). (*D*) Huh7-Lunet cells were coelectroporated with the HCVrep-Luc replicon and the siRNAs. The percentage of relative luciferase light units compared with siIrr-transfected cells is shown at the day of most efficient silencing (mean ± SEM; $n = 2$). (*E*) Three days after transfection of silenced Huh7.5 cells with HCVcc RNA, the HCVcc infectivity in the supernatant was titrated by a limited dilution assay (*Left*). The accumulation of intracellular HCVcc mRNA was analyzed by quantitative RT-PCR (*Right*). Both values were normalized to the amount of transfected RNA (mean ± SEM; $n = 3$) and are shown relative to siIrr-transfected cells.

5′UTR, a luciferase reporter or neomycin phosphotransferase selection marker, the internal ribosome entry site (IRES) of the encephalomyocarditis virus (EMCV) followed by the HCV genes for the nonstructural proteins and the HCV 3′UTR. The infectious virus HCVcc has a 2a genotype and was used as such or with a luciferase reporter.

We first set up the silencing conditions for the cellular proteins by using specific siRNAs (Fig. 1*B*). With respect to the LSm1–7 cytoplasmic ring, we focused on the LSm1 subunit that defines the role of the ring in decapping. The other subunits when complexed with LSm8 are also part of a nuclear complex involved in splicing. In all cases, siRNA-mediated silencing resulted in a specific 80–85% reduction of the corresponding proteins when using the nontargeting siRNA siIrr as a reference (Fig. 1*B*). Importantly, silencing of Rck/p54, LSm1, and PatL1 did not affect cell growth or viability measured by sequential counting of the number of cells, propidium iodide staining or in an ATP assay (Fig. 1 *C* and *D* and Fig. S1). In addition, type I-interferons were not induced as judged by lack of MxA protein expression (Fig. S2).

To test whether silencing of any of these factors affects HCV replication, Huh7-Lunet cells were coelectroporated with HCVrep-Luc and a specific siRNA or with siIrr as a negative control. An additional mock-transfected control and a siRNA directed against the replicon-encoded luciferase gene (siLuc) were included. Luciferase values were then measured at times of maximum silencing (Fig. 1*D*). Down-regulation of Rck/p54, LSm1, and PatL1 resulted in a marked reduction of the luciferase activity by ≈80%, 70%, and 60%, respectively. Similar results were obtained with the HCVrep-Neo replicon (Fig. S3). This reduction was comparable to the 84% reduction observed by directly targeting the HCV replicon with siLuc. With transfection efficiencies of ≈90% and protein knockdowns of 80 to 85%, the values obtained are close to the maximal possible reduction. This strongly suggests that Rck/p54, LSm1, and PatL1 play an important role in HCV replicon amplification. Because the replicon system does not include RNA encapsidation, the observed effects can be explained by defects in HCV RNA translation and/or replication.

Next we tested whether this role is also detectable with an infectious HCV. At the time of maximal silencing, Huh7.5 cells were transfected with HCVcc RNA. Three days later, cellular supernatants were harvested for titration of infectious particles whereas intracellular HCV RNAs were quantified by quantitative RT-PCR (Fig. 1*E*). In all cases, HCV production from siRck-, siLsm1- and siPatL1-transfected cells was significantly reduced, the infectious titers being 50-, 10-, and 10-fold lower than in the siIrr control, respectively. Moreover, intracellular HCV RNA levels were also reduced. An inhibition in both, particle production and viral RNA accumulation, is expected for defects in an early step of the viral life cycle such as translation and replication. However, an additional effect on RNA encapsidation, particle morphogenesis or release cannot be excluded. Because all developed systems that allow to study HCV particle production depend on active translation and replication, this possibility was not explored further.

Depletion of the Proteins Dcp2 and Xrn1 Does Not Affect Infectious HCV Production. In the 5′-3′-deadenylation-dependent mRNA decay pathway, mRNA exit from translation and shortening of the poly(A) tail by deadenylases is followed by decapping via the Dcp1/Dcp2 decapping enzyme and 5′ to 3′ degradation via the exonuclease Xrn1 (6). To test the effect of some late components from this pathway on HCV replication, we selected Xrn1 and Dcp2. Silencing conditions were established, cell toxicity excluded (Fig. 1*B* and Fig. S1), and the effect on replication of HCVcc was assayed as before (Fig. 1*E*). No significant differences in the virus titer of the supernatants or in the level of the

intracellular HCV RNA between siXrn1-, siDcp2-, or siIrr-transfected cells were observed. These results argue that it is not the decapping and degradation process itself which is important for the HCV life cycle but the proteins acting upstream of it.

Rck/p54, LSm1–7, and PatL1 Affect Translation of the HCV RNA Genome via the 5′ and 3′UTRs. Rck/p54, LSm1 and PatL1 may affect HCV propagation by acting on HCV RNA translation, replication, or both. Most of the HCV proteins required for replication function in *cis*. As a consequence, one can measure either translation plus replication effects by using a replication-competent HCV derivative as above or only translation effects by using a nonreplicative HCV derivative. To investigate a putative role in translation, we used (i) a HCVrep-Luc replicon and (ii) a derivative of the HCVcc that contains the luciferase ORF fused to the NS2 gene (Fig. 1*A*). In both cases the NS5B polymerase carries a mutation that inhibits replication and, consequently, any luciferase activity of these derivatives can be attributed solely to translation of the transfected HCV RNA. Rck/p54-, LSm1-, or PatL1-silenced cells were transfected with the corresponding HCV RNAs and luciferase activities were measured 4 h later. When normalized to the abundance of intracellular HCV RNAs, activity reductions by ≈65%, 55%, and 48% were observed with the HCV replicon (Fig. 2*A*) whereas the reductions with HCVcc were 63%, 59%, and 79% (Fig. 2*B*), respectively. It is important to note, that the stability of HCV RNA was not significantly affected under these conditions (Fig. S4). By metabolic labeling we could exclude a generalized effect on cellular mRNA translation (Fig. 2*G*). In addition, translation of a luciferase mRNA flanked by a 5′cap and a 3′poly-(A) tail and with 5′and 3′UTR of nonviral origin was not affected by Rck/p54-, LSm1-, or PatL1- silencing (Fig. 2*F*). Since the major *cis*-signals controlling HCV RNA translation and replication are located in the 5′ and 3′UTRs of the HCV genome, we carried out a similar translation analysis with a genotype 1b HCV RNA derivative that contains only the HCV 5′UTR followed by a luciferase ORF and the HCV 3′UTR (Fig. 1*A*). The luciferase values were comparable to the ones obtained with the complete replicon (Fig. 2*C*). To test whether the observed translation inhibition depended on the HCV 5′UTR, HCV 3′UTR or both, we generated luciferase-reporter derivatives in which either the HCV 3′UTR was exhanged by a 3′poly(A)-tail or the HCV 5′UTR by a capped, unrelated 5′UTR. Silencing of Rck/p54, LSm1, and PatL1 had no significant effect on the translation of any of these RNAs (Fig. 2 *D* and *E*). In addition, EMCV-IRES mediated translation was also not significantly inhibited by silencing (Fig. S5). This suggests that HCV RNA translation specifically depends on Rck/p54, LSm1, and PatL1, and that this dependence is linked to the presence of both UTRs.

Reconstituted LSm1–7 Rings Bind Directly and Specifically to Translation/Replication Regulatory Signals in the HCV 5′ and 3′UTRs. At least 2 possible models can be considered by which Rck/p54, LSm1, and PatL1 can act on the HCV life cycle. First, silencing of these proteins may alter the host physiology thereby exerting a nonspecific effect on HCV replication. The toxicity tests performed in Rck-, LSm1-, and PatL1-silenced cells, however, render this possibility unlikely. Alternatively, these proteins may have a direct and specific effect on the virus and hence directly interact with viral RNA or proteins. In yeast cells, the corresponding proteins Dhh1, Pat1, and the LSm1–7 ring have been shown to interact in vivo (6), and there is evidence of a direct interaction of the LSm1–7 ring with deadenylated cellular mRNAs (8, 12). Considering a direct interaction model, it seemed possible that the LSm1–7 ring could interact with the 5′ and 3′UTRs of HCV since they are essential regions in the regulation of viral translation and replication (13), and our translation

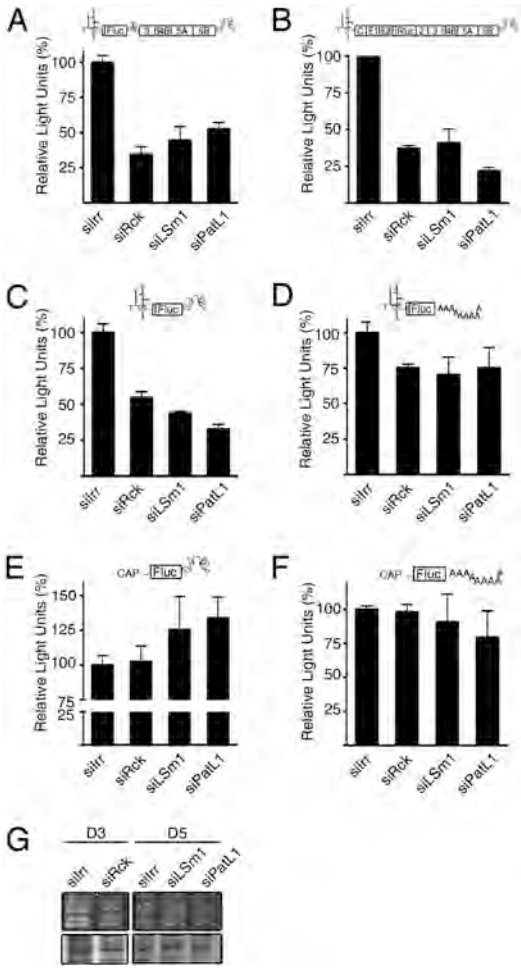


Fig. 2. Rck/p54, LSm1 and PatL1 silencing influences HCV RNA translation. Huh7-Lunet and Huh7.5 cells were transfected with siRNAs targeting Rck/p54, LSm1, PatL1, or a nontargeting siRNA, siIrr. The silenced cells were further transfected with (*A*) a nonreplicating bicistronic Luciferase replicon (HCVrep-Luc-GND), (*B*) a nonreplicating Luciferase-HCVcc (HCVcc-Luc-GNN), (*C*) a derivative (HCV-UTRs-Luc) containing the HCV 5′ and 3′UTRs from genotype 1b flanking the firefly luciferase ORF, (*D*) a derivative from HCV-UTRs-Luc in which the HCV 3′ UTR was exchanged by a poly(A) tail, (*E*) a derivative from HCV-UTRs-Luc in which the HCV 5′ UTR was exchanged by capped, nonviral 5′UTR, and (*F*) a derivative [CAP-Luc-Poly(A)] containing the 5′capped, nonviral 5′UTR followed by the firefly luciferase ORF and a poly(A) tail. The luciferase activity was measured 4 h after transfection and normalized to the respective intracellular RNA levels measured by quantitative RT-PCR (mean ± SEM; $n = 3$). (*G*) To examine the influence of Rck/p54-, LSm1- and PatL1-silencing on the synthesis of cellular proteins, silenced cells were labeled with [^{35}S]methionine for 30 min, separated on a denaturating polyacrylamide gel and visualized by autoradiography (*Lower*). Gels were coomassie-stained to visualize protein-loading (*Upper*).

results suggested a functional link to these sequences. To examine this possibility, we reconstituted functional human LSm1–7 rings according to a recently reported strategy (14) (Fig. S6), and performed electromobility shift assays with HCV RNA fragments (Fig. 3). Incubation of the LSm1–7 rings with the corresponding ^{32}P-labeled transcripts demonstrated strong binding to both UTR regions reflected by a complete band shift (Fig. 3 *B* and *C*). This binding was specific because addition of excess unlabeled 5′ or 3′UTR sequences resulted in binding competition (Fig. 3*D*), whereas addition of excess unlabeled nonbinding HCV RNA sequences did not.

To identify the viral RNA motifs involved in the interactions with the LSm1–7 rings, we systematically deleted domains of

MICROBIOLOGY

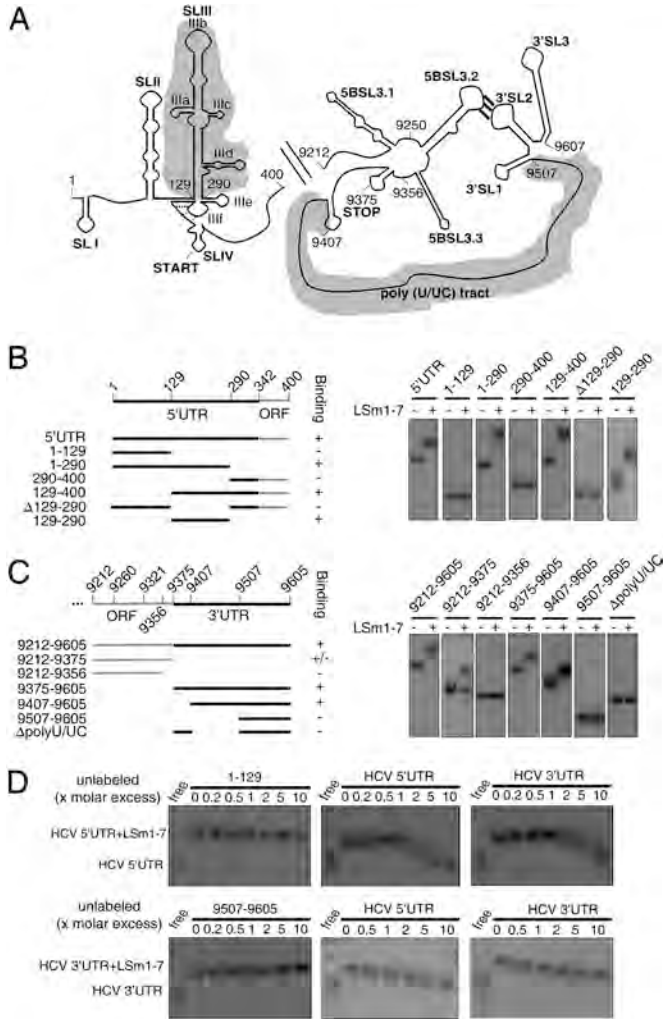


Fig. 3. Reconstituted LSm1–7 rings bind to specific HCV 5′ and 3′UTR regions. (*A*) Schematic representation of the secondary structures of the HCV 5′ and 3′ ends. Upstream of the 3′UTR, the NS5B coding sequence containing an RNA cruciform structure is shown. This structure includes the 5BSL3.2 loop. Its long-range interaction with the 3′SL3 loop in the 3′Xtail of the 3′UTR is essential for replication. Shadowed regions highlight the binding sites of the LSm1–7 rings. (*B* and *C*, *Left*) The constructs used in the electromobility shift assays are shown. The numbers refer to the nucleotide positions in the genome of the HCV Con1 strain. (*B* and *C*, *Right*) Radiolabeled, gel-purified RNA transcripts were incubated with the reconstituted LSm1–7 rings. After complex formation, products were separated on a nondenaturating polyacrylamide gel and visualized by autoradiography. (*D*) Labeled HCV 5′ and 3′UTR RNAs (HCV sequences 1–400 and 9,375–9,605, respectively) were incubated with reconstituted LSm1–7 rings in the presence of increasing amounts of unlabeled HCV 5′ and 3′UTR transcripts. As noncompeting controls, unlabeled RNAs negative for LSm1–7 ring binding (HCV sequences 1–129 and 9,507–9,605) were used. After complex formation, products were treated as in *B* and *C*.

defined RNA structure and function from the HCV 5′ and 3′UTRs respectively. The HCV 5′UTR contains the 4 stem loop structures SLI to SLIV (Fig. 3*A*). SLI is required in replication but is dispensable for translation. SLII, SLIII and SLIV form the internal ribosomal entry site. From these, SLIII is proposed to interact with the 40S ribosomal subunits thus playing a key role in translation initiation. The SLII and SLIII stem loops function in replication as well (13, 15). Electromobility shift analysis showed that the SLIII was both necessary and sufficient for binding of the LSm1–7 ring to the 5′UTR region (Fig. 3*B*) as evidenced by binding to this RNA motif (positions 129–290) and by loss of binding upon removal of it.

The HCV 3′UTR consists of a variable region, a poly (U/UC) tract and a highly conserved terminal region termed 3′X tail that consists of 3 stem loop structures (13). Importantly, the 3′UTR is not only required for replication but also for efficient translation (16). The electromobility shift analysis revealed a robust area of binding corresponding to the poly(U/UC) tract. Binding to the LSm1–7 ring was lost when this area was deleted (transcripts Δ polyU/UC) and gained when it was added (transcripts 9,507–9,605 and 9,407–9,605). The length and sequence composition of this region has recently been shown to have an important function in HCV RNA replication and the binding of host factors that could regulate this function has been suggested (17). An additional weak binding was observed to the NS5B coding sequence that includes a RNA cruciform structure (transcript 9,212–9,375). This binding was lost when the transcript was reduced to the positions 9,212–9,356. Because this remaining HCV sequence still contained the cruciform structure, either this was not the target of binding or the complete sequence might be required for proper folding to allow binding. Thus, in summary, the LSm1–7 ring binds robustly and directly to 2 important motifs in the 5′ and 3′UTR regions that are involved in the regulation of translation and replication of HCV.

Discussion

Our data demonstrate that HCV translation and replication decreases when the levels of Rck/p54, LSm1–7, and PatL1 are down-regulated. This was observed for the HCV genotype 1b and 2a for which replicons and viral derivatives were readily available. Most strikingly, the respective homologs play a similar role in BMV translation and replication in yeast, and Hfq, a homolog of LSm1 in bacteria, is required for the replication of the (+)RNA bacteriophage Qß (18). Furthermore, efficient retrotransposition of the yeast retrovirus-like elements Ty-1 and Ty-3 also depends on Dhh1, Lsm1, and Pat1 (19). One common theme for (+)RNA viruses and retroviruses is that both need to regulate the transition of the genomic RNA from translation to encapsidation (20). Thus, the dependence of at least some members of both virus groups on particularly those host proteins that are involved in the transit of cellular mRNAs from translation to another fate suggests that they have hijacked this function for their own benefit. Interestingly, Rck/p54, LSm1–7, and PatL1 are core components of P-bodies. These foci are sites where nontranslating mRNAs accumulate for different fates such as degradation, storage or returning to translation. Whether P-body formation itself is required for the HCV life cycle is an interesting issue (19) yet to be resolved in subsequent studies.

The function of LSm1–7 rings as activators of decapping of cellular mRNAs seems to involve their binding to short oligo (A) tracts at the 3′end of deadenylated cellular mRNAs. This binding then inhibits trimming of the 3′end while simultaneously promotes decapping and subsequent 5′ to 3′ degradation (8, 12, 21). However, the role of LSm1–7 rings on virus life cycles may be different because viral RNAs have different requirements for their eventual fates. In case of HCV RNA, the LSm1–7 rings are required for efficient translation. This function might be mediated by the direct interaction of the LSm1–7 rings with sequences in both, the 5′ and 3′UTR regions (Fig. 3). These interactions could facilitate rearrangements in the viral RNP structure and composition, recruiting proteins such as Rck/p54 and PatL1 from the cellular mRNA repression/decay machinery and, instead of promoting decay, might promote HCV RNA translation and subsequent transfer to replication. This view is consistent with a recent proposal made for the regulation of mRNAs generated by poxviridae. Viruses of this family generate viral mRNAs with an additional oligo(A) tract located at their 5′ends. Bergman and colleagues have shown that binding of LSm1–7 rings to such a tract at the 5′end of reporter RNAs does not result in mRNA decay but rather in RNA stabilization through inhi-

bition of decapping and degradation (22). This effect was proposed to be mediated by the simultaneous binding of Lsm1–7 rings to the 5′ and the 3′ends.

Silencing of Rck/p54, LSm1, or PatL1 affects HCV RNA translation and intracellular HCV RNA accumulation. This may be explained by an effect solely at the translation level or by an independent effect on both translation and replication as observed in the BMV model. Such an apparently antagonistic function, to promote both translation and exit from translation, is not without precedent as cellular proteins acting in 2 antagonistic processes such as translation initiation and translation repression have been described (23). An advantage of using a single complex for opposing outcomes seems to be the possibility of responding rapidly to different cellular requirements. A similar advantage might apply for the regulation of the viral life cycle.

In conclusion, the functional conservation of cellular and viral regulatory circuits across kingdoms and virus groups mark a weak spot that can be exploited for the generation of broad-spectrum antiviral drugs. Our observation that the individual, transient knock-down of Rck/p54, LSm1–7, and PatL1 proteins in human cells is not toxic and the fact that the respective yeast knockout strains are viable, stress the feasibility of such an approach for the future.

Materials and Methods

Plasmids, siRNA, and Antibodies. We present the plasmids, siRNA and the antibodies used in this study in *SI Text* and in Table S1 and Table S2.

In Vitro Transcription and Capping Reaction. In vitro transcripts of HCVcc, HCV replicons, and Luciferase reporter derivatives were performed by using RNAMaxx High Yield Transcription Kit (Stratagene) or MEGAScript Kit (Ambion) with T3 or T7 polymerase according to the manufacturer's instructions. After the in vitro transcription capped RNA was generated using the ScriptCap m^7G Capping System (Epicentre Biotechnologies). Transcripts used in electromobility shift assays were in vitro transcribed using T7 and SP6 polymerases (Fermentas GmbH) and labeled with [α-^{32}P]UTP.

Cell Culture, RNA Transfections, and Knockdown of Host Factors by RNAi. The Huh7.5 and Huh7-Lunet cells, subclones of the hepatoma cell line Huh7, were described (24, 25). Two different RNA transfection protocols were used, transfection by lipofectamine 2000 (Invitrogen) and electroporation (26). For silencing, 50 nM siRNA was optimal for Lipofectamine 2000 and 1 μM or 4 μM siRNA for electroporation. In all cases siRNA transfection efficiencies were determined 4 h after transfection using fluorescence-labeled siRNA and cytometry. The knockdown of Rck/p54, Xrn1, and Dcp2 required 1 siRNA transfection and efficient silencing was achieved 3 to 4 days after. The knockdown of Lsm1 required 2 to 3 successive transfections and efficient silencing was achieved 6 to 7 days after the initial transfection. A similar procedure was used for the transient knockdown of PatL1. The viability of the silenced cells was assessed by quantification of propidium iodide (PI) (MBL International,), by measurement of intracellular ATP-levels using CellTiterGlo (Promega) or by growth rate, counting cells up to 6 days after transfection. The ATP assay (Fig. S1) was used to analyze cell viability of the lipofectamine-transfected cells whereas growth rate and propidium iodide incorporation was used to analyze viability of the electroporated cells (Fig. 1*C*).

HCV-Replication Assays. Huh7-Lunet or Huh7 cells were coelectroporated with 10 μg yeast RNA, 1 μg of the corresponding replicon, and siRNA. For HCVrep-Luc, 1 μM of the corresponding siRNAs was used whereas for HCVrep-Neo, either 1 or 4 μM of siRNAs was used (27). Replication was measured either in colony-formation assays (HCVrep-Neo) or by quantification of intracellular replicon-encoded Luciferase (HCVrep-Luc and HCVrep-Luc-GND) as described (26, 28, 29).

To investigate the effect of the knockdown of the analyzed proteins on HCVcc replication, silenced cells were transfected by Lipofectamine with HCVcc RNA at the time of most efficient silencing. To maintain the protein knockdown of LSm1 and PatL1, an additional transfection with siRNA was required 24 h later. Intracellular HCVcc RNA levels and infectious HCVcc particles in the supernatant of transfected cells were quantified at various time points up to 72 h after transfection. The obtained values were standardized to the amount of transfected RNA quantified 4 h after transfection to equalize transfection efficiencies.

HCV-Translation Assays. Analysis of HCV translation was performed with different luciferase constructs. After transfection by Lipofectamine of the respective RNAs, luciferase activities were measured 4 h later and normalized to the total amount of protein. Then this value was corrected by the amount of the HCV RNA that was obtained by qRT-PCR using specific Taq-Man primers and probes (Table S3) after normalization to internal 18S RNA.

Titration of Infectious HCVcc Particles and RNA Quantification. Titration of infectious particles in the supernatant of HCVcc RNA-transfected cells was performed as described in ref. 28. For RNA quantification, 40 ng total RNA were reverse transcribed using random primers and SuperScript III according to manufacturer's recommendations (Invitrogen). The cDNA was amplified with specific primers and probes (listed in Table S3) using an ABI Prism 900HT sequence detection system (Applied Biosystems). The amplifications were standardized to an internal 18S control (ABI Taqman HS99999901 s1*; Applied Biosystems) using a relative quantification analysis from the SDS 2.3 software (Applied Biosystems).

Electromobility Shift Assays. Expression, purification of individual LSm proteins and reconstitution of complexes were performed as described in ref. 14. Three hundred cpm of gel-purified in vitro transcribed HCV-RNAs were incubated with 10 pmol LSm protein heptameric complexes in a buffer containing 20 mM Hepes-NaOH, pH 7.5, 200 mM NaCl, 2 mM $MgCl_2$, 0.1 U/μL RNasin, and 0.1 μg/μL yeast tRNA at 30 °C for 1 h. Samples were loaded on prerun 5% native polyacrylamide gels, and run at 4 °C for 2 h and 30 mA. Gels were autoradiographed on maximum sensitivity films (KODAK Biomax MS). For assays that included RNA competition, increasing amounts of RNA competitor were added to the reactions (Fig. 3*D*). The assays for binding to *Xenopus* U1 and U6 snRNAs were performed as described (14).

ACKNOWLEDGMENTS. We thank Apath LLC, R. Bartenschlager (Univ. of Heidelberg, Germany), O. Haller (Univ. of Freiburg, Germany), C. Kambach (Paur Scherrer Institut, Villigen, Switzerland), M. Kiledjian (Rutgers Univ., Piscataway, NJ), J. Lykke-Andersen (Univ. of Colorado, Boulder, CO), I.W. Mattaj (European Molecular Biology Laboratory, Heidelberg), F. Gebauer (Centre de Regulació Genòmica, Barcelona), S. Bradrick (Duke Univ. Medical Center, Durham, NC), and M. Niepmann (Justus-Liebig-Univ., Giessen, Germany) for reagents; and B. Lindenbach for critically reading the manuscript and for helpful discussions. This work was supported by Spanish Ministerio de Educación y Ciencia Grants BFU2004–00654 and BFU2007–66933/BMC, the Deutsche Forschungsgemeinschaft Me 1061/4 as part of the clinical research network KFO 129, SFB581, and FOR-855.

1. Gamarnik AV, Andino R (1998) Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes Dev* 12:2293–2304.
2. Alves-Rodrigues I, Galao RP, Meyerhans A, Diez J (2006) Saccharomyces cerevisiae: A useful model host to study fundamental biology of viral replication. *Virus Res* 120:49–56.
3. Diez J, Ishikawa M, Kaido M, Ahlquist P (2000) Identification and characterization of a host protein required for efficient template selection in viral RNA replication. *Proc Natl Acad Sci USA* 97:3913–3918.
4. Mas A, et al. (2006) Host deadenylation-dependent mRNA decapping factors are required for a key step in brome mosaic virus RNA replication. *J Virol* 80: 246–251.
5. Noueiry AO, et al. (2003) Yeast Lsm1p–7p/Pat1p deadenylation-dependent mRNA-decapping factors are required for brome mosaic virus genomic RNA translation. *Mol Cell Biol* 23:4094–4106.
6. Coller J, Parker R (2004) Eukaryotic mRNA decapping. *Annu Rev Biochem* 73:861–890.
7. Coller J, Parker R (2005) General translational repression by activators of mRNA decapping. *Cell* 122:875–886.
8. Tharun S, Parker R (2001) Targeting an mRNA for decapping: Displacement of translation factors and association of the Lsm1p–7p complex on deadenylated yeast mRNAs. *Mol Cell* 8:1075–1083.
9. Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007) P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* 27:3970–3981.
10. Parker R, Sheth U (2007) P bodies and the control of mRNA translation and degradation. *Mol Cell* 25:635–646.
11. Scheller N, et al. (2007) Identification of PatL1, a human homolog to yeast P body component Pat1. *Biochim Biophys Acta* 1773:1786–1792.
12. Chowdhury A, Mukhopadhyay J, Tharun S (2007) The decapping activator Lsm1p–7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *Rna* 13:998–1016.

13. Tellinghuisen TL, et al. (2007) Studying hepatitis C virus: Making the best of a bad virus. *J Virol* 81:8853–8867.
14. Zaric B, et al. (2005) Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* 280:16066–16075.
15. Isken O, et al. (2007) Nuclear factors are involved in hepatitis C virus RNA replication. *RNA* 13:1675–1692.
16. Song Y, et al. (2006) The hepatitis C virus RNA 3′-untranslated region strongly enhances translation directed by the internal ribosome entry site. *J Virol* 80:11579–11588.
17. You S, Rice CM (2008) 3′ RNA elements in hepatitis C virus replication: Kissing partners and long poly(U). *J Virol* 82:184–195.
18. Franze de Fernandez MT, Eoyang L, August JT (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* 219:588.
19. Beckham CJ, Parker R (2008) P bodies, stress granules, and viral life cycles. *Cell Host Microbe* 3:206–212.
20. Ahlquist P (2006) Parallels among positive-strand RNA viruses, reverse-transcribing viruses, and double-stranded RNA viruses. *Nat Rev Microbiol* 4:371–382.
21. He W, Parker R (2001) The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3′ termini from partial degradation. *Genetics* 158:1445–1455.
22. Bergman N, et al. (2007) Lsm proteins bind and stabilize RNAs containing 5′ poly(A) tracts. *Nat Struct Mol Biol* 14:824–831.
23. Abaza I, Gebauer F (2008) Trading translation with RNA-binding proteins. *RNA* 14:404–409.
24. Quinkert D, Bartenschlager R, Lohmann V (2005) Quantitative analysis of the hepatitis C virus replication complex. *J Virol* 79:13594–13605.
25. Blight KJ, McKeating JA, Rice CM (2002) Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol* 76:13001–13014.
26. Lohmann V, Korner F, Dobierzewska A, Bartenschlager R (2001) Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *J Virol* 75:1437–1449.
27. Wilson JA, Richardson CD (2005) Hepatitis C virus replicons escape RNA interference induced by a short interfering RNA directed against the NS5b coding region. *J Virol* 79:7050–7058.
28. Lindenbach BD, et al. (2005) Complete replication of hepatitis C virus in cell culture. *Science* 309:623–626.
29. Lohmann V, et al. (2003) Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J Virol* 77:3007–3019.

Supporting Information

Scheller et al. 10.1073/pnas.0906413106

SI Text

HCV Replicons and Viruses. The plasmids pFKi389neoNS3–3′ ET, pFKi389FLucNS3–3′ ET and pFKi389FLucNS3–3′GND were kindly provided by Ralf Bartenschlager (University of Heidelberg, Germany) and used for generating the corresponding HCV Con1 (Genotype 1b) strain (GenBank accession number AJ238799) RNA replicons HCVrep-Neo, HCVrep-Luc, and HCVrep-Luc-GND, respectively. These constructs carry, besides the reporter-genes neomycin phosphotransferase (HCVrep-Neo) or firefly (*Photinus pyralis*) luciferase (HCVrep-Luc and HCVrep-Luc-GND), the highly adaptive mutations in NS4B (K1846T) and NS3 (E1202G+T1280I) (1). The plasmids pFL-J6/JFH1/JC1 and pFL-J6/JFH/JC1p7Rluc2a (GNN) used to generate HCVcc and HCVcc-Luc-GNN RNA were kindly provided by Apath LLC. Both contain the 5′UTR and nonstructural regions NS3 through NS5B from the JFH1 (Genotype 2a) strain (GenBank accession number AB047639) and structural regions through NS2 and the 3′UTR from the J6 (Genotype 2a) strain (GenBank accession number AF177036). pFL-J6/JFH/JC1p7Rluc2a (GNN) contains a renilla luciferase (*Renilla reniformis*) gene fused in frame with the N terminus of J6 NS2. The GND and the GNN mutation in the active site of the NS5B gene in pFKi389FLucNS3–3′GND and pFL-J6/JFH/JC1p7Rluc2aGNN impair replication (2).

Luciferase-reporter Derivatives. The luciferase reporter construct HCV-3′UTR, kindly provided by Michael Niepmann (3), was used to generate HCV-UTRs-Luc and contains the Con1 5′UTR and 3′UTR flanking a firefly luciferase gene. The plasmid LUC cassette clon 2 was a gift from Fatima Gebauer and used to generate CAP-Luc-Poly(A), which contains a capped, nonviral 5′UTR followed by the firefly luciferase gene and a poly(A) tail of 73 nucleotides. To generate the derivative HCV-5′UTR-Luc-Poly(A) containing the HCV 5′UTR from genotype 1b and the firefly luciferase ORF followed by the poly(A) tail, the *Bam*H1/*Eco*RV fragment from the HCV-3′UTR was treated with Mung Bean Nuclease and cloned into the LUC cassette clon2 digested with *Sma*I and *Eco*RV. The derivative CAP-Luc-HCV-3′UTR was used to generate the 5′capped firefly luciferase transcript followed by the HCV 3′UTR. It was obtained by exchanging the *Hin*dIII/*Eco*RV fragment from the LUC cassette clon2 by the *Hin*dIII/*Eco*RV fragment from the Con1 3′UTR. The renilla luciferase reporter construct pEMCV(A50) was kindly provided by Shelton Bradrick (4) and used to generate EMCV-IRES-Luc-Poly(A) which contains the EMCV-IRES followed by the renilla luciferase ORF and a poly(A) tail of 50 nucleotides.

Constructs Used in Electromobility Shift Assays. RNA sequences of interest were amplified by PCR using a complete Con1 HCV genome (GenBank accession number AJ238799) as DNA template. Forward primers contain an *Eco*RI restriction site at the 5′ end followed by the T7 promoter and the HCV sequence. Reverse primers contain a *Bam*HI restriction site at the 3′ end. The resulting *Eco*RI/*Bam*HI PCR-fragments were subcloned into pUC18 (New England Biolabs). The oligonucleotide sequences used for amplifying the different areas are given in Table S1. The DNA fragments Δ129–290 and ΔpolyU/UC were obtained by a 2-step PCR procedure. For Δ129–290, the fragments 1–129 and 290–400 were amplified and then used as templates for another amplification round with the primers HCV pos1 fwd and HCV 400 rev. For ΔpolyU/UC, the fragments 9,375–9,407 and 9,507–9,605 were amplified first and then combined and reamplified in a second PCR with the oligonucleotides 9375fwd and 9605rev. The plasmid used to transcribe nt 9,212–9,356 of HCV RNA (pBSK-T7–3.1–3.3) was kindly provided by Ralf Bartenschlager. The plasmids used for expression and purification of individual LSm proteins and for transcription of the *Xenopus* U6 and U1 snRNA were already described (5, 6).

siRNAs and Antibodies. The siRNAs specific for Rck/p54, LSm1, PatL1, Xrn1, and Dcp2 mRNA were purchased from Ambion or Dharmacon. The sense strand sequences of the siRNAs are given in Table S2. The following controls were used: fluorescent-labeled-siRNA and siRNAs specific for Renilla Luciferase (siRLuc) were both from Ambion; siRNA specific for Firefly Luciferase (anti-Luc siRNA 1) and nontargeting siRNA # 1 (siIrrD) were both from Dharmacon. Anti-NS5a antibody was kindly provided by Apath LLC. The anti-LSm1, anti-Rck/p54, mouse anti-β-actin and anti-pyruvate kinase antibodies were purchased by Genway, Biozol, Sigma–Aldrich, and Chemicon, respectively. Polyclonal rabbit anti-Xrn1, rabbit anti-Dcp2 and mouse anti-MxA antibodies were already described and kindly provided (7–9).

1. Lohmann V, Hoffmann S, Herian U, Penin F, Bartenschlager R (2003) Viral and cellular determinants of hepatitis C virus RNA replication in cell culture. *J Virol* 77:3007–3019.
2. Lohmann V, Roos A, Korner F, Koch JO, Bartenschlager R (2000) Biochemical and structural analysis of the NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *J Viral Hepat* 7:167–174.
3. Song Y, et al. (2006) The hepatitis C virus RNA 3′-untranslated region strongly enhances translation directed by the internal ribosome entry site. *J Virol* 80:11579–11588.
4. Bradrick SS, Dobrikova EY, Kaiser C, Shveygert M, Gromeier M (2007) Poly(A)-binding protein is differentially required for translation mediated by viral internal ribosome entry sites. *RNA* 13:1582–1593.
5. Zaric B, et al. (2005) Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* 280:16066–16075.
6. Hamm J, Darzynkiewicz E, Tahara SM, Mattaj IW (1990) The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* 62:569–577.
7. Fenger-Gron M, Fillman C, Norrild B, Lykke-Andersen J (2005) Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol Cell* 20:905–915.
8. Wang Z, Jiao X, Carr-Schmid A, Kiledjian M (2002) The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc Natl Acad Sci USA* 99:12663–12668.
9. Flohr F, Schneider-Schaulies S, Haller O, Kochs G (1999) The central interactive region of human MxA GTPase is involved in GTPase activation and interaction with viral target structures. *FEBS Lett* 463:24–28.
10. Stoecklin G, Mayo T, Anderson P (2006) ARE-mRNA degradation requires the 5′-3′ decay pathway. *EMBO Rep* 7:72–77.
11. Scheller N, et al. (2007) Identification of PatL1, a human homolog to yeast P body component Pat1. *Biochim Biophys Acta* 1773:1786–1792.
12. Kronke J, et al. (2004) Alternative approaches for efficient inhibition of hepatitis C virus RNA replication by small interfering RNAs. *J Virol* 78:3436–3446.
13. Chen LS, Tassone F, Sahota P, Hagerman PJ (2003) The (CGG)n repeat element within the 5′ untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum Mol Genet* 12:3067–3074.

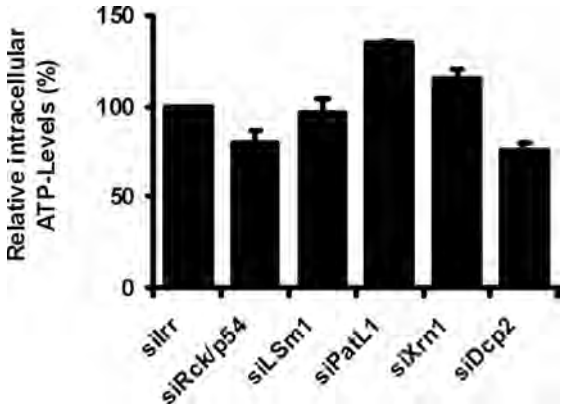


Fig. S1. Relative intracellular ATP levels after knockdown of Rck/p54, LSm1, PatL1, Xrn1, and Dcp2. The intracellular levels of ATP were measured in Huh7.5 cells at the time of most efficient knockdown of Rck/p54, LSm1, Xrn1, or Dcp2 by siRNA transfection. The ATP levels of control siRNA (siIrr) -transfected cells were set to 100%. For PatL1, the ATP levels were measured at the time point of most efficient knockdown of LSm1 (mean ± SEM; $n = 3$).

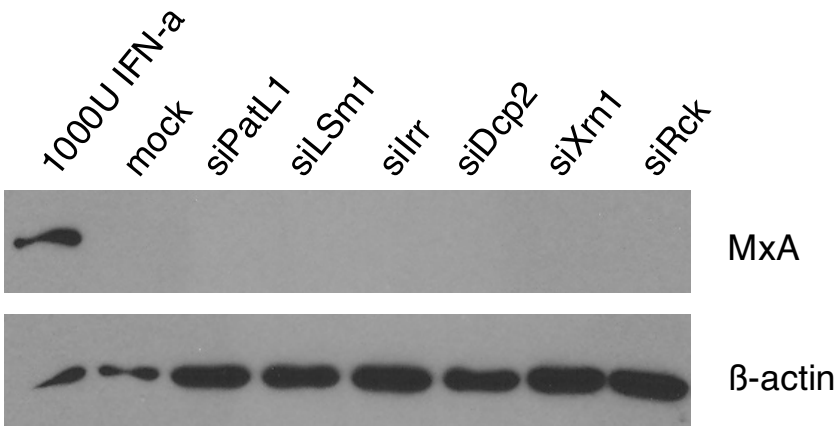


Fig. S2. Transient knockdown of Rck/p54, LSm1, PatL1, Dcp2, and Xrn1 does not induce MxA expression. Huh7.5 cells were transfected with siRNA targeting Rck/p54, LSm1, PatL1, Xrn1, Dcp2, or a nontargeting siRNA (siIrr). At the time point of most efficient protein-knockdown, levels of intracellular MxA expression were visualized by immunoblot showing MxA and β-actin. Controls were left untreated (mock) or were incubated 24 h with 1,000 U/mL of IFN-α.

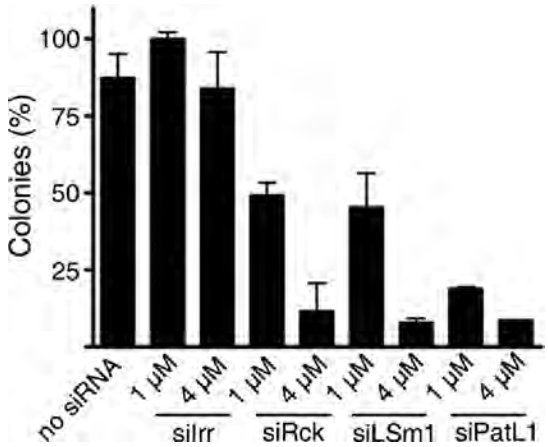


Fig. S3. Effect of transient knockdown of Rck/p54, LSm1, and PatL1 on HCV replication measured by a colony-formation assay. Huh7 cells were cotransfected with the in vitro-transcribed HCVrep-Neo and control siRNA (siIrr), siRNA specific for Rck/p54, LSm1, or PatL1 (1 μM or 4 μM). G418-resistant colonies were quantified 3 weeks later. The percentage of G418-resistant colonies compared with siIrr-transfected cells is shown (mean $\pm$ SEM).

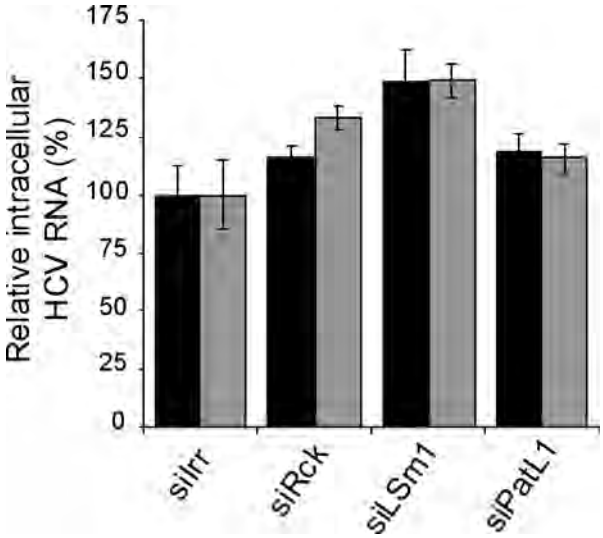


Fig. S4. Effect of transient knockdown of Rck/p54, LSm1, and PatL1 on HCV RNA stability measured by quantitative RT-PCR specific for the 5′UTR and 3′UTR of HCV. Huh7-Lunet cells were transfected with siRNAs targeting Rck/p54, LSm1, PatL1, or with a nontargeting siRNA, siIrr. The silenced cells were further transfected with a RNA derivative (HCV-UTRs-Luc) containing the 5′ and 3′ HCV UTRs from genotype 1b flanking the luciferase ORF. The accumulation of intracellular HCV mRNA was analyzed 4 h after transfection by quantitative RT-PCR specific for the 5′UTR (black bars) or the 3′UTR (gray bars) of HCV. Both values are shown relative to siIrr-transfected cells (mean ± SEM, $n = 3$). Similar results were obtained when silenced cells were transfected with HCV replicon RNA or HCVcc RNA.

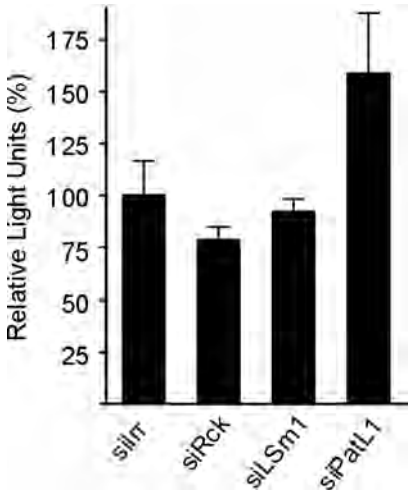


Fig. S5. Effect of transient knockdown of Rck/p54, LSm1, and PatL1 on Encephalomyocarditis virus (EMCV)-IRES mediated translation. Huh7-Lunet were transfected with siRNAs targeting Rck/p54, LSm1, PatL1, or with a nontargeting siRNA, siIrr. The silenced cells were further transfected with a EMCV-IRES-Luc-Poly(A), a mRNA containing the EMCV-IRES follow by the renilla luciferase ORF and a poly(A) tail. The luciferase activity was measured 4 h after transfection and normalized to relative intracellular RNA levels measured by quantitative RT-PCR (mean ± SEM, $n = 3$).

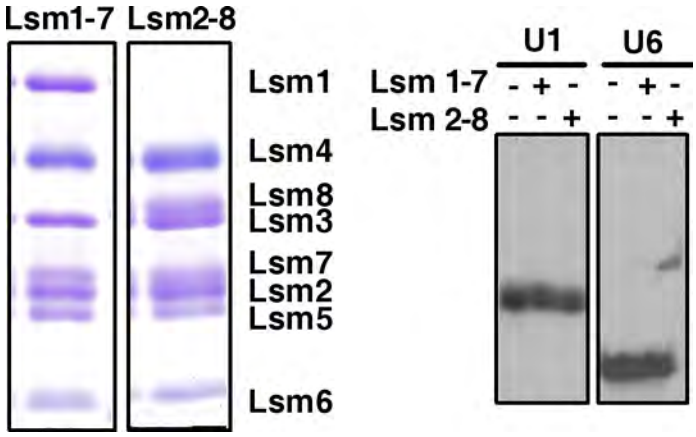


Fig. S6. Analysis of the reconstituted LSm1–7 and LSm2–8 complexes. (*Left*) Comassie-stained SDS/PAGE showing the separated components of the purified, reconstituted LSm1–7 and LSm2–8 complexes. (*Right*) Since no specific binding site for the LSm1–7 ring had yet been identified, binding specificity of complexes were determined, as described in ref. 5, by in vitro electromobility shift assays with [α^{32}P]-UTP labeled U1 and U6 snRNAs. As expected, the reconstituted LSm2–8 rings bind to U6 but not to U1 snRNAS whereas the Lsm1–7 ring did not bind to either of the 2 snRNAs.

Table S1. Primer Sequences used to generate constructs for the electromobility shift assays

Amplified UTR Region	Primer Name	Sequence from 5′ to 3′ end
5′ UTR, 1–129, 1–290, Δ129–290	HCVpos1fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GCC AGC CCC
5′ UTR, 129–400, 290–400, Δ129–290	HCV 400 rev	G**GG ATC C**TG TGG GCG GCG GTT GGT GTT AC
1–129	HCV 129 rev	G**GG ATC C**GG AGG GGG GGT CCT GGA G
1–290	HCV 290 rev	G**GG ATC C**AG TAC CAC AAG GCC TTT CGC G
129–400	HCV 129 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GCC GGG AGA
290–400	HCV 290 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GTG CCT GAT A
1–129	ΔSLIII 129 rev	AAG CAC CCT ATC TGG CAG GAG GGG GGG GTC C
290–400	ΔSLIII 290 fwd	TGC CTG ATA GGG TGC TTG CGA GTG
9375–9605, 9375–9407, ΔpolyU/UC	HCV 9375 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GAC GGG GAG C
9212–9605, 9375–9605, 9407–9605, ΔpolyU/UC	HCV 9605 rev	G**GG ATC C**AC TTG ATC TGC AGA GAG GCC
9407–9605	HCV 9407 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GAT CCT GTT
9507–9605	HCV 9507 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GTG GT GGC TCC
9375–9407	ΔpolyU/UC 9407 rev	GCT AAG ATG GAG CCA CCA GAA AAA AAG AGG AT
9507–9605	ΔpolyU/UC 9507 fwd	TGG TGG CTC CAT CTT AGC CTA GTC
9212–9605, 9212–9375	HCV 9212 fwd	GG**G AAT TC***G TAA TAC GAC TCA CTA TAG* GAA TCC CGG CTG CGT TCC
9212–9375	HCV 9375 rev	G**GG ATC C**TT CAT CGG TTG GGG AGT AGA

Italicized: T7 promoter; bold: restriction sites.

Table S2. Sense-strand sequences of siRNAs used in this study

Targeted gene	siRNA name	Sequence from 5′ to 3′ end	Reference
LSm1	siL1	CAA ACU UAG UGC UAC AUC AdTdT	9
Rck/p54	si5076	GGA GGA GAG CAU UCC CAU UdTdT	10
PatL1	siPat #1	CUA GAA GAU CCA GCU AUU AdTdT	10
Xrn1	siXrn1	AGA UGA ACU UAC CGU AGA AdTdT	9
Dcp2	siDcp2	GAA AUU GCC UUG UCA UAG AdTdT	9
HCV-Con1-sequence	siHCV	CCU CAA AGA AAA ACC AAA CdTdT	11

Table S3. Oligonucleotides and Oligonucleotide-probes used for quantitative RT-PCR

Construct	Targeted sequence	Primers and probes	Sequence from 5′ to 3′ end with indicated modifications	Reference (this study unless noted)
HCVcc		Primer sense	TCTTCACGCAGAAAGCGTCTAG	
HCVcc-Luc-GNN	HCV 5′UTR	Primer anti-sense	CGGGTTGATCCAAGAAAGGA	
HCVrep-Luc-GND		Probe	CGGGAGAGCCATAGTG-6-FAM-MGBNFQ	
HCV-UTRs-Luc				
HCVcc		Primer sense	GGCTCCATCTTAGCCCTAGTCA	
HCVcc-Luc-GNN	HCV 3′UTR	Primer anti-sense	CAGTATCAGCACTCTCTGCAG	
HCVrep-Luc-GND		Probe	CTAGCTGTGAAAGGTC-6-FAM-MGBNFQ	
HCV-UTRs-Luc				
HCV-5′UTR-Luc-Poly(A)		Primer sense	CCGTTGTTGTTTTGGAGC	
CAP-Luc-HCV-3′UTR	Firefly Luciferase ORF	Primer anti-sense	CGATGACGCCGGTGAACT	
CAP-Luc-Poly(A)		Probe	CTGGCGACGTAATCCACGAT-6-FAM-MGBNFQ	
		Primer sense	GCTGTTATTTTTTTACATGGTAACGC	12
EMCV-IRES-Luc-Poly(A)	Renilla Luciferase ORF	Primer anti-sense	CGCGCTACTGGCTCAATATG	12
		Probe	CCTCTTCTTATTTATGGCGACAT-6-FAM-MGBNFQ	12

LSm1-7 complexes bind to specific sites in viral RNA genomes and regulate their translation and replication

Rui Pedro Galão[1], Ashwin Chari[2], Isabel Alves-Rodrigues[1], Daniela Lobão[1], Antonio Mas[3], Christian Kambach[4], Utz Fischer[2], Juana Díez[1]

[1]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain. [2]Department of Biochemistry, Biocenter, University of Wuerzburg, Am Hubland, D-97074 Wuerzburg, Germany. [3]Centro Regional de Investigaciones Biomédicas – Facultad de Medicina, Universidad de Castilla La Mancha, 02006 Albacete, Spain. [4]Paul Scherrer Institut, Biomolecular Research, CH5232 Villigen, Switzerland.

Correspondence should be adressed to:
Juana Díez
Universitat Pompeu Fabra - Department of Experimental and Health Sciences
Dr Aiguader 88, 08003 Barcelona, Spain
Tel: 00-34-933. 160862
Fax: 00-34-933. 160901
e-mail address: juana.diez@upf.edu

Running head: LSm1-7 binds to viral RNAs regulating their fates

Keywords: BMV/ LSm1-7 / RNA virus / Sm proteins / translation / replication

ABSTRACT

LSm1-7 complexes promote cellular mRNA degradation, in addition to translation and replication of positive-strand RNA viruses such as the Brome mosaic virus (BMV). Yet, how LSm1-7 complexes act on their targets remains elusive. Here we report that reconstituted recombinant LSm1-7 complexes directly bind to two distinct RNA-target sequences in the BMV genome, a tRNA-like structure at the 3´-untranslated region and two internal **A-rich single-stranded regions**. Importantly, *in vivo* analysis shows that these sequences regulate the translation and replication of the BMV genome. Furthermore, both RNA target-sequences resemble those found for Hfq, the LSm counterpart in bacteria, suggesting a conservation through evolution. Our results provide the first evidence that LSm1-7 complexes interact directly with viral RNA genomes and open new perspectives in the understanding of LSm1-7 functions.

INTRODUCTION

The Sm and Sm-like (LSm) proteins constitute a conserved family whose members function in multiple aspects of RNA metabolism (reviewed in (Beggs 2005; Khusial et al. 2005; Wilusz and Wilusz 2005)). They are characterized by the presence of a so-called Sm fold that enables their association into homomeric (in eubacteria) or heteromeric (in eukaryotes) ring-shaped complexes. Sm proteins form a stable complex on U-rich small nuclear RNA (U snRNA) and, as part of the spliceosome, function in pre-mRNA processing. LSm complexes, in contrast, bind transiently to a broad spectrum of RNAs and influence their fate. Eight of the LSm proteins (LSm1 to LSm8) are conserved from yeast to humans and form two distinct heteroheptameric rings that differ only by the exchange of the LSm1 and LSm8 subunits. The LSm2-8 complex is localized in the nucleus and functions in pre-mRNA maturation (as part of the U6 and U6ATAC snRNPs), pre-mRNA decay, and processing of pre-tRNAs, pre-small nucleolar RNA (snoRNA) and pre-rRNAs. Conversely, the LSm1-7 complex is localized in the cytoplasm and plays a key role in mRNA decay (reviewed in (Beggs 2005; Khusial et al. 2005; Wilusz and Wilusz 2005)).

In eukaryotes, there are two major pathways of mRNA decay that are both initiated by the shortening of the poly(A) tail (Garneau et al. 2007). In the 3´-to-5´ decay pathway, deadenylated mRNAs are degraded by a large complex of 3´-5´exonucleases, the exosome. Conversely, in the 5´-to-3´ decay pathway, deadenylation triggers decapping and subsequent 5´-3´ exonucleolytic decay by Xrn1. In *Saccharomyces cerevisiae* it has been shown that LSm1-7 complexes act as activators of decapping in the 5´-3´decay pathway (Bouveret et al. 2000; Tharun et al. 2000) and as inhibitors of the 3´-5´ exonucleolytic decay (He and Parker 2001; Tharun et al. 2005). In mammalian cells, LSm1-7 complexes are also required for decapping of mRNAs. In

particular they have been shown to participate in the decay of mRNAs containing AU-rich elements (ARE) and histone mRNAs, whose 3´untranslated regions (UTRs) end in a conserved stem-loop instead instead of a poly(A) tail (Mukherjee et al. 2002; Mullen and Marzluff 2008; Stoecklin et al. 2006). The mechanism by which the LSm1-7 complexes promote mRNA decapping and their mode of interaction with the target RNA remain elusive. However, several *in vivo* and *in vitro* binding studies in yeast suggest a direct interaction with deadenylated 3´ UTRs (Chowdhury et al. 2007; He and Parker 2001; Tharun et al. 2005; Tharun and Parker 2001).

Viruses are obligatory intracellular parasites that depend on the host machinery to multiply. As such, they are useful tools to provide functional insights into cellular regulatory pathways. We have previously observed that the LSm1-7 complex plays a fundamental role in the replication of positive-strand RNA ((+)RNA) viruses (Diez et al. 2000; Mas et al. 2006; Noueiry et al. 2003). This viral group includes serious plant, animal and human pathogens, such as the hepatitis C virus and the SARS-coronavirus. Their genomes are single-stranded RNA molecules that are replicated in the cytoplasm of the host. Early in infection, (+)RNA viral genomes perform two essential functions. They act as mRNAs for expression of viral replicases and as templates for replication. As these two functions are mutually exclusive, a key step in the replication of all (+)RNA viruses is the regulated exit of their genomic RNAs from the cellular translation machinery to a membrane-associated viral replication complex, a process referred to as recruitment. **The molecular features underlying such transition are poorly understood.**

The propagation of the plant Brome mosaic virus in the yeast *Saccharomyces cerevisiae* has proven to be an excellent model system for studying common and

fundamental steps of (+)RNA virus biology in a relatively simple genetic background (Alves-Rodrigues et al. 2006). The genome of BMV consists of three RNAs with 5´-terminal m^7G-caps and a tRNA-like structure (TLS) located at the end of the 3´UTRs (Ahlquist 1992) (Fig. 1A). The *cis*-acting signals in BMV RNA translation and replication have been extensively characterized (reviewed in Noueiry and Ahlquist 2003). In brief, both the 5´ and 3´ UTRs of all BMV genomic RNAs contain partly overlapping sequences that control translation and initiation of negative- (in the 3´UTR) or positive-strand RNA synthesis (in the 5´UTR). Moreover, an element, the recruitment element (RE), at the 5´terminal ends of RNAs 1 and 2 and in the intergenic region of RNA3 is necessary and sufficient for proper selection and recruitment of the viral RNAs. RNA1 and RNA2 encode the replicase 1a, and the polymerase 2a, respectively. 1a is the only BMV protein required for the recruitment of the BMV genome (Fig. 1B) (Janda and Ahlquist 1998). RNA3 encodes the movement protein 3a and, through a subgenomic RNA generated during replication, the coat protein. Both proteins are required for systemic infection of plants but not for viral replication.

With the BMV/yeast system we previously reported that LSm1-7 complexes were required for both the translation and the recruitment of the BMV genome (Diez et al. 2000; Mas et al. 2006; Noueiry et al. 2003) (Fig. 1). Yet how these complexes act mechanistically on the BMV genome is unknown. With the availability of recombinant LSm1-7 complexes, we were now able to address this issue experimentally. Using an *in vitro* binding approach we provide evidence that LSm1-7 directly interacts with BMV RNAs via two specific RNA-targeting elements, namely a tRNA-like structure located in the 3´untranslated region (UTR) and two internal **A-rich single-stranded regions**. Functional *in vivo* analysis showed **that these LSm1-7**

RNA target sequences act as regulators of translation and recruitment of the BMV genome suggesting that the intrinsic RNA-binding characteristics of the LSm1-7 complexes determine their function. Moreover, the newly defined LSm1-7/BMV RNA interactions resemble those found for the Hfq complex, the bacterial counterpart of the LSm proteins. Hfq was first identified as a factor required for the replication of the (+)RNA bacteriophage Qβ. Together our results have important implications for the understanding of the function of LSm1-7 complexes and their conservation through evolution.

RESULTS

The dependence of RNA3 translation on LSm1-7 complexes is functionally linked to the 3´UTR of the BMV genome.

From the three BMV genomic RNAs, it is the RNA3 whose regulation of the translation /recruitment steps is characterized best. With the BMV/yeast system we have previously shown that the substitution of the 3´UTR of BMV RNA3 by a poly (A) tail suppressed the requirement of LSm1-7 complexes for RNA3 recruitment. This indicates a functional link between the presence of the natural 3´UTR and the LSm1-7 complex (Diez et al. 2000; Mas et al. 2006).To examine if the same 3'UTR region was also linked to the function of LSm1-7 complexes on RNA3 translation, we transformed WT yeast and the deletion strain *lsm1Δ* with plasmids expressing the BMV RNA3 or a RNA3 derivative in which the 3´UTR was substituted by the yeast *ADH1* polyadenylation signal (Fig. 2A). **Since neither the replicase 1a nor the polymerase 2a were expressed these RNAs will be translated but not recruited and replicated.** We focused on *lsm1Δ* because LSm1 is the subunit that defines the function of the LSm1-7 complex. After induction of expression, total protein and RNA were extracted and analyzed. RNA3 translation efficiency was obtained by normalizing the 3a protein levels detected via Western blot by the respective RNA3 levels detected via Northern blot (see histogram in Fig. 2B). As expected, RNA3 translation was strongly inhibited in the *lsm1Δ* deletion strain relative to WT (Noueiry et al. 2003) (Fig. 2B, lanes 1 and 3). However, upon replacement of the 3' UTR by a poly(A) tail similar RNA3 translation levels were observed in WT and *lsm1Δ* cells (Fig. 2B**,** lanes 2 and 4). In WT yeast, RNA3 was more efficiently translated than its RNA3p(A) derivative (Fig. 2B**,** lanes 1 and 2) highlighting the relevance of the natural BMV 3´UTR in translation. Thus, as previously observed for RNA3

recruitment, the 3´UTR of BMV RNA3 confers dependence on LSm1-7 for RNA3 translation.

Reconstituted LSm1-7 complexes bind specifically to the 3'UTR of BMV RNAs

LSm1-7 complexes are believed to influence the metabolism of cellular mRNAs by interacting directly with their 3´UTR (Chowdhury et al. 2007). Hence, we considered the possibility that LSm1-7 complexes regulate translation and replication of BMV through a direct interaction with the viral genome. To test this possibility, we carried out *in vitro* RNA-binding experiments with reconstituted recombinant human LSm1-7 complexes (Zaric et al. 2005). We focused our binding studies on the 3´ and 5´UTRs of the three genomic BMV RNAs because they contain conserved and structured sequences that are involved in the regulation of BMV RNA translation and replication. The reconstituted LSm1-7 complexes were incubated with the respective radiolabeled RNA molecules and binding was analyzed by gel-shift assays in native polyacrylamide gels. First, we determined the specificity of the LSm complexes used in this study (Fig. 3A). The RNA-binding specificity of LSm1-7 complexes is not known whereas binding of LSm2-8 complexes to U6 snRNA has been documented. Thus, we used the binding pattern of LSm2-8 as a control to evaluate the ability of the reconstituted complexes to discriminate between cognate and non-cognate RNAs. Consistently, we observed specific binding of the LSm2-8 to U6 but not to U1 snRNA, while the LSm1-7 complex did not bind to any of the two (Fig. 3A). We next tested the binding of the LSm1-7 complex to radiolabeled BMV RNA derivatives using identical conditions as determined for the LSm2-8 complex (Fig. 3B-D). Interestingly, a robust binding to the 3´UTR of all three genomic RNAs was observed. The 5´UTR, in contrast, failed to bind under the same conditions irrespective of whether they were m^7G-capped or uncapped (Fig. 3C and data not shown). Thus,

these experiments reveal a direct binding of the LSm1-7 complexes to the 3´UTR of the three BMV RNAs, i.e. those sequences that our *in vivo* data link to the dependence on LSm1-7 for BMV RNA translation and recruitment.

LSm1-7 complexes bind to the tRNA-like structure within the 3´UTR of BMV RNAs

All 3'UTRs of the three BMV RNAs contain two sequence elements, one with a non-defined structure named non-tRNA-like (NTLS) and the other with a highly conserved tRNA-like structure (TLS). The latter is 98% identical among the three RNAs and fulfills crucial functions in the BMV life cycle. It is required for efficient translation, initiation of replication and encapsidation of the BMV RNA genome (Barends et al. 2004; Choi and Rao 2003; Noueiry and Ahlquist 2003). Gel-shift assays were conducted to test which of the sequence elements interacts with the LSm1-7 complex. As shown in Fig. 3D (lanes 4, 8 and 12), interaction of the LSm1-7 complex with the TLS of all three RNAs was observed. This interaction was highly specific as no binding was observed to a structurally related plant tRNA (tRNA-tyr (Zerfass and Beier 1992)). In contrast, the NTLS regions either failed to interact completely (RNA1 and RNA2) or bound LSm1-7 only very weakly (RNA3) (Fig. 3D, lanes 2, 6 and 10).

Binding of the LSm1-7 complex to the TLS sequence was weaker as compared to the complete 3´UTR suggesting that additional sequence elements contribute to this interaction. To determine the minimal 3'UTR sequence required to obtain a complete band shift, we designed a set of RNA3 constructs starting either from the NTLS sequence and inserting consecutive domains from the TLS region or from the TLS sequence and inserting equivalent segments of the NTLS region (Fig. 4). In both cases

the addition of consecutive sequence elements gradually increased LSm1-7 binding capability. However, a complete band shift was only observed with the entire 3'UTR (Fig. 4B, C). Thus, even though the TLS alone is sufficient to bind LSm1-7 complex, the flanking sequences strongly increase its affinity. **The need of this 5´ flanking sequence might be sequence specific or alternatively length related.** As LSm1-7 and TLS are both required for BMV RNA translation and replication, our data suggest that the LSm1-7 effect on the BMV genome is linked to its ability to bind to the TLS region.

LSm1-7 complexes bind to the Intergenic Region of BMV RNA3

The BMV RNA3 contains an Intergenic Region between its ORFs that include the recruitment element (RE) and probably other unidentified regulatory sequences. The RE is required for 1a-mediated recruitment of the viral RNA to replication complexes (Sullivan and Ahlquist 1999). Chemical mapping together with computer predictions and phylogenetic studies have shown that the Intergenic Region folds into a long stem loop structure (Baumstark and Ahlquist 2001) (Fig. 5A). Interestingly, the LSm1-7 complex also interacts with this important regulatory region, as determined by gel-shift assays (Fig. 5B, C**,** lane 2). Two sequences, termed loops L1 and L2, were identified to be essential for LSm1-7 binding. L1 contains an A-rich region with **a 16-nt oligo(A) stretch followed by a UUAUUA sequence, while L2 contains a shorter A-rich region (Fig. 5A)**. Both of them are only present in RNA3 and no regulatory function has been ascribed to them yet. *In silico* RNA folding analysis suggested that removal of the loops had no effect on the overall structure of the Intergenic Region. LSm1-7 binding was not affected by removal of either L1 or L2, however, the simultaneous deletion of both entire loops totally abolished binding (Fig. 5C, compare lanes 4, 6 and 8). Deletion of the oligo(A) stretch in L1 plus the AU stretch in L2

strongly inhibited binding indicating the importance of these sequences for the interaction with LSm1-7 (Fig. 5C, lane 12). **The two bands observed for the IR3ΔA RNA reflect two conformations under native conditions and only the slower migrating one is able to interact with LSm1-7**. The RE element did not interact with LSm1-7 complexes (Fig. 5C, lane 8). This is consistent with the finding that neither the 5´UTR of RNA1 nor that of RNA2, which both contain RE elements, interacted with LSm1-7 (Fig. 3C, lanes 2 and 6). Taken together, the results show that the LSm1-7 rings are able to bind **A-rich sequences** located internally in RNA3.

LSm1-7 complexes specifically bind to the 3´UTRs and the Intergenic Region through the same RNA binding site

To determine the specificity of the observed LSm1-7 binding interactions, we performed competition assays. The interaction of LSm1-7 with the 3´UTRs and the IR region was specific since addition of excess unlabeled 3´UTR and IR, respectively, resulted in binding competition while addition of excess non-binding unlabeled 5´UTR did not (Fig. 6).

To explore whether the two determined RNA-target sequences interact with the same or distinct RNA binding sites in the LSm1-7 complex, we perform cross-competition assays. Radiolabeled RNA substrates derived from the RNA3 3'UTR or IR were incubated with LSm1-7 rings in the presence of increasing amounts of unlabeled IR or 3´UTR, respectively (Fig. 6). In both cases the binding was efficiently competed and this competition was comparable to the one obtained with the self-unlabeled RNA. These results strongly suggest that both the 3´UTR and the IR interact with the same RNA binding site of the LSm1-7 complex.

Loops L1 and L2 in the Intergenic Region of RNA3 regulate translation and recruitment to replication *in vivo*

The putative function of the loops L1 and L2 in the Intergenic Region on the regulation of the BMV RNA3 translation and recruitment has not been explored. With this aim, we generated yeast plasmids that express BMV RNA3 derivatives in which L1, L2 or both were deleted (ΔL1, ΔL2 and ΔL1L2, respectively) (Fig. 7A). To study putative effects on translation, WT and Δ*lsm1* yeast strains were transformed with the corresponding plasmids and the levels of 3a protein and RNA3 derivatives were analyzed. If the defined *in vitro* binding of LSm1-7 to these loops has a functional meaning, one would predict that the lack of the LSm1-7 complex would have a comparable effect as the deletion of the loops. Indeed, similar to the effect observed for the Δ*lsm1* strain in RNA3 translation (Fig. 7B, compare lanes 1 and 5), the single deletion of L1, L2 or both simultaneously in RNA3 resulted in a strong inhibition of RNA3 translation in WT yeast (Fig. 7B, lanes 2-4). In the Δ*lsm1* strain, deletion of the loops resulted in an even higher inhibition suggesting that additional regulatory factors might be binding to the loops (Fig. 7B, lanes 6-8). Therefore, the obtained results indicate that L1 and L2 loops are translation regulatory sites.

Next, we analyzed the effect of the deletion of the internal loops on RNA3 recruitment to the sites of replication. This 1a-dependent recruitment process protects the viral RNAs from degradation and thus dramatically increases their stability and subsequent accumulation (Janda and Ahlquist 1993; Mas et al. 2006; Schwartz et al. 2002; Sullivan and Ahlquist 1999). Accordingly, when RNA3 and 1a protein are co-expressed, differences in RNA3 accumulation reflects differences in recruitment. WT and *lsm1Δ* yeast strains were transformed with plasmids transcribing wt RNA3 or the corresponding deletion derivatives plus or minus another plasmid expressing the

BMV 1a protein. In the absence of 1a, a similar steady state level of RNA3 was obtained for all the constructs in WT and *lsm1Δ* yeast (data not shown). When 1a was co-expressed, the recruitment of RNA3 was inhibited in *lsm1Δ* when compared to WT yeast (Fig. 7C, upper panel and histogram, lanes 1 and 5, and Diez et al. (2000)). However, deletion of the loops resulted in a 2-3 fold increase of recruitment when compared to wt RNA3 in both WT and *lsm1Δ* yeast strains. (Fig. 7C, compare histograms corresponding to lane 1 with 2-4, and to lane 5 with 6-8). These effects were not related to differences in 1a expression levels (Fig. 7C, lower panel). Therefore, deletion of the loops L1 and L2 enhance recruitment independently of the presence of LSm1-7 complexes. Taken together, our experiments indicate that L1 and L2, the LSm1-7 recognition signals in the IR region, play an essential role in the regulation of BMV RNA3 translation as well as a secondary role in its recruitment.

DISCUSSION

The LSm1-7 complex, which functions in cellular mRNA decay, specifically binds to BMV genomes and instead of decay, promotes their translation and replication. By using *in vitro* binding assays with reconstituted LSm1-7 complexes and BMV RNA sequences, we have identified two specific LSm1-7 interaction sites in the BMV genome. Importantly, *in vivo* analysis showed that these sites are required for the regulation of translation and recruitment of BMV RNA. These results are consistent with a role of LSm1-7 complexes in the BMV life-cycle and suggest that the intrinsic RNA-binding characteristics of these complexes determine their function. Our data open new perspectives in the mode of action of these versatile complexes in viral and cellular RNA biology.

Reconstituted recombinant human LSm1-7 complexes specifically recognized two translation/replication regulatory signals in the BMV genome, the TLS region in the 3´UTR of all three genomic RNAs and two internal single-stranded **A-rich sequences (L1 and L2)** in the Intergenic Region of RNA3 that are located ~1 kb upstream of the 3´end. These are the first defined LSm1-7 binding sites in viral genomes and represent novel LSm1-7 interaction sites. Only two previous studies have addressed the binding specificity of LSm1-7 to RNA sequences other than homopolymers. Binding assays with purified yeast LSm1-7/Pat1 complexes have shown a preferential binding to oligoadenylated *versus* polyadenylated RNAs and a direct interaction near or at 3´ends that contain oligo(U) stretches (Chowdhury et al. 2007). However, since the protein Pat1 was present in the complex, its contribution to the identified RNA binding properties remains unclear. The other study was performed with reconstituted human LSm1-7 rings and showed their direct interaction and subsequent stabilization of reporter RNAs containing 5´poly(A) tracts. Although these studies were performed

with non-natural RNA sequences, **a function of LSm1-7 in the stabilization of poxviridae messenger RNA, which contain 5´poly(A) tracts, was suggested via a 5´-3´circularization by binding of the LSm1-7 complexes to the 3´ and 5´ends (Bergman et al. 2007).**

For BMV RNA3, substitution of the TLS region by a poly(A) tail suppressed the requirement of LSm1-7 complexes for RNA3 translation and recruitment. Since poly(A) tails in cellular mRNAs mediate 5´-3´circularization via the binding of the poly(A)binding protein, the LSm1-7/TLS binding may establish such 5´-3´interactions in BMV RNAs. Proteins containing Sm folds act as chaperones facilitating a variety of RNA-RNA and RNA protein interactions (Wilusz and Wilusz 2005). Accordingly, the binding of LSm1-7 to TLS and the L1 and L2 internal loops might facilitate rearrangements in the viral ribonucleoprotein (RNP) structure required for the 5´-3´interaction. This circularization would then stabilize the RNA permitting both efficient translation and 1a-dependent recruitment via recognition of the RE element (Fig. 8). The fact that the lack of LSm1-7 complexes inhibits both BMV RNA3 translation and recruitment suggests that binding of LSm1-7 is required for both processes. Since translation and recruitment are two antagonist processes, changes such as the presence of 1a protein expression or the absence of some of the LSm1-7 interaction sites in the viral genomes would result in favouring one over the other. In line with this, we observed that deletion of the L1 and L2 loops inhibit translation but favours recruitment. This may be an indirect effect due the absence of the competing translation function. Alternatively, deletion of L1 and L2 could affect the RNP reorganization in such a way that the RE signal would be more accessible to the

1a protein. On the other hand, the observed effects may be also due to the binding to the loops L1 and L2 of additional host factors not identified yet.

Deletion of both L1 and L2 loops in the intergenic region of the RNA3 was required to abolish LSm1-7 binding (Fig. 5), however deletion of either L1 or L2 inhibited translation and enhanced recruitment (Fig. 7). It is yet unclear which are the basis of this particular disconnection between the *in vitro* binding assays and the *in vivo* functional data. However, it might be possible that in the *in vivo* setting, where the local concentration of viral RNA and LSm protein is low, the presence of both L1 and L2 sites are important to promote efficient LSm1-7 binding. Hence, under *in vivo* conditions, a suboptimal binding site would not suffice to allow LSm binding and function. In the *in vitro* setting, however, we may simply override this subtle effect.

The newly defined RNA-target sites of LSm1-7 rings resemble those found for Hfq, the bacterial counterpart of Sm/LSm proteins (Lee and Feig 2008; Wilusz and Wilusz 2005). Hfq forms an homohexameric ring that functions in degradation but also in stabilization and translation of cellular RNAs by facilitating RNA-RNA and RNA-protein interactions (Wilusz and Wilusz 2005). Interestingly, Hfq was first identified as a factor required for the replication of the (+)RNA bacteriophage Qß (Franze de Fernandez et al. 1968). By promoting changes in the 3´secondary structure of the Qß RNA genome, it allows the association of the viral replicase with the viral genome and the subsequent initiation of replication (Schuppli et al. 1997). In contrast to the LSm1-7 complex, a crystal structure of the Hfq complex is available (Schumacher et al. 2002). Two faces of the ring have been shown to bind RNA substrates with different specificities. The distal face binds to poly(A) tails while the proximal face

binds to internal AU-rich regions located close to stem-loop structures in non-coding RNAs (ncRNAs) or to tRNAs (Lee and Feig 2008; Mikulecky et al. 2004). As for Hfq, LSm1-7 complexes bound to internal **regions, in this case A-rich stretches, that flank a stem-loop RNA structure a**nd to a tRNA-like sequence (TLS) in the BMV 3´UTR**.** In addition, our cross-competition binding analysis showed that both sequences bound through the same surface in the LSm1-7 complex (Fig. 6). Mapping of the minimal BMV 3´UTR sequence necessary for optimal binding suggests that, as described for the Hfq/tRNA interaction, it is not a sequence motif *per se* but a structural motif which promotes binding (Fig. 4). The TLS is tyrosylated in vivo by host enzymes (Kohl and Hall 1974), however the LSm1-7 complex did not bind the related tRNAtyr (Fig. 3). Thus, the observed LSm1-7/TLS interactions are not solely dependent on the structural mimicry of tRNA, but instead might involve additional residues or structures absent in canonical tRNAs. The fact that the described binding properties and functions of LSm1-7 complexes are similar to those observed for the bacterial Hfq, suggest a strong conservation through evolution.

Viruses require the cellular machinery to multiply. In order to express their genomes and regulate the viral life cycle, a complex interplay between the cellular translation and degradation machinery is starting to emerge (Sokoloski et al. 2006). Our data indicate that BMV genomes usurp the LSm1-7 complex, which function in cellular mRNA decay, to regulate their translation and replication. Given the common strategies of (+)RNA virus replication and that at least three viruses from this group, the bacterial phage Qß, the plant BMV, and the human HCV virus (**Scheller et al. 2009)** require this complex to regulate their replication, it is likely that LSm1-7 is broadly used within this viral group. The obtained results open new perspectives on the role and way of function of LSm1-7 complexes not only on viral genomes but

possibly also on specific cellular RNAs. For example, LSm1 protein has been found to be over-expressed in several human cancers (Fraser et al. 2005; Kelley et al. 2003; Streicher et al. 2007). Recent findings show that the long non-coding RNA (ncRNA) MALAT1, which is overexpressed in many human carcinomas, generates a small ncRNA that adopt a tRNA-like structure and localizes to the cytoplasm (Wilusz et al. 2008). The putative role of LSm1-7 complexes in the stability and function of such ncRNAs and other cellular RNAs with special features remains to be explored further and might uncover new functionalities of this versatile complex.

MATERIALS AND METHODS

Yeast cells

Standard yeast genetics and media were used (Guthrie and Fink 1991). For recruitment experiments**,** wild-type (WT) *Saccharomyces cerevisiae* strain YPH500 (*MATα ura3-52 lys2-801 ade2-101 trp1Δ63 his3-Δ200 leu2-Δ1*) and the derivative *lsm1Δ* deletion strain were used (Diez et al. 2000). For translation experiments, WT BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) and the derivative *lsm1Δ* deletion strain were used (Winzeler et al. 1999) and purchased from EUROSCARF, Frankfurt, Germany.

Plasmids

Plasmids for in vivo experiments. For recruitment assays, plasmid pJDSal1 was used to transcribe wt BMV RNA3 by using a *CUP1* promoter (Diez et al. 2000). Plasmids pIR1, pIR2 and pIR3 were derived from pJDSal1 by using the QuicKchange XL II site-directed mutagenesis kit (Stratagene, La Jolla, CA) to delete the loop L1 (nt 1194-1226), loop L2 (nt 1000-1020), or both loops simultaneously. An additional mutation A998T was inserted into pIR2 and pIR3 to introduce a stop codon at the end of the 3a ORF. The primers used are listed in Supplementary Table1. Plasmids pIR1, pIR2 and pIR3 were then used to transcribe the BMV RNA3 derivatives ΔL1, ΔL2 and ΔL1L2, respectively. To achieve similar 1a expression levels in WT and Δ*lsm1* yeast, the 1a replicase was expressed from either pB1CT19 or pB1YT3H under the *ADH1* or *GAL1* promoter, respectively, as described before (Mas et al. 2006). In these plasmids the 5´UTR and the 3´UTR of the BMV RNA1 were replaced by cellular leader sequences and the *ADH1*-polyadenylation signals (Ahola et al. 2000; Janda and Ahlquist 1993). All the generated plasmids used to test effects on 3a protein translation were derived from pB3RQ39-URA, a yeast plasmid transcribing wt BMV

RNA3 under the *GAL1* promoter (Beckham et al. 2007). pB3RQ39JDURA was constructed by cloning the *SalI*/*EcoRI* fragment from pB3RQ39JD (Diez et al. 2000) into pB3RQ39-URA and used to transcribe a BMV RNA3 derivative (RNA3p(A)) with the yeast *ADH1* polyadenylation site substituting the original viral 3'end. Plasmids pIR7, pIR8 and pIR9 transcribe from the *GAL1* promoter the BMV RNA3 derivatives ΔL1, ΔL2 and ΔL1L2, respectively. To construct these plasmids, pIR1, pIR2 and pIR3 were *ClaI*/*StuI* digested and the resulting ~1,2 Kb fragments cloned into pB3RQ39 (Ishikawa et al. 1997). Following, a digestion with *EcoRI*/*MscI* resulted in ~2KB fragments that were cloned into the same sites in pB3RQ39-URA.

Constructs used in gel-shift assays. Unless otherwise mentioned, constructs were generated by PCR amplification of the fragments of interest from BMV RNAs 1, 2 and 3 genomes cloned, respectively, in the plasmids pB1tp3, pB2tp5 and pB3tp8 (French and Ahlquist 1988; Janda et al. 1987). Forward primers carried a T7 promoter sequence at their 5'-end. Amplified fragments were cloned into *SmaI* digested pUC18 (New England Biolabs, Ipswich, MA) or pGEM-T vectors (Promega, Madison, WI). Fragments IR A and IR 3ΔA were synthetically produced by Geneart (Regensburg, Germany) and cloned into plasmid pPCR-Script-ampR (Stratagene) using *KpnI* and *SacI* restriction sites. Finally, constructs used to generate the transcritps IRΔL1, IRΔL2 and IRΔL1+2, were obtained by PCR amplification of the relevant parts of pIR1, pIR2 and pIR3, respectively, and cloned into pGEM-T vector (Promega, Madison, WI). Standard procedures were used for DNA manipulations and all generated constructs were verified by sequencing. Sequences of the primers used for each construct are given in Supplementary Table 2. BMV sequence numbering refers to the position of the corresponding nucleotides in the complete BMV RNA genomes (GenBank acession numbers NC_002026, NC_002027 and NC_002028). Plasmids used for expression and purification of individual human LSm proteins, and for

transcription of the *Xenopus* U6 and U1 snRNA and the tRNA-tyr from *Nicotania* were already described (Hamm et al. 1990; Zaric et al. 2005; Zerfass and Beier 1992).

Recruitment and translation assays.

To analyze recruitment, yeast cells were transformed with the corresponding plasmids and grown in 0.5 mM $CuSO_4$, 2% dextrose minimal media at 30°C until mid-log phase and then induced in 2% galactose for 72h. Total RNA from yeast cells was isolated by the hot-phenol method and analyzed by Northern blot as previously described (Janda and Ahlquist 1993). Specific probes to detect positive-strand BMV RNA3 and 18S rRNA has been described (Alves-Rodrigues et al. 2007; Ishikawa et al. 1997) and were generated using the MAXIscript In Vitro Transcription Kit (Ambion, Austin, TX). Northern blots were imaged on a Typhoon 8600 Instrument (Amersham Biosciences, Piscataway, NJ) and band intensities were quantified using the ImageQuant Software (Molecular Dynamics, Piscataway, NJ).

To evaluate BMV 3a protein translation, yeast cells were transformed with the corresponding plasmids and grown in 2% galactose at 30°C until mid-log phase. Total protein was extracted from yeast cells as previously described (Ishikawa et al. 1997), separated on 10% SDS-polyacrylamide gels and immunoblotted. Antibodies against BMV 1a (Restrepo-Hartwig and Ahlquist 1999), BMV 3a (Noueiry et al. 2003) and Phophoglycerate kinase (PGK) (Molecular Probes, Eugene, OR) were used for Western blot. Protein detection and quantification was made using the infrared imaging system Odyssey (LI-COR Biosciences, Lincoln, NE). Total RNA was extracted and RNA3 accumulation was measured as explained above.

LSm recombinant proteins and RNA gel shift assays

Expression and purification of individual LSm proteins and sub-complexes, as well as the reconstitution and purification of LSm1-7 and LSm2-8 rings, were performed essentially as described (Zaric et al. 2005), with the exception of LSm8, which was expressed singly and purified as described for LSm1. Transcripts used to perform gel shift assays were transcribed *in vitro* and labelled with [α-^{32}P]UTP. **In a typical binding experiment, 50fmol of labeled RNA is incubated with 20pmol of reconstituted LSm complexes** in a buffer containing 20mM HEPES-NaOH pH 7.5, 200mM NaCl, 2mM MgCl2, 0.1U/µl RNAsin and 0.1 µg/µl yeast tRNA in a 5 µl assay at 30°C for 1hr. Samples were loaded on previoulsy pre-run 5% native polyacrrylamide gels, and run at 4°C for 2h, 30mA. Gels were autoradiographed at -80°C on maximum sensivity Kodak Biomax Films (Sigma-Aldrich, Munich, Germany). For gel shift assays that included RNA competition, the indicated amount of RNA competitor was added to the reactions.

ACKNOWLEDGMENTS

We thank F. Gebauer for helpful discussions and critical reading of the manuscript. This work was supported by grants from the Spanish Ministerio de Educación y Ciencia BFU2007-66933/BMC. I. A-R and D.L. were supported by Fundação para a Ciência e Tecnología (SARH/BD/9630/2002; SFRH/BD/37047/2007), Portugal.

References

Ahlquist P. 1992. Bromovirus RNA replication and transcription. *Curr Opin Genet Dev* **2**: 71-76.

Ahola T, den Boon JA, Ahlquist P. 2000. Helicase and capping enzyme active site mutations in brome mosaic virus protein 1a cause defects in template recruitment, negative-strand RNA synthesis, and viral RNA capping. *J Virol* **74**: 8803-8811.

Alves-Rodrigues I, Galao RP, Meyerhans A, Diez J. 2006. Saccharomyces cerevisiae: a useful model host to study fundamental biology of viral replication. *Virus Res* **120**: 49-56.

Alves-Rodrigues I, Mas A, Diez J. 2007. Xenopus Xp54 and human RCK/p54 helicases functionally replace yeast Dhh1p in brome mosaic virus RNA replication. *J Virol* **81**: 4378-4380.

Barends S, Rudinger-Thirion J, Florentz C, Giege R, Pleij CW, Kraal B. 2004. tRNA-like structure regulates translation of Brome mosaic virus RNA. *J Virol*, **78**: 4003-4010.

Baumstark T, Ahlquist P. 2001. The brome mosaic virus RNA3 intergenic replication enhancer folds to mimic a tRNA TpsiC-stem loop and is modified in vivo. *Rna* **7**: 1652-1670.

Beckham CJ, Light HR, Nissan TA, Ahlquist P, Parker R, Noueiry A. 2007. Interactions between brome mosaic virus RNAs and cytoplasmic processing bodies. *J Virol* **81**: 9759-9768.

Beggs JD. 2005. Lsm proteins and RNA processing. *Biochem Soc Trans* **33**: 433-438.

Bergman N, Moraes KC, Anderson JR, Zaric B, Kambach C, Schneider RJ, Wilusz, CJ, Wilusz J. 2007. Lsm proteins bind and stabilize RNAs containing 5' poly(A) tracts. *Nat Struct Mol Biol* **14**: 824-831.

Bouveret E, Rigaut G, Shevchenko A, Wilm M, Seraphin B. 2000. A Sm-like protein complex that participates in mRNA degradation. *Embo J* **19**: 1661-1671.

Choi YG, Rao AL. 2003. Packaging of brome mosaic virus RNA3 is mediated through a bipartite signal. *J Virol* **77**: 9750-9757.

Chowdhury A, Mukhopadhyay J, Tharun S. 2007. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *Rna* **13**: 998-1016.

Diez J, Ishikawa M, Kaido M, Ahlquist P. 2000. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. *Proc Natl Acad Sci U S A* **97**: 3913-3918.

Franze de Fernandez MT, Eoyang L, August JT. 1968. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **219**: 588-590.

Fraser MM, Watson PM, Fraig MM, Kelley JR, Nelson PS, Boylan AM, Cole DJ, Watson DK. 2005. CaSm-mediated cellular transformation is associated with altered gene expression and messenger RNA stability. *Cancer Res* **65**: 6228-6236.

French R, Ahlquist P. 1988. Characterization and engineering of sequences controlling in vivo synthesis of brome mosaic virus subgenomic RNA. *J Virol* **62**: 2411-2420.

Garneau NL, Wilusz J, Wilusz CJ. 2007. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**: 113-126.

Guthrie C, Fink GR (eds). 1991. *Guide to yeast genetics and molecular biology*. Academic Press, Inc., San Diego, CA.

Hamm J, Darzynkiewicz E, Tahara SM, Mattaj IW. 1990. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* **62**: 569-577.

He W, Parker R. 2001. The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3' termini from partial degradation. *Genetics* **158**: 1445-1455.

Ishikawa M, Janda M, Krol MA, Ahlquist P. 1997. In vivo DNA expression of functional brome mosaic virus RNA replicons in Saccharomyces cerevisiae. *J Virol* **71**: 7781-7790.

Janda M, Ahlquist P. 1993. RNA-dependent replication, transcription, and persistence of brome mosaic virus RNA replicons in S. cerevisiae. *Cell* **72**: 961-970.

Janda M, Ahlquist P. 1998. Brome mosaic virus RNA replication protein 1a dramatically increases in vivo stability but not translation of viral genomic RNA3. *Proc Natl Acad Sci U S A* **95**: 2227-2232.

Janda M, French R, Ahlquist P. 1987. High Efficiency T7 Polymerase Synthesis of Infectious RNA from Cloned Brome Mosaic Virus cDNA and Effects of 5' Extensions on Transcript Infectivity. *Virology* **158**: 259-262.

Kelley JR, Fraser MM, Hubbard JM, Watson DK, Cole DJ. 2003. CaSm antisense gene therapy: a novel approach for the treatment of pancreatic cancer. *Anticancer Res* **23**: 2007-2013.

Khusial P, Plaag R, Zieve GW. 2005. LSm proteins form heptameric rings that bind to RNA via repeating motifs. *Trends Biochem Sci* **30**: 522-528.

Kohl RJ, Hall TC. 1974. Aminoacylation of RNA from several viruses: amino acid specificity and differential activity of plant, yeast and bacterial synthetases. *J Gen Virol* **25**: 257-261.

Lee T, Feig AL. 2008, The RNA binding protein Hfq interacts specifically with tRNAs. *Rna* **14**: 514-523.

Mas A, Alves-Rodrigues I, Noueiry A, Ahlquist P, Diez J. 2006. Host deadenylation-dependent mRNA decapping factors are required for a key step in brome mosaic virus RNA replication. *J Virol* **80**: 246-251.

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL. 2004. Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* **11**: 1206-1214.

Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn G, Lutz CS, Wilusz, J. 2002. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *Embo J* **21**: 165-174.

Mullen TE, Marzluff WF. 2008. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev* **22**: 50-65.

Noueiry AO, Ahlquist P. 2003. Brome mosaic virus RNA replication: revealing the role of the host in RNA virus replication. *Annu Rev Phytopathol* **41**: 77-98.

Noueiry AO, Diez J, Falk SP, Chen J, Ahlquist P. 2003. Yeast Lsm1p-7p/Pat1p deadenylation-dependent mRNA-decapping factors are required for brome mosaic virus genomic RNA translation. *Mol Cell Biol* **23**: 4094-4106.

Restrepo-Hartwig M, Ahlquist P. 1999. Brome mosaic virus RNA replication proteins 1a and 2a colocalize and 1a independently localizes on the yeast endoplasmic reticulum. *J Virol* **73**: 10303-10309.

Scheller, N., Mina, L.B., Galão, R.P., Chari, A., Gímenez-Barcons, M., Noueiry, A., Fischer, U., Meyerhans, A., Díez, J. Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates. PNAS doi:10.1073/pnas.0906413106, (in press).

Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG. 2002. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *Embo J* **21**: 3546-3556.

Schuppli D, Miranda G, Tsui HC, Winkler ME, Sogo JM, Weber H. 1997. Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less Escherichia coli. *Proc Natl Acad Sci U S A* **94**: 10239-10242.

Schwartz M, Chen J, Janda M, Sullivan M, den Boon J, Ahlquist P. 2002. A positive-strand RNA virus replication complex parallels form and function of retrovirus capsids. *Mol Cell* **9**: 505-514.

Sokoloski KJ, Wilusz CJ, Wilusz J. 2006. Viruses: overturning RNA turnover. *RNA Biol* **3**: 140-144.

Stoecklin G, Mayo T, Anderson P. 2006. ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep* **7**: 72-77.

Streicher KL, Yang ZQ, Draghici S, Ethier SP. 2007. Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon. *Oncogene* **26**: 2104-2114.

Sullivan ML, Ahlquist P. 1999. A brome mosaic virus intergenic RNA3 replication signal functions with viral replication protein 1a to dramatically stabilize RNA in vivo. *J Virol* **73**: 2622-2632.

Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R. 2000. Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* **404**: 515-518.

Tharun S, Muhlrad D, Chowdhury A, Parker R. 2005. Mutations in the Saccharomyces cerevisiae LSM1 gene that affect mRNA decapping and 3' end protection. *Genetics* **170**: 33-46.

Tharun S, Parker R. 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* **8**: 1075-1083.

Wilusz CJ, Wilusz J. 2005. Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* **12**: 1031-1036.

Wilusz JE, Freier SM, Spector DL. 2008. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**: 919-932.

Winzeler EA, et al. 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**: 901-906.

Zaric B, Chami M, Remigy H, Engel A, Ballmer-Hofer K, Winkler FK, Kambach, C. 2005. Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J Biol Chem* **280**: 16066-16075.

Zerfass K, Beier H. 1992. Pseudouridine in the anticodon G psi A of plant cytoplasmic tRNA(Tyr) is required for UAG and UAA suppression in the TMV-specific context. *Nucleic Acids Res* **20**: 5911-5918.

FIGURE LEGENDS

FIGURE 1. The BMV genome and the initial BMV replication steps. **(a)** Schematic diagram of the BMV tripartite genome showing the ORFs (solid black boxes) and the untranslated regions (UTRs) (single lines). The three BMV RNAs are capped (m^7G) and end in a tRNA-like structure (cloverleaf structure). The location of the recruitment element (RE) is shown for the three RNAs. **(b)** After translation of the viral proteins, the 1a replicase recognizes the RE sequence to specifically recruit the viral genomes from the cellular translation machinery to viral-induced invaginations in the endoplasmic reticulum where replication occurs. The LSm1-7 complex plays a key role in the regulation of these processes.

FIGURE 2. A 3'poly(A) tail suppresses the requirement of LSm1-7 complexes for RNA3 translation. **(a)** Schematic representation of the RNA3 and the RNA3-poly(A) derivative (RNA3p(A)) in which the 3´UTR was replaced by the yeast *ADH1* polyadenylation site. **(b)** WT and *lsm1Δ* yeast strains were transformed with a plasmid harboring RNA3 or RNA3p(A). Upper panels, Western blot analysis of the 3a protein expression. As a control for equal loading of total protein, expression of the phosphoglycerate kinase protein (PGK) was also analyzed. Lower panels, Northern blot analysis of RNA3 accumulation. Detection of the 18S ribosomal RNA was used to assure equal loading and sample quality. Histograms show average and standard error of the mean of the accumulation of 3a protein relative to the amount of RNA3 for at least three independent colonies. The average value obtained for RNA3 in WT cells was set to 100.

FIGURE 3. Binding of LSm1-7 complexes to the 3´UTRs of the three BMV RNAs depends on a tRNA-like structure. **(a)** Left, comassie blue-stained SDS-polyacrylamide gel of purified reconstituted LSm1-7 and LSm2-8 complexes. Right, binding specificity of LSm1-7 and LSm2-8 complexes were determined by gel-shift analysis using recombinant rings and radiolabeled gel-purified U1 and U6 snRNAs. After complex formation, samples were loaded on a nondenaturating polyacrylamide gel and visualized by autoradiography. **(b)** Schematic representation of generated transcripts. The numeration refers to the position of the corresponding nucleotides in the complete BMV RNA genomes. **(c, d)** Gel-shift assays were performed as described above. LSm1-7 complexes were incubated with radiolabeled transcribed RNAs corresponding to the 5' and 3'UTRs **(c)** or with the 3´UTR TLS and NTLS sequences of the three genomic BMV RNAs **(d)**. The plant tRNA corresponds to Tyr-tRNA of *Nicotania bentamiana*. Asterisks indicate the position of the gel-shifted RNAs.

FIGURE 4. The complete 3'UTR is required for optimal interaction with LSm1-7 complexes. **(a)** Secondary structure of the TLS from the 3'UTR of BMV RNA3. Numbers refer to corresponding nucleotide positions in the BMV RNA3 genome. The various stem-loop structures are indicated (A to E). To facilitate their identification some of them are highlighted with different shadowing (adapted from (Barends et al., 2004)). **(b)** RNA substrates corresponding to the NTLS region of the RNA3 3´UTR and to the same sequence plus consecutive TLS domains or **(c)** RNA substrates corresponding to the TLS region of the RNA3 3´UTR and to the same sequence plus additional stretches from the NTLS region were incubated with purified LSm1-7 complexes and analyzed by gel-shift assays as described in Fig 3.

FIGURE 5. Binding of LSm1-7 complexes to the Intergenic Region of RNA3 depends on the **A-rich** loops L1 and L2**. (a)** Secondary structure of the Intergenic Region of BMV RNA3. Numbers refer to the corresponding nucleotide positions in the BMV RNA3 genome. The recruitment element and the loops L1 and L2 are shown. In bold are indicated the stop codon of the 3a ORF and the start codon of the coat protein that limit the Intergenic Region. **(b)** Schematic diagram of the generated RNA transcripts used to map the sequences in the IR that mediate the interaction with the LSm1-7 complexes. **(c)** The corresponding radiolabeled RNAs were used to perform gel-shift assays with LSm1-7 complexes as described in Figure 3.

FIGURE 6. LSm1-7 complexes bind to the 3'UTR and the Intergenic region (IR) specifically and through a common binding site. Radiolabeled RNA3 3'UTR (upper panels) and RNA3 IR (lower panels) were incubated with reconstituted LSm1-7 complexes in the presence of increasing amounts of unlabeled RNA1 5'UTR, RNA3 3'UTR or RNA3 IR transcripts as competitors.

FIGURE 7. Deletion of loops L1 and L2 in the Intergenic Region inhibits translation and favors recruitment of BMV RNA3. **(a)** Generated BMV RNA3 derivatives in which the internal loop1 (ΔL1), loop2 (ΔL2) or both (ΔL1L2) were deleted. **(b)** WT and *lsm1Δ* strains were transformed with a plasmid harboring the RNA3 or the RNA3 derivatives (ΔL1, ΔL2 or ΔL1L2) and the accumulation of the 3a protein and RNA3 were analyzed by Western and Northern blotting, respectively, as in Fig 2. Histograms show average and standard error of the mean of the accumulation of 3a protein relative to the amount of RNA3 from at least three independent colonies. The average 3a/RNA3 value in WT yeast was set to 100. **(c)** WT and *lsm1Δ* yeast strains were transformed with a plasmid expressing the 1a protein plus a plasmid transcribing the

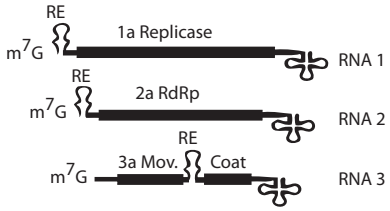
wt RNA3 or the RNA3 derivatives (ΔL1, ΔL2 or ΔL1L2). RNA3 accumulation was analyzed by Northern blotting and 1a protein expression by Western blotting. Histograms represent relative accumulation of RNA3 and the RNA 3 derivatives in wt and *lsm1Δ* cells of three independent colonies. The average accumulation of RNA3 in WT yeast was set to 100.

FIGURE 8.

Model of LSm1-7 function on BMV RNA3. Binding of LSm1-7 complexes to the tRNA-like structure in the 3´UTR and the internal loops in the intergenic region would regulate translation and recruitment to replication of BMV RNA3. Given that the LSm1-7 ring does not bind simultaneously to both RNA sequences, the interaction with two rings might be considered.

Figure 1

A


RE
1a Replicase
m7G
RNA 1
RE
2a RdRp
m7G
RNA 2
RE
3a Mov.
Coat
m7G
RNA 3

B

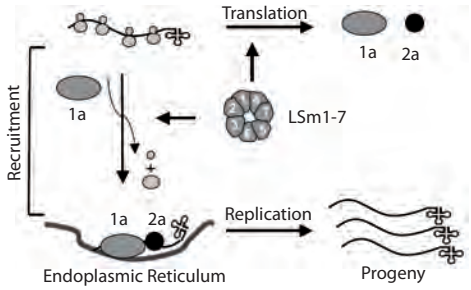

Translation
1a
2a
Recruitment
1a
LSm1-7
1a
2a
Replication
Endoplasmic Reticulum
Progeny

Figure 2

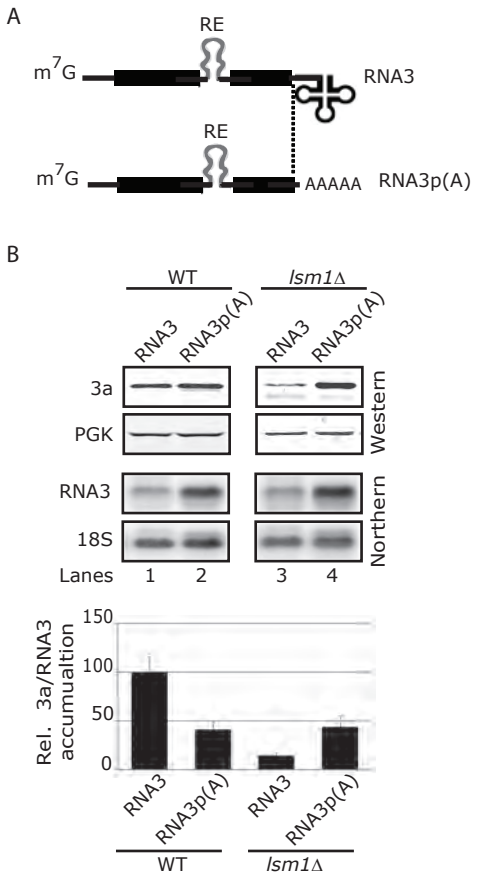

A
RE
m7G
RNA3
RE
m7G
AAAAA
RNA3p(A)
B
WT
lsm1Δ
RNA3
RNA3p(A)
RNA3
RNA3p(A)
3a
PGK
Western
RNA3
18S
Northern
Lanes 1 2 3 4
Rel. 3a/RNA3 accumualtion
150
100
50
0
RNA3
RNA3p(A)
RNA3
RNA3p(A)
WT
lsm1Δ

Figure 3

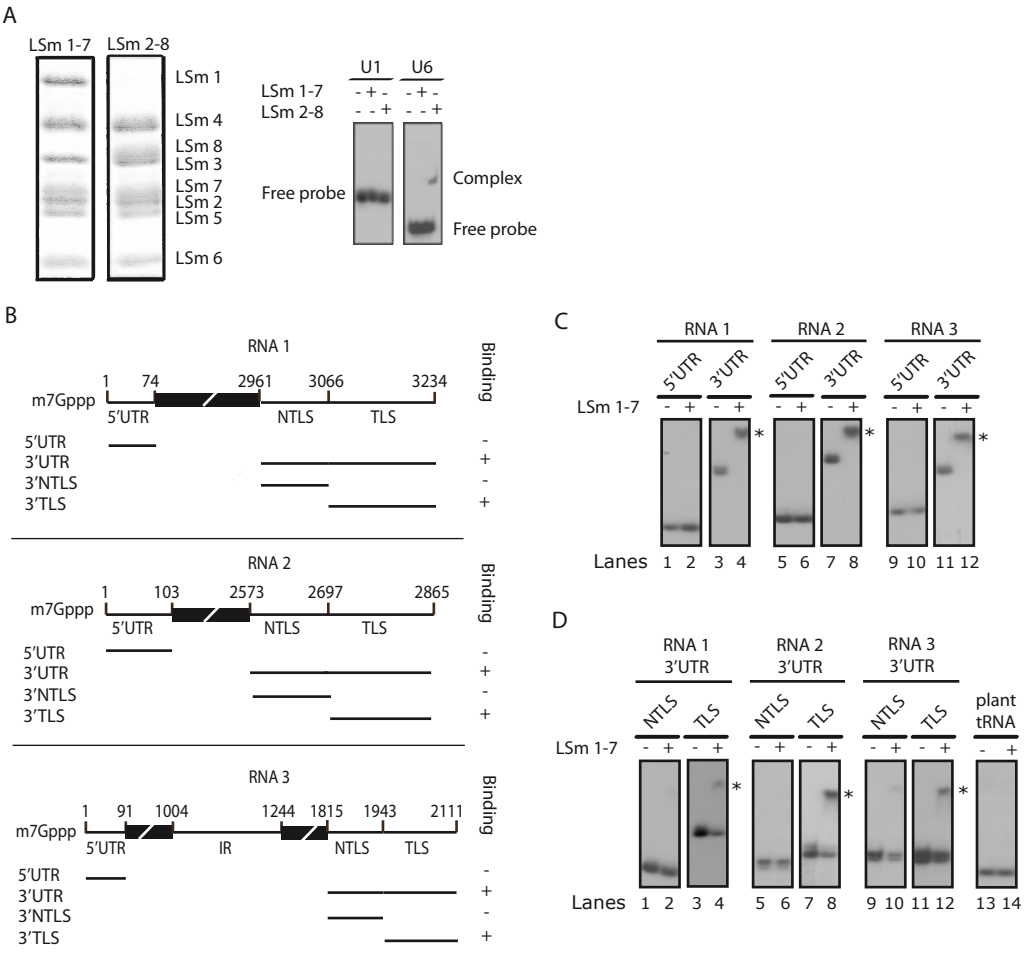
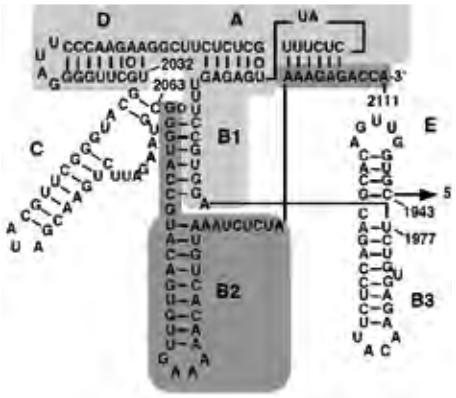
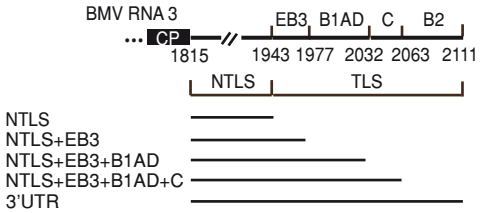

A
LSm 1-7
LSm 2-8
LSm 1
LSm 4
LSm 8
LSm 3
LSm 7
LSm 2
LSm 5
LSm 6
U1
U6
LSm 1-7
LSm 2-8
Free probe
Complex
Free probe
B
RNA 1
m7Gppp
1
74
2961
3066
3234
5'UTR
NTLS
TLS
Binding
5'UTR
3'UTR
3'NTLS
3'TLS
RNA 2
m7Gppp
1
103
2573
2697
2865
5'UTR
NTLS
TLS
RNA 3
m7Gppp
1
91
1004
1244
1815
1943
2111
5'UTR
IR
NTLS
TLS
C
RNA 1
RNA 2
RNA 3
5'UTR
3'UTR
LSm 1-7
Lanes 1 2 3 4 5 6 7 8 9 10 11 12
D
RNA 1 3'UTR
RNA 2 3'UTR
RNA 3 3'UTR
NTLS
TLS
plant tRNA
LSm 1-7
Lanes 1 2 3 4 5 6 7 8 9 10 11 12 13 14

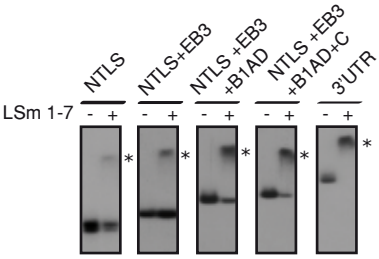
Figure 4

A

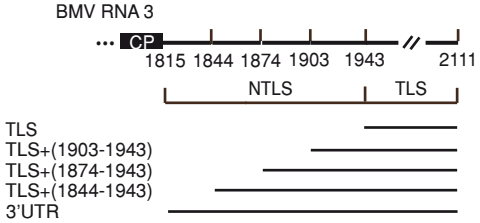

D
A
UA
CCCAAGAAGGCUUCUCUCG
UUUCUC
GGGUUCGU
2032
GAGAGU
AAAGAGACCA-3'
2063
2111
B1
C
E
5
1943
1977
B2
B3

B


BMV RNA 3
CP
EB3
B1AD
C
B2
1815
1943 1977 2032 2063 2111
NTLS
TLS
NTLS
NTLS+EB3
NTLS+EB3+B1AD
NTLS+EB3+B1AD+C
3'UTR


NTLS
NTLS+EB3
NTLS +EB3 +B1AD
NTLS +EB3 +B1AD+C
3'UTR
LSm 1-7
- +
*

C


BMV RNA 3
CP
1815 1844 1874 1903 1943 2111
NTLS
TLS
TLS
TLS+(1903-1943)
TLS+(1874-1943)
TLS+(1844-1943)
3'UTR

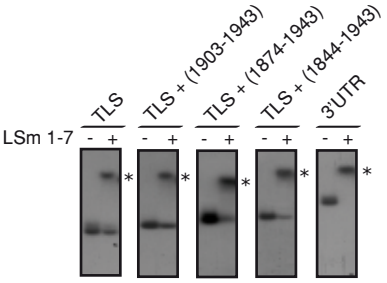

TLS
TLS + (1903-1943)
TLS + (1874-1943)
TLS + (1844-1943)
3'UTR
LSm 1-7
- +
*

Figure 5

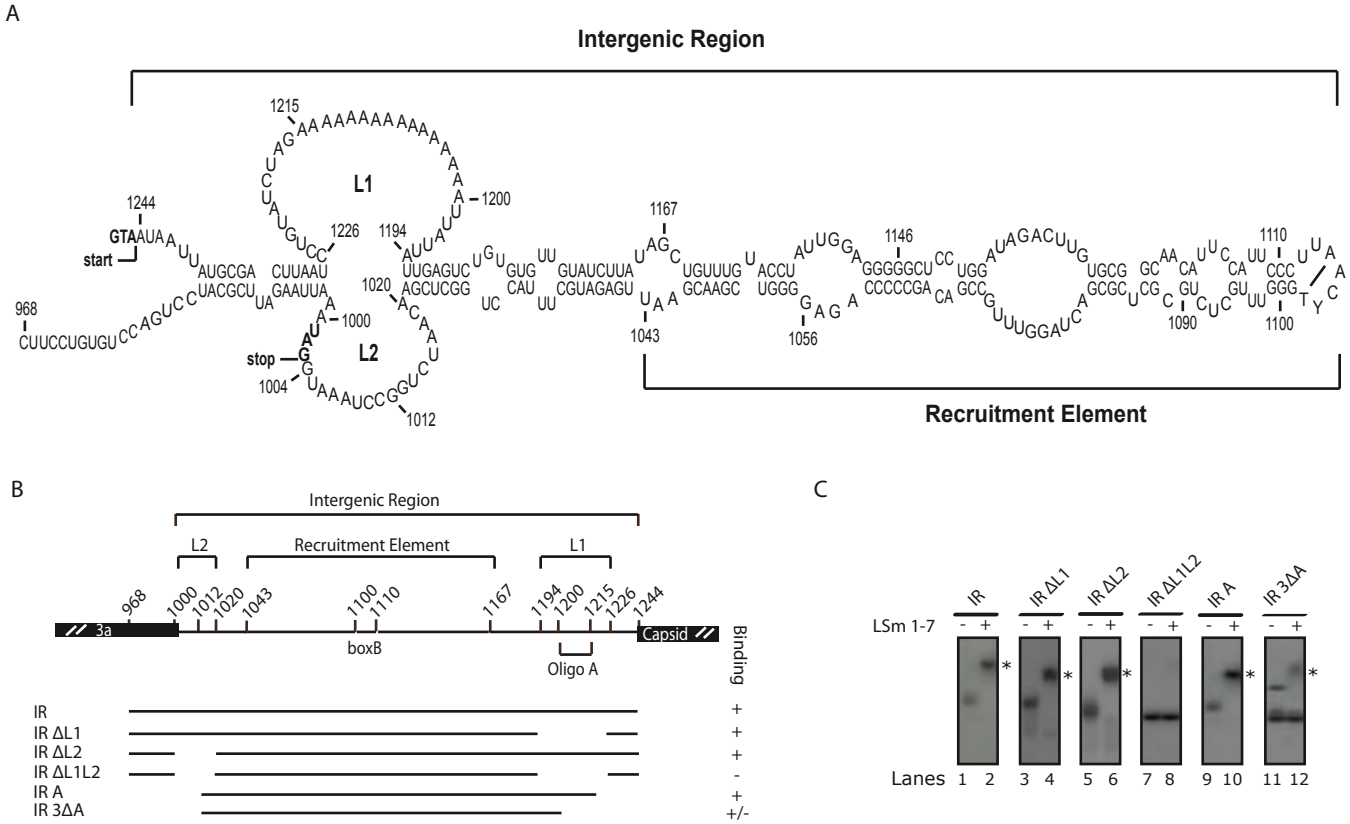

A
Intergenic Region
L1
L2
start
stop
GTA
Recruitment Element
B
Intergenic Region
L2
Recruitment Element
L1
3a
boxB
Oligo A
Capsid
Binding
IR
IR ΔL1
IR ΔL2
IR ΔL1L2
IR A
IR 3ΔA
C
LSm 1-7
Lanes 1 2 3 4 5 6 7 8 9 10 11 12

Figure 6

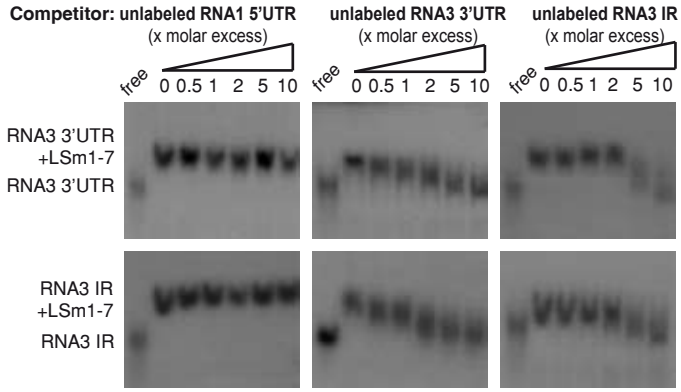

Competitor: unlabeled RNA1 5'UTR
(x molar excess)
free 0 0.5 1 2 5 10
unlabeled RNA3 3'UTR
(x molar excess)
free 0 0.5 1 2 5 10
unlabeled RNA3 IR
(x molar excess)
free 0 0.5 1 2 5 10
RNA3 3'UTR
+LSm1-7
RNA3 3'UTR
RNA3 IR
+LSm1-7
RNA3 IR

Figure 7

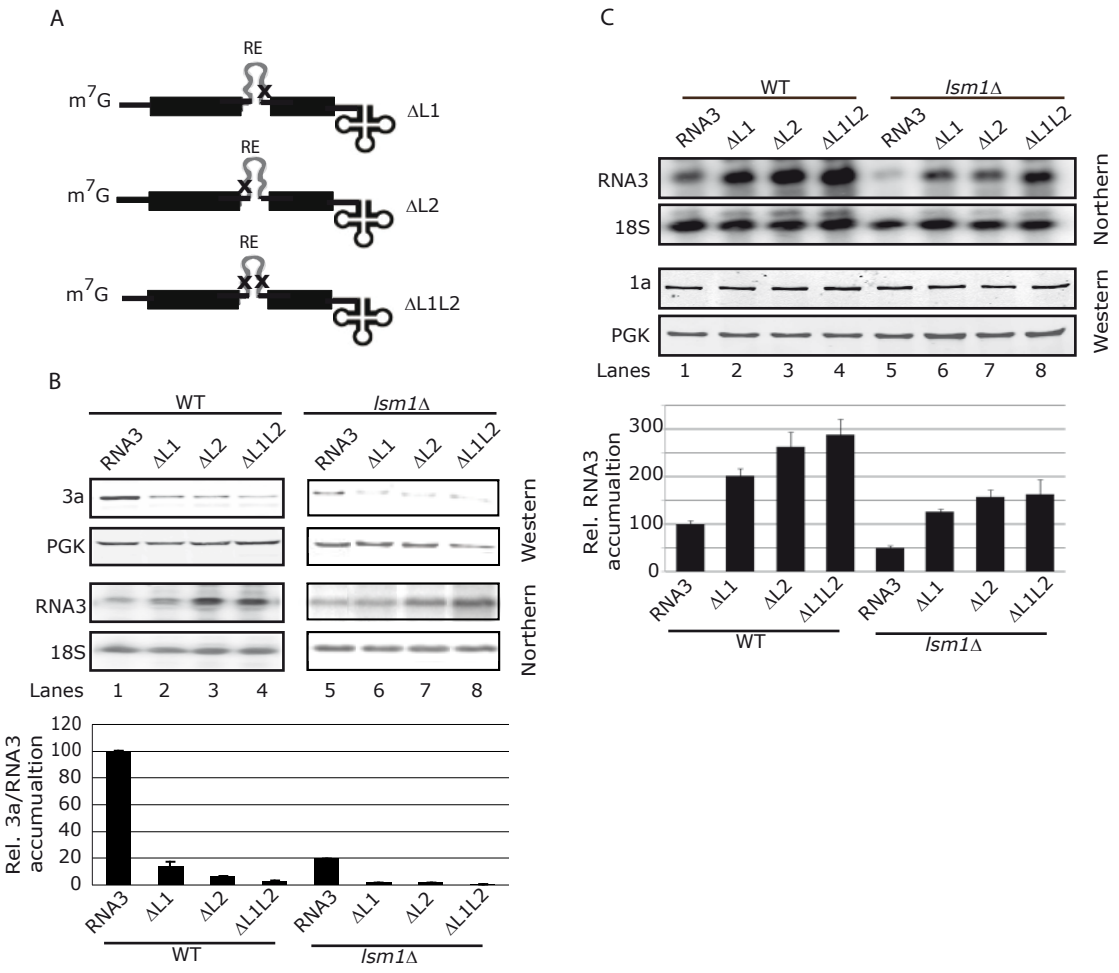
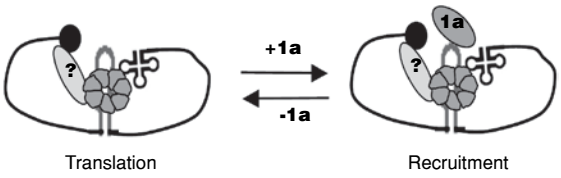

A
RE
m7G
ΔL1
ΔL2
ΔL1L2
B
WT
lsm1Δ
RNA3
ΔL1
ΔL2
ΔL1L2
3a
PGK
RNA3
18S
Western
Northern
Lanes 1 2 3 4 5 6 7 8
Rel. 3a/RNA3 accumualtion
120
100
80
60
40
20
0
C
1a
Rel. RNA3 accumualtion
300
200
100
0

Figure 8



Supplementary Table 1: Primer Sequences used to generate constructs for *in vivo* experiments

Construct	Primer Name	Sequence from 5`- to 3`end
pIR1 and pIR3	del75-116	CCA TGT TTG TGG ATA TTC TAT GTT GTG TGT CTG AGT TTA ATT CAG CGT ATT AAT AAT GTC GAC TTC AGG AAC
pIR1 and pIR3	del75-116 rev	GTT CCT GAA GTC GAC ATT ATT AAT ACG CTG AAT TAA ACT CAG ACA CAC AAC ATA GAA TAT CCA CAA ACA TGG
pIR2	del55-75	GGA CTT CCT GTG TCC AGT CCT ACG CTT AGA ATT TAA GCT CGG TCC ATT TCG TAG AG
pIR2	del55-75 rev	CTC TAC GAA ATG GAC CGA GCT TAA ATT CTA AGC GTA GGA CTG GAC ACA GGA AGT CC

Supplementary Table 2: Primer Sequences used to generate constructs for gel shift assays

Amplified Region	Primer Name	Sequence from 5`- to 3`end
BMV 1 5'UTR	BMV1_5'UTR_fwd:	TAA TAC GAC TCA CTA TAG GGT AGA CCA CGG AAC GAG GTT C
BMV 1 5'UTR	BMV1_5'UTR_rev:	TTT GTT GGT GAA AAA CAA AGA ACA
BMV 2 5'UTR	BMV2_5'UTR_fwd:	TAA TAC GAC TCA CTA TAG GGT AAA CCA CGG AAC GAG GTT C
BMV 2 5'UTR	BMV2_5'UTR_rev:	CTT GGT GAT AGT AGA AAG AAC AAG CAC
BMV 3 5'UTR	BMV3_5'UTR_fwd:	TAA TAC GAC TCA CTA TAG GGT AAA ATA CCA ACT AAT TCT CGT TCG AT
BMV 3 5'UTR	BMV3_5'UTR_rev:	CGG GAA CAA AAA CAG TAT CAC TAC TG
BMV 1 3'UTR, BMV1 NTLS	BMV1_3'UTR_fwd:	TAA TAC GAC TCA CTA TAG GTG ATG CGC TTG TCT CTG TGT
BMV 2 3'UTR, BMV2 NTLS	BMV2_3'UTR_fwd:	TAA TAC GAC TCA CTA TAG GTG ATC GGT TCT ATG ATA TAT GAA CCT AAG
BMV 3 3'UTR, BMV3 NTLS, NTLS+EB3, NTLS+EB3+B1AD, NTLS+EB3+B1AD+C	BMV3_3'UTR_fwd:	TAA TAC GAC TCA CTA TAG GTA GTG CCC CTG CTC GGA
BMV 1 3'UTR, BMV 2 3'UTR, BMV 3 3'UTR, BMV 1 TLS, BMV 2 TLS, BMV 3 TLS, TLS+(1903-1943), TLS+(1874-1943), TLS+(1844-1943)	BMV _3'UTR_rev:	TGG TCT CTT TTA GAG ATT TAC AGT GTT TT
BMV 1 NTLS	BMV1 _NTLS_rev:	ACT ATC GGT TAT AGC ACC G
BMV 2 NTLS, BMV 3 NTLS	BMV2 _3_NTLS_rev:	ACT ATC AGT TAT TGT ACT GAT TCA ACA AGC
BMV 1 TLS, BMV 2 TLS, BMV 3 TLS	BMV_TLS_fwd:	TAA TAC GAC TCA CTA TAG GCG TGG TTG ACA CGC AGA CCT C
NTLS+EB3	BMV3_1977_rev	AGA CAC TCT TGT AAG AGG TC
NTLS+EB3+B1AD	BMV3_2032_rev	ACG AAC CCC TAA GGG TTC TTC CG

NTLS+EB3+B1AD+C	BMV3_2063_rev	GCA TTC TAA GAC TTG CTA TG
TLS+(1903-1943)	BMV_1903_fwd	TAA TAC GAC TCA CTA TAG GAA GGT TAA AAG CTT GTT GAA
TLS+(1874-1943)	BMV_1874_fwd	TAA TAC GAC TCA CTA TAG GAG ACCCTGTCCAGGTAGGA
TLS+(1844-1943)	BMV_1844_fwd	TAA TAC GAC TCA CTA TAG GTT AAA GTC ACA GGC CCC TT
IR, IRΔL1, IRΔL2, IRΔL1L2	BMV3_968_fwd	GGG ATC CGT AAT ACT CAC TAT AGG CTT CCT GTG TCC AG
IR, IRΔL1, IRΔL2, IRΔL1L2	BMV3_1244_rev	GGA ATT CTA TTA ATA CGC TG

Underlined: T7 promoter

8.2 Arginine Methylation of Mammalian Pre-mRNA Cleavage Factor I

Arginine methylation in subunits of mammalian pre-mRNA cleavage factor I

Martin G, Ostareck-Lederer A, Chari A, Neuenkirchen N, Dettwiler S, Blank D, Rüegsegger U, Fischer U, Keller W: **RNA**, in revision

Thesis author´s contribution:

Conception:	20 %
Experimental contribution:	10 %
Formulation of results:	10 %

Arginine methylation in subunits of mammalian pre-mRNA cleavage factor I

Georges Martin[1], Antje Ostareck-Lederer[4], Ashwin Chari[5], Nils Neuenkirchen[5], Sabine Dettwiler[3], Diana Blank[2], Ursula Rüegsegger[3], Utz Fischer[5] and Walter Keller[3*]

[1]Computational & Systems Biology, [2]Structural Biology & Biophysics, [3]Growth & Development, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

[4]Department of Surgical Intensive Care Medicine, University Hospital, RWTH Aachen, D-52074 Aachen, Germany

[5]Theodor Boveri Institute, Biocenter at the University of Würzburg, D-97074 Würzburg, Germany

*Corresponding author:
Tel.: +41 61 267 2060, Fax: +41 61 267 2079
E-mail address: walter.keller@unibas.ch

Keywords: protein arginine methyl transferase, PRMT2, PRMT5, mammalian cleavage factor I, CF I_m

Abstract

The 3' ends of eukaryotic messenger RNA precursors are processed by the 3' processing complex. Part of this complex in higher eukaryotes is cleavage factor I, which is composed of subunits of 25 and either 59 or 68 kDa (CF I_m25, CF I_m59, CF I_m68) whereas yeast does not have any homologues of either CFI_m subunits. To investigate posttranslational modifications in factors of the 3' processing complex, we systematically searched for protein arginine methyl transferases (PRMTs), enzymes that transfer methyl groups from S-adenosyl methionine to arginine residues within polypeptide chains resulting in mono- or dimethylated arginines. We found that CF I_m68 and the poly(A) binding protein PABPN1 were methylated by Hela cell extracts *in vitro*. We further showed that the catalytic subunit PRMT5 of the "methylosome" requires WD45 for symmetric dimethylation of CF I_m68, whereas pICln, the third polypeptide of the complex, is stimulatory. Sites of methylation in CF I_m68 were found exclusively in a GGRGRGRF motif that is conserved in vertebrates.

Yeast two hybrid assays revealed an interaction of CF I_m59 with the arginine methyl transferase PRMT2. However, neither recombinant PRMT2 nor extracts of Hela cells had methylation activity on CF I_m59 as substrate. Antibodies against symmetrically dimethylated arginines reacted with purified CF I_m68 but not with CF I_m59. In contrast, antibodies against asymmetrically methylated arginines mainly reacted with purified CF I_m59. Our results suggest the existence of an elaborate posttranslational modification system in these RNA processing complexes.

Introduction

Messenger RNA precursors (pre-mRNAs) in eukaryotes are processed extensively before they serve as templates for protein synthesis. Pre-mRNAs are cleaved at specific sites downstream of the open reading frame by the cleavage and polyadenylation complex, followed by addition of a poly(A) extension to the upstream cleavage fragment. Factors in this complex that have been characterized extensively include the cleavage and polyadenylation specificity factor (CPSF) consisting of eight polypeptides (CPSF-30, -73, -100, and -160, as well as hFip1, WDR33, Rbbp6 and PP1), three polypeptides of the cleavage stimulation factor (CstF), (mammalian) cleavage factor I and cleavage factor II (CF I_m and CF II_m, two subunits each) and a poly(A) polymerase (PAP; for reviews see (Wahle & Keller, 1996; Zhao et al., 1999; Martin & Keller, 2007; Shi et al., 2009). CF I_m is a complex that contains one large polypeptide of either 59 or 68 kDa (CF I_m59, CF I_m68; Suppl. Figure S1) and a 25 kDa homodimer (CF I_m25) (Rüegsegger et al., 1996; Coseno et al., 2008). Splice variants of both factors were also described: CF I_m59_S1 as a short form of CF I_m59 and CF I_m68_L1 as long form of CF I_m68 with a size of 72 kDa (previously called CF I_m72 kDa; Suppl. Figure S2 A, B; Figure 1 A). CF I_m59/59_S1 and CF I_m68/68_L1 are encoded by two different genes (Rüegsegger et al., 1996), yet share a high degree of homology. Both contain an RNA recognition motif (RRM) fold at the N-terminus, a proline-rich region in the center and an SR-like domain at the C-terminus (Suppl. Figures S1 and S2). CF I_m25 together with one of the two larger subunits is sufficient to stimulate cleavage although it is not known in which function the large subunits differ (Rüegsegger et al., 1996; Rüegsegger et al., 1998). Analysis of the subnuclear localization of CF I_m68 has identified it in speckles and paraspeckles and found that its location varied during the cell cycle (Cardinale et al., 2007). Moreover, sequence specific RNA binding of the CF I_m68 subunit to motifs near the cleavage

site of its own pre-mRNA have been reported to suppress cleavage (Brown & Gilmartin, 2003). CF II is composed of the subunits hClp1 and hPcf11. Recent purification and proteomic analysis of the 3' processing complex suggests that this structure is much more complex and contains many additional polypeptides that interact with the core complex (Shi et al., 2009).

After cleavage of the pre-mRNA, a poly(A) tail of approximately 250 residues is added to the upstream cleavage fragment by a PAP. The polyadenylation reaction is stimulated by processivity factors and the length of the tail is measured by nuclear poly(A) binding protein 1 (PABPN1) which is also part of the cleavage and polyadenylation complex (Wahle, 1991).

It has been postulated that interactions between CPSF and U2 snRNP result in the coupling of pre-mRNA 3' end formation and splicing (Kyburz et al., 2006). In addition, regulatory interactions between PAP and components of the splicing machinery have been found. The effects were found to consist of either suppression of PAP activity by U1A (Gunderson et al., 1994) or U1-70K (Gunderson et al., 1998) or stimulation of splicing by the PAP-U2AF65 interaction (Vagner et al., 2000). In this context, a class of regulatory splicing factors, called SR proteins, was described and members of the family were found to be involved in splice site selection, splicing regulation and regulation of alternative splicing (Graveley, 2000). These SR proteins contain characteristic stretches of repetitive arginine/serine (RS) motifs. CF I_m59 and CF I_m68 were similarly found to have RS-like motifs that contain in addition to arginine/serine frequent arginine/aspartate (RD) repeats (Rüegsegger et al., 1996). In addition, the RS-like domain in CF I_m59 was found to be involved in the functional coordination of splicing and 3' end processing (Millevoi et al., 2006) and the RS-like domain of CF I_m68 was demonstrated to interact with several SR proteins (Dettwiler et al., 2004). Moreover, CF I_m68 was found to interact with CF I_m25 at its N-terminal RRM fold and this

interaction strongly stimulated RNA binding (Dettwiler et al., 2004). Similar effects were seen with the RRM of CF I_m59 in complex with CF I_m25 (S. Dettwiler and W. Keller, unpublished). Furthermore, distinct motifs within the proline-rich domains present in CF I_m59 and CF I_m68 serve as interaction sites for WW proteins (Ingham et al., 2005). This indicates the existence of a larger interaction cascade between CF I_m and other processes such as transcription, splicing, chromatin remodeling and actin polymerization.

Post-translational modifications are found in many factors belonging to the mammalian pre-mRNA 3' end formation complex (reviewed in (Ryan & Bauer, 2008). They include phosphorylation as in CF I_m59, Cstf-64 and CstF-77, Pcf11, CPSF-100 and CPSF-160 (Olsen et al., 2006) and PAP (Colgan et al., 1996) and lysine acetylation that regulates the interaction between CFI_m25 and PAP (Shimazu et al., 2007). In addition, CF I_m68 was found to recruit the CBP acetyltransferase to modulate the localization of PAP through an acetylation- deacetylation cycle (Shimazu et al., 2007). Other examples are sumoylation, such as in PAP (Vethantham et al., 2008) and arginine methylation in CPSF-100, CF I_m25, CF I_m68 and PABPN1 as found in a screen with the sym10 antibody (Boisvert et al., 2003). All these modifications are thought to be essential for proper functioning and regulation of 3' end processing.

Arginine methylation is a widespread posttranslational modification of proteins in eukaryotes. The enzymes responsible for this activity belong to a family of protein arginine methyltransferases (PRMTs) of which about ten different members have been characterized (reviewed in (Bedford et al., 2000; Bedford & Richard, 2005; Lee & Stallcup, 2009). PRMTs can catalyze the transfer of one or two methyl groups from S-adenosylmethionine (SAM) to arginine guanidino nitrogen atoms. Intermediates with single methyl additions are ω-N^G monomethyl arginines. There are two classes of PRMTs that differ in the mode of attachment

of the second methyl group (Figure 2). Type I PRMTs produce asymmetric ω-N^G,N^G-dimethylarginine (ADMA) and include PRMT1, PRMT2, PRMT3, PRMT4/CARM1, PRMT6 and PRMT8 (Lin et al., 1996; Tang et al., 1998; Schurter et al., 2001; Frankel et al., 2002; Qi et al., 2002; Lee et al., 2005a). The type II PRMTs PRMT5, PRMT7 and PRMT9 promote the formation of symmetric ω-N^G,N'^G-dimethylarginines (SDMA) (Branscombe et al., 2001; Miranda et al., 2004; Lee et al., 2005b). No activity could be demonstrated so far for PRMT2 and PRMT8 and their classification is based on homology. Substrates for PRMTs are often glycine and arginine-rich (GAR) sequence motifs (see review by (Bedford & Richard, 2005) and references therein (Figure 1 B). PRMTs can recognize different arginine containing substrates as recently illustrated in a crystal structure of PRMT1 in complex with the reaction products S-adenosylhomocysteine (SAH) and a GAR motif (Zhang & Cheng, 2003). The structure contains three different peptide binding channels as possible alternate binding sites for varying substrates.

Cellular functions of PRMT activities are very diverse and can involve transcriptional regulation by histone arginine methylation, RNA processing, signal transduction, DNA repair and protein-protein interactions to name just a few (reviewed in (Pahlich et al., 2006). In pre-mRNA processing and mRNA transport several proteins have been described to be modified by PRMTs. An example is the poly(A)-binding factor PABPN1 that is modified by PRMT1, PRMT3 and PRMT6. There, arginine methylation does not prevent self-association of PABPN1 and it is unclear whether arginine methylation has any effect on PABPN1 aggregation that causes oculopharyngeal muscular dystrophy (Fronz et al., 2008). It has been found that PRMTs fulfil distinctive, non-redundant functions in cells and that some PRMTs display different subcellular localization in different cell types, implicating cell- and tissue-specific mechanisms for regulating PRMT functions (Herrmann et al., 2009).

Heterogeneous nuclear ribonucleoprotein K (hnRNP K) has been identified to be functionally modified by PRMT1. HnRNP K is an activator of the tyrosine kinase c-Src. Asymmetric dimethylation of five distinct arginine residues in hnRNP K reduces the interaction with and activation of c-Src and thereby inhibits the c-Src dependent phosphorylation of hnRNP K (Ostareck-Lederer et al., 2006). hnRNP K was also found to stimulate the transcription stimulatory activity of p53 when methylated as a result of UV radiation (Chen et al., 2008). PRMT1 appears to be the main PRMT in human cells with the largest spectrum of substrates (Pahlich et al., 2006).

Another well documented PRMT, PRMT5, acts in a complex (called the methylosome) with pICln1 and WD45 (also named methylosome protein 50 kDa or MEP50) on many different protein substrates (Meister et al., 2001; Friesen et al., 2002). The methylosome is also involved in methylation and subsequent assembly of the spliceosomal U snRNPs, a process that is assisted by a second complex containing the survival of motor neuron protein SMN (Meister et al., 2002; Neuenkirchen et al., 2008).

In this study, we demonstrate that recombinant PRMT5/WD45 as well as the full PRMT5/WD45/pICln methylosome complex can symmetrically dimethylate CF I_m68 at three arginines within a GAR motif. In contrast, homologous CF I_m59 was found to interact with the type I methyl transferase PRMT2, yet no methyl addition with recombinant PRMT2 or with cell extracts could be observed. Western blots and probing with specific antisera confirm mainly symmetric dimethylation in CF I_m68 and asymmetric dimethylation in CF I_m59, modifications reminiscent of type II PRMT5 and type I PRMT2, respectively. Possible roles of arginine modifications by methylation in CF I_m are discussed below.

Results

Systematic search for methyltransferase activity acting on 3' processing factors

We initiated our study with a systematic search for methyltransferase activity towards most of the known factors of the mammalian 3' processing complex available as recombinant proteins or proteins purified from Hela cells. Because CF I_m59 and CF I_m68 can only be expressed as fragments and not as full length proteins in *E. coli* we also used these fragments for the assays (Suppl. Table 1 and Figure 1 C). In addition, full length CF I_m proteins were expressed in insect cells (see Materials and Methods). Methylation reactions typically contained Hela cell nuclear or cytoplasmic extract, [^{3}H]-SAM and the purified proteins as substrates. Results are summarized in Suppl. Figure S3 A-C and show that with our assay conditions PABPN1 and a fragment of CF I_m68 (GST-68-GRP) were methylated. The 68-GRP fragment is encompassing a region between amino acids 191 and 407 downstream of the RRM and including the proline rich domain (Suppl. Figure 3 A, B and D). In contrast, fragments containing the N-terminal RRM or the C-terminal RS-like domain of CF I_m68 were not methylated in these tests. We also found methylation of CF I_m68_L1, a CF I_m68 splice variant (see below). Interestingly, with both Hela cell nuclear and cytoplasmic extracts no methylation of the recombinant CF I_m59 protein or its fragments was detected. Also, no methyl groups were incorporated into CF I_m68 purified from Hela cells, suggesting that these proteins were already fully methylated prior to purification (Suppl. Figure S3 B; CFI_m purif.). A strong signal of methylation was obtained with PABPN1 as substrate and this was more intense with nuclear than cytoplasmic extract whereas CF I_m68_GR was more intensely methylated by cytoplasmic extract (Figure 4). CF I_m25was not methylated by any extract or recombinant PRMT5 (Figure 4). Furthermore, PABPN1 purified from Hela extract (comparable to purified CF I_m68, see above), probably was fully methylated in cells and did

not accept further methylation of arginines ("PABPN1 purif" in Suppl. Figure S3 D). The results of the methylation screen are summarized in Supplementary Table 1.

PRMT1 was found to methylate factors involved in mRNA translational control, such as hnRNP K (Ostareck-Lederer et al., 2006). Because PRMT1 appeared as a likely candidate for methylation of CF I_m we tested methyl transfer to the GST-68_GRP fragment in the PRMT1 deletion cell line ES (-/-) by incubation of substrates with cell extract. GST-68_GRP was methylated by both the ES(+/+) and the ES(-/-) extracts whereas hnRNP-K as control substrate for PRMT1 was not methylated by the ES(-/-) extract (lanes 4 and 8). Thus, PRMT1 could be excluded as CF I_m68-specific methyl transferase. When PABPN1 was incubated with ES(-/-) extract it generated a signal most likely coming from remaining PRMT3, PRMT5 or PRMT6 (lanes 3 and 7, Figure 3), PRMTs acting on PABPN1 in addition to PRMT1 (Fronz et al., 2008) (and Figure 4; see below). Methylation of PABPN1 was extensively characterized recently (Kühn et al., 2003; Fronz et al., 2008) and the protein was found to be asymetrically dimethylated by PRMT1, PRMT3 and PRMT6.

Two hybrid screen with CF I_m59 and CF I_m68

To search for factors interacting with CF I_m59 and CF I_m68 we performed a yeast two-hybrid screen (not shown, see Materials and Methods for details). We found an interaction between the SH3 domain of protein arginine methyl transferase PRMT2 and CF I_m59 whereas no interaction of any PRMT with CF I_m68 was detected in this screen. PRMT2 was determined to be a PRMT based on sequence homology, yet no activity has been reported for this protein to date. PRMT2 contains an SH3 domain in its N-terminus and the smallest common interaction domain in the several two-hybrid positives we found to be within the SH3 domain (Figure 5A).

To confirm the interaction of PRMT2 with CF I_m59 we performed GST pulldown experiments with recombinant human PRMT2. The results indicate strong specific binding of $[H]_6$-PRMT2 to the N-terminal fragment of CF I_m59 and weaker binding to a C-terminal fragment (Figure 5B). Weak binding to several fragments of CF I_m68 was detected as well. These interactions seem unspecific because in our two-hybrid experiments no interaction was found between PRMT2 and CF I_m68 and no such interaction has been described in the databases. In contrast, the interaction between CF I_m59 and PRMT2 was also described in a proteome scale screen (Rual et al., 2005).

Partial purification of methyltransferase activity from Hela cells and identification of methyl transferase by mass spectrometry

In order to directly identify CF I_m-specific methyl transferases and to determine the enzyme responsible for the activity by mass spectral analysis we enriched the activity in Hela extracts by chromatographic methods. Hela nuclear and cytoplasmic fractions were prepared as described (Rüegsegger et al., 1996). We used cytoplasmic extract for the purification of the activity because we consistently found higher methyltransferase activity specific for CF I_m68 in this fraction compared to nuclear extract (Suppl. Figure S3 B). In contrast, very weak activity specific for CF I_m59 was detected with fractionated cell extracts and was consistently lost during the subsequent purification steps.

Starting with 20 ml cytoplasmic extract, CF I_m68-specific PRMT activity was enriched according to a scheme described in Materials and Methods with GST-68_GRP as test substrate. All affinity columns were low pressure type (except for the final Phenyl Superose and MonoQ columns) and all were run on a FPLC system (from GE Healthcare, see Materials and Methods). Fractions with peak activity from the gel filtration column were then

submitted to mass spectral analysis. Consistently, PRMT5 and WD45 appeared in the list of identified peptides at positions with high ranks (Table 1). When the methylation activity was purified over six columns PRMT5 and WD45 were always present with several peptides. In addition, when column fractions were separated on SDS PAGE gels and gel slices in the range between 68 to 75 kDa (predicted M.W. of PRMT5 is 73 kDa) were analyzed, the number of peptides originating from PRMT5 with high scores was further increased. No other member of the PRMT family and no peptides of any 3' processing factors appeared in the mass spectra. In addition, the third subunit of the methylosome, pICln, was never found in the mass data. This was a first indication that WD45 is essential and that the PRMT5/WD45 complex is sufficient for methyl transferase activity acting on CF I_m68 (see below). Apparently, pICln is loosely associated with PRMT5 and dissociated during purification under our conditions. In support of these findings, pICln has recently been reported to be an assembly chaperone specific for spliceosomal Sm proteins and appears to be only transiently bound to PRMT5/WD45 (Chari et al., 2008). Furthermore, fraction 7 of the gel filtration column (results of mass spectral analysis for this fraction listed in Table 1) was also used for methylation assays depicted in Suppl. Figure S3 D with substrate PABPN1 and GST-68_GRP.

These results suggested that the GST-68_GRP substrate was methylated by the PRMT5/WD45 complex and that this fragment must have a GAR motif to be a true substrate. In fact, a predicted GAR motif (GGRGRGRF) is located at residues 201-207 of CF I_m68. This is consistent with results of a proteomic analysis of arginine methylated protein complexes in which CF I_m68 was identified as one of the symmetrically dimethylated proteins and to have modifications at this motif (Boisvert et al., 2003).

Generation of recombinant methylosome subunits

To facilitate the study of methylation by the methylosome we made several attempts to reconstitute a recombinant methylosome complex. Expression of the three subunits in *E. coli* resulted in efficient expression of PRMT5 and pICln but with insoluble WD45. Surprisingly, the resulting PRMT5/pICln complex was highly active in methylating the Sm protein D1, but inactive with the GST-68_GRP substrate (Figure 6). Finally, a dicistronic vector construct with PRMT5 and WD45 could successfully be expressed in insect cells, resulting in soluble proteins. The complex was fully reconstituted with bacterially expressed pICln if required. As expected from the mass spectrometry data, the PRMT5/WD45 complex had high levels of activity with the GST-68_GR protein substrate (Figure 6) and addition of pICln to the reaction further stimulated this activity. Under these experimental conditions, pICln was as efficient to stimulate the activity of PRMT5 as was WD45 or both together (Figure 6).

Since the reactions in these tests went to completion, differences in activity with the various subcomplexes are not obvious, such that kinetics tests were needed to visualize the dependence on substrate concentrations. Results of steady state kinetics tests clarified the contribution of pICln for CF I_m68 methylation (Table 2). A protein-substrate-dependent Michaelis constant K_M was several fold lower when pICln was included in the reaction in addition to PRMT5 and WD45. This suggests that the complete methylosome with all three polypeptides is responsible for methylation of CF I_m68 *in vivo*. Also, V_{max} was more than four times higher with the ternary complex, indicating that the complete methylosome acts at the full catalytic rate. Moreover, the results of a time course experiment indicate a stimulation of the second methyl addition to monomethylated arginines when the full methylosome complex was present in the reaction (results not shown).

PABPN1 could also be methylated in vitro by the PRMT5/WD45/pICln complex (Figure 4) or the PRMT5/WD45 subcomplex (Suppl. Figure S3 D). Interestingly, this reaction was strongly inhibited by the addition of SV40 or poly(A) RNA to recombinant PABPN1, suggesting a function of methyl modifications in protein/RNA interaction (Suppl. Figure S3 D).

Identification of methylated arginines in CF I_m68 by deletion analysis and site directed mutagenesis

To determine the substrate arginines for PRMT5 within CF I_m68, further subfragments of CF I_m68_GRP were generated and the proline rich domain was separated from the GGRGRGRF motif located at position 201-207 (Figure 1 C). Methylation was found exclusively in the recombinant protein fragment containing the GAR motif (GST-68_GR) whereas the proline rich domain (GST-68_P) that contains four single RG or GR motifs between amino acids 375-400, and also fragments representing the RS domain downstream of residue 405 were not methylated in the in vitro assays (Figure 7).

Mutation of single arginines to alanine within the GAR motif of CFIm68 resulted in variable effects on methylation activity when 100 ng of recombinant PRMT5/WD45 were present in the assay (Figure 7A). Mutation of R206 or of all three arginines (R202A, R204A, R206A) resulted in complete loss of methylation whereas mutation of R202 and R204 had less effect. When the concentration of the PRMT5/WD45 complex in the reaction was increased to 1 µg, methylation activity was completely lost only when all three arginines were mutated to alanine (Figure 7B). And obviously, when the products were probed for symmetric dimethylation, mutation of all three arginines to alanine was required to abolish the activity (Figure 7C).

Because these assays were done with protein substrate concentrations of 3.3 mM (almost 15-fold the K_M; Table 2), we defined the kinetics for the second step of the methylation reaction resulting in symmetric dimethyl arginines. Figure 8 depicts the results of an assay where the concentration of the protein substrate was titrated between 0.1 and 1.8 μM. Resulting dimethyl groups on arginines were visualized by Western blots probed with sym10 antibody. The results indicate that an arginine replaced by alanine (R204A) in the center of a group of three arginines has a rather strong inhibitory effect on addition of a second methyl group. This seems to result from a low reaction rate indicated by a flat response to substrate stimulation. In contrast, mutants R204A and R206A show a high apparent K_M, indicated by a steep curve of response and a possibly high catalytic rate.

Mass spectral analysis of an in vitro methylated recombinant CF I_m68_GR fragment

To further characterize the GAR motif with respect to its methylation pattern the recombinant GST-68_GR fusion protein was methylated with the recombinant PRMT5/WD45/pICln complex. Products containing SDMAs were immunoprecipitated with SDMA-specific sym10 antibodies, separated by SDS gel electrophoresis and bands corresponding to the GST-68_GR protein excised and submitted to mass spectral analysis (see Materials and Methods for details). The purpose of the gel was to separate the GST-68_GR protein from the methylosome components.

One tryptic peptide was detected with the GAR motif in three peaks of additional masses accounting for the addition of two, three or four methyl groups (Figure 9). This indicates that one of the arginines must be dimethylated; only this would allow immuno-precipitation by the sym10 antibody. The two other species may have either one (for triple methylation) or two additional monomethylations or a second dimethyl group, resulting in quadruple

methylation. The arginine at the C-terminus of the peptide is assumed to be unmethylated because trypsin does not cleave C-terminal to methylated arginines. Apparently, the presence of one doubly methylated arginine in the GGRGRGR sequence seems to prevent cleavage or leads to inefficient cleavage at the two other arginines within the motif.

Identification of differential methylation in purified CF I factors

Specific antibodies were used to test recombinant and purified CF I_m59 and CF I_m68 proteins for the presence of symmetric or asymmetric dimethylation (Figure 10). In addition to CF I_m68 expressed in Picchia pastoris and a fraction of partially purified CF II_m fractions that still contain a large proportion of CF I_m, fractions of CF I_m purified from Hela cells (Rüegsegger et al., 1996) were separated on SDS gels and probed with polyclonal antibodies specific for symmetrically (sym10) and asymmetrically methylated arginines (asym24) (Boisvert et al., 2003). CF I_m68 was found to be mainly symmetrically dimethylated; this is consistent with the fact that CF I_m68 is a substrate for PRMT5 (Figure 10, sym10 panel). Also, the longer form of CF I_m68, CF I_m68_L1 is symmetrically dimethylated; the signal is weaker because this form of the protein is present in smaller amounts in this fraction (fraction 40, Figure 3B in (Rüegsegger et al., 1996).

In contrast, CF I_m59 reacted with antibodies against ADMA. Although this is consistent with our results that CF I_m59 is not a substrate for PRMT5, it could be a substrate for type I PRMTs that include PRMT2. Activity for PRMT2 has not been found to date and our attempts to methylate CF I_m59 by recombinant PRMT2 failed (not shown). In addition, a weak signal of ADMA is also seen in CF I_m68 or CF I_m68_L1 (Figure 10, lane 3). CF I_m59 and CF I_m68 appear to be mono-methylated when expressed in *P. pastoris* (Figure 10, lane 4).

Discussion

The cleavage factor subunits CF I_m59 and CF I_m68 were found to undergo multiple functional interactions with adaptor proteins such as WW-proteins and SR splicing factors (Ingham et al., 2005). This suggests a role for CF I as a specialized adaptor in a pre-mRNA transcription and processing interactome. The assembly of such large protein complexes is often regulated by posttranslational modifications of the different subunits.

To investigate enzymatic events that lead to post-translational modifications in 3' processing factors we systematically investigated addition of methyl groups to purified and recombinant proteins by methyl transferases present in nuclear and cytoplasmic extracts. A first in vitro methylation screen identified two proteins as substrates, PABPN1 and CF I_m68. Chromatographic enrichment of methylation activity towards CF I_m68 in cytoplasmic extracts led us to the mass spectral identification of the arginine methyl transferase PRMT5 and a cofactor WD45, both belonging to the methylosome. The fact that no other PRMT was found in the mass spectral tests strongly suggests PRMT5 to be the major enzyme to be responsible for CF I_m68 methylation.

The site of methyl addition in CF I_m68 could be tracked down to a single GAR motif containing three characteristic tandem GR motifs. We could partially dimethylate the GAR motif arginines in vitro by use of a reconstituted recombinant methylosome. The two subunits PRMT5 and WD45 were essential and apparently sufficient for activity on CF I_m68. Steady state kinetics tests indicated that the catalytic rate with the full PRMT5/WD45/pICln complex to be higher with the GST-68_GR substrate than for the PRMT5/WD45 complex but also the K_M value to be lower, resulting in a higher catalytic efficiency, expressed as k_{cat}/K_m, for the ternary complex. This suggests a function for WD45 as specificity factor in the methylation of CF I_m68. In contrast, with spliceosomal Sm D1/D2 heterodimer we found both the

PRMT5/pICln and the PRMT5/WD45 subcomplex to be active. pICln was recently identified as an assembly chaperone that promotes spliceosomal snRNP formation together with the SMN complex (Chari et al., 2008). The stimulating effect of pICln on activity with CF I_m68 is most likely the result of a more stable complex. It remains to be determined whether the WD45 subunit is a specificity factor for proteins other than CF I_m68.

Interactions with CF I_m68 that could be regulated by methylation of the GAR motif could be those with SR proteins 9G8, Srp20 and hTra2b (Dettwiler et al., 2004). More likely, interactions could be regulated by WW proteins, factors found to bind to proline rich motifs in their interactors (Ingham et al., 2005). However, methylation of SAM68, a Src-kinase adaptor protein that also contains a GAR motif upstream of a proline-rich region similarly to CF I_m68 was found to bind less efficiently to an SH3 domain-containing binding partner via its proline rich domain when methylated. In contrast, binding to a WW-domain containing binding partner was not affected by methylation (Bedford et al., 2000).

A candidate CF I_m68 interactor containing three SH3 domains is Nck2 (Figure 11). Nck1 and Nck2 are adaptor proteins composed of one SH2 and three SH3 domains. They are believed to couple activated receptor kinases and their substrates to downstream effectors through their SH domains and they were mainly found in actin cytoskeleton reorganization (reviewed in (Buday et al., 2002). In a proteome scale interaction screen both CF I_m59 and CF I_m68 were found to interact with Nck2 (Rual et al., 2005). In contrast, in this screen only CF I_m59 was found to interact with the SH3 domain of PRMT2 but not CF I_m68. Nck2 was also found to act as an adapter protein involved in the control of eIF2a phosphorylation and eIF2-dependent signaling response to ER stress. Nck2 maintains low levels of eIF2-S1 phosphorylation by promoting its dephosphorylation by PP1 (Latreille & Larose, 2006). Nck2, in addition to three SH3 domains, contains one SH2 domain that was found to bind to

specific protein motifs with a phosphotyrosine in the center (Frese et al., 2006). Both CF I_m59 and CF I_m68 contain a sequence conforming to this consensus at their N-terminus (Suppl. Figure S4) and it remains to be shown whether Nck2 does indeed interact with these two cleavage factor subunits and in this way could form a link to translation and signaling. Moreover, in a recent proteomic characterization of a mammalian 3' processing complex, several translation factors, including eIF2 were present (Shi et al., 2009).

In a two-hybrid screen we found CF I_m59 to bind PRMT2, a PRMT with an SH3 domain. A possible function of this interaction is unclear; in particular because no activity for PRMT2 is known. In addition, the binding site for the PRMT2 SH3 domain within CF I_m59 is not known but is is likely to bind to motifs in the proline rich domain. Furthermore, the two subunits CF I_m59 and CF I_m68 appear to have quite different interaction partners in particular at the RS-like domain (Figure 11). The question remains whether methylation in the GAR motif of CF I_m68 and possibly in CF I_m59 regulates interactions also at the RS-like domain located at the C-terminus. Also, CF I_m59 was found to interact in vitro with U2AF65 whereas CF I_m68 was not (Millevoi et al., 2006). However, these interactions were detected between the RS-like domain fragment of CF I_m59 and an RS domain in U2AF65 and thus were independent of the presence of a GAR motif.

We made several attempts to find a function for methylated arginines in the GAR motif of CF I_m68. One possibility to generate the two methylation states is *in vitro* translation of proteins in a coupled reticulocyte lysate system in the presence of either SAM or S-adenosyl-homocysteine (SAH, a methyltransferase inhibitor) in order to force or inhibit methylation in the reaction, respectively (for details see Materials and Methods). Reticulocyte lysate contains several active PRMTs, in particular PRMT5 (Denman, 2008). In these tests we could not find an influence of methylation of the GAR motif upon the interaction between

CF I_m68 and CF I_m25, and also not when RNA was present in the reaction (not shown). In GST pulldowns we initially found a reduction of the interaction of in vitro methylated CF I_m59, CF I_m59S1, CF I_m68, CF I_m68_L1 and 68-GR with GST-CF I_m25 when compared to CF I_m proteins that were in vitro translated in the presence of the methylation inhibitor SAH. However, when either RNAse or total Hela RNA was included in the reaction, the correlation with methylation was no longer significant (not shown). We also found no indication that either short or long forms of CF I_m59 and CF I_m68 resulting from exon skipping or inclusion had an influence on the interaction with CF I_m25 (results not shown).

To confirm that PRMT5 is responsible for methylation of CF I_m68 we used RNAi to knock down PRMT5 expression in HEK293 cells. The results indicated that PRMT5 is a very stable protein in the cells. Although we were able to dramatically reduce the level of PRMT5 after 3-5 days by siRNA transfection, the level of symmetric dimethylation in immunoprecipitated CF I_m68 or CF I_m68_L1 was not reduced (Suppl. Figure S5). This indicates that CF I_m proteins are not only very stable but that their modifications are also very stable. Also, our observation that dimethylation in CF I_m and of PABPN1 purified from Hela cells is stable and is not lost during the purification process and in addition can not be further dimethylated in vitro (Suppl. Figure S3 B) indicates that these modifications are extremely stable and may not be reversible (Suppl. Figure S3 B and C). In addition, it seems that symmetric dimethylation has a positive effect on CF I_m cleavage stimulatory activity and that during purification of CF I_m unmethylated CF I_m (that is assumed to be inactive) was purified away from methylated and active CF I_m (Rüegsegger et al., 1996). However, it has been demonstrated that cleavage activity can be reconstituted with recombinant CF I_m expressed in *P. pastoris* and insect cells, a form of CF I_m that is only monomethylated (our results in Figure 10; (Rüegsegger et al., 1998; Dettwiler et al., 2004).

We also found interesting modifications in CPSF fractions purified from Hela cells. When we purified CPSF we found that sometimes these fractions were inactive in cleavage activity without an obvious reason (I. Kaufmann and W. Keller, unpublished data). When comparing active and inactive CPSF preparations on Western blots that were probed with anti-phosphotyrosine antibodies, we observed that active fractions contained CFSF-100 and CPSF-160 subunits phosphorylated at tyrosines whereas both subunits were not phosphorylated in inactive CPSF preparations. In contrast, inactive fractions of CPSF were found to be symmetrically and asymmetrically dimethylated when probed with sym10 and asym24 antibodies, modifications that are not found in active CPSF fractions (Suppl. Figure S6). This could be an indication for regulation of cleavage activity in CPSF by phosphorylation and methylation. These results suggest that the addition of phosphatase inhibitors during the purification of CPSF could be of advantage. This is consistent with recent work that demonstrated that phosphatase inhibitors could stimulate in vitro cleavage activity (Ryan et al., 2009).

In summary, our results give early insights into a complex system of posttranslational modifications in mammalian pre-mRNA cleavage factors with focus on CF I_m59 and CF I_m68. We have determined the basic determinants of symmetric arginine methylation by PRMT5 in CF I_m68. In CF I_m59, that we found mainly asymmetrically dimethylated, the role of the interacting PRMT2 is unclear. PRMT2, for which no enzymatic activity nor substrates have yet been identified, could also act as an adaptor to other factors if bound to CF I_m59. It is likely that arginine modifications in CF I_m contribute to the fine tuning of pre-mRNA processing, RNA transport or translation.

Table 1. Mass spectral analysis of samples from purification of methyl transferase activity

Sample	Protein	Accession	Position	Score	Top score	Peptides
GF-S200-6	PRMT5	Q4R5M3	95	10.21	118.22	1
GF-S200-6	WD45	XP_537028	145	10.14	118.22	1
GF-S200-7	PRMT5	Q4R5M3	16	40.21	150.15	4
GF-S200-7	WD45	XP_537028	51	20.18	150.15	2
GF-S200-6, gel	PRMT5	Q4R5M3	5	30.15	50.18	3
GF-S200-6, gel	PRMT5	Q4R5M3	6	30.16	90.2	3

In the column "Sample", GF-200-6/7 indicates gel filtration column Superdex-200, fraction 6 or 7 and gel indicates when trypsin digest was done in isolated gel slice. The column "Position" lists the position of the "Protein" (see there) in the list of identified peptides. The "Score" is the score given to the peptide of the protein; this value can be compared to the "Top score" (score of peptide at position 1).

Table 2. GST-68_GR -dependent steady-state kinetic parameters defining methyl transfer by two different PRMT5 complexes.

	K_M (μM)	V_{max} (nmol/min/mg)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
PRMT5-WD45	0.24	0.48	0.015	75'000
PRMT5-WD45-pICln	0.037	1.97	0.059	1'594'594

MATERIALS AND METHODS

Proteins and antibodies. Nuclear and cytoplasmic extracts were prepared with minor modifications as described (Dignam et al., 1983). For the cloning of CF I_m59, peptide sequencing was carried out from the same pool of human CF I_m and according to the same procedure as described for CF I_m68 (Rüegsegger et al., 1998). The sequence of CF I_m59 was submitted to the genbank database (accession AJ275970). CF I_m59 is also annotated under SwissProt/Uniprot accession Q8N684 or CPSF7_HUMAN and the sequence corresponding to the one depicted in Suppl. Figure S1 was designated as the "canonical sequence". The N-terminus of the 59 kDa protein was produced by digestion of the full length CF I_m59 clone with HindIII, resulting in a fragment encoding amino acids 1-250 and the product was subcloned into the Hind III site of the pGDV vector (Dichtl & Keller, 2001).

Reconstituted recombinant CF I_m25 and CF I_m68 subunits expressed in Picchia pastoris were described (Rüegsegger et al., 1998). For the Western blot shown in Figure 10, lane 4, 2 µl of fraction 25 from the original Phenyl Superose column were used.

Purified CF I_m from Hela cells was described (Rüegsegger et al., 1996). 5 µl of fraction 40 of the final MonoQ eluate was used for Western blots depicted in Figure 10 (CF I_m purif, lane 3). Partially purified CF I/II_m was a gift from Henk de Vries and was described (de Vries et al., 2000). Fraction 43 (10 µl) of Phenyl Superose column 4 was used for Western blots (Figure 10; CF II_m part, lane 5). CPSF-4 (fraction 4 of CPSF purified from Hela cells, inactive in cleavage) was a gift from Isabelle Kaufmann, cleavage competent CPSF was purified from Hela cells (Bienroth et al., 1991).

Because both CF I_m59 and CF I_m68 could not be expressed as full length proteins in bacteria, they were produced in insect cells or P. pastoris (used for Western analysis, Figure 10, lane 4). For baculovirus/insect cell expression of proteins in the Bac-To-Bac expression system

(Invitrogen), coding regions or fragments of CF I_m were inserted into the transfer vector p-FASTBAC-Hta and generation of recombinant proteins was done as described (Dettwiler et al., 2004).

For expression of CF I_m fragment fusions with GST (glutathione-S-transferase), the corresponding sections of the ORFs were cloned into the pGDV vector and expressed and purified as described (Dettwiler et al., 2004).

PRMT5 and WD45 were co-expressed in Sf21 insect cells with the MultiBac system (Berger et al., 2004). PRMT5 contained an N-terminal His-tag. Harvested cells were resuspended in 20 mM Hepes-NaOH (pH 8), 0.5 M NaCl, 10 mM imidazole, lysed by sonication and the lysate clarified by centrifugation at 40,000 rpm in a 45Ti rotor. The PRMT5/WD45 heterodimer was purified from the cleared lysate by Ni-NTA chromatography (Qiagen), followed by dialysis and anion exchange chromatography on a 1 ml HiTrap Q column (GE Healthcare). The final purification step was gel filtration on a Superdex 200 column (GE Healthcare) run in 20 mM Hepes-NaOH (pH 7.5), 0.2 M NaCl, 5 mM DTT.

The PRMT5/pICln heterodimer was expressed in BL21(DE3) cells induced with 1 mM IPTG at 25 °C for 5 hours from a tricistronic expression plasmid encoding PRMT5, WD45 and pICln with an N-terminal His-tag on pICln. WD45 was insoluble when expressed with this construct. Harvested cells were resuspended in 20 mM Hepes-NaOH (pH 8), 0.5 M NaCl, 10 mM imidazole, lysed by sonication and the lysate clarified by centrifugation at 40,000 rpm in a 45Ti rotor. The PRMT5/pICln heterodimer was purified from the cleared lysate by Ni-NTA chromatography (Qiagen). The final purification step was gel filtration on a Superdex 200 column (GE Healthcare) run in 20 mM Hepes-NaOH (pH 7.5), 0.2 M NaCl, 5 mM DTT.

Details for the bacterial expression and purification of pICln have been described elsewhere (Chari et al., 2008). For reconstitution of the heterotrimeric PRMT5/WD45/pICln complex,

the PRMT5/WD45 heterodimer was incubated with an equimolar amount of bacterially expressed pICln at 4 °C over night. The reconstitution mixture was then purified by gel filtration chromatography on a Superdex 200 column (GE Healthcare) run in 20 mM Hepes-NaOH (pH 7.5), 0.2 M NaCl, 5 mM DTT.

The $[His]_6$-PRMT2 clone was generated from reverse transcribed Hela cell total RNA and PCR amplification of the open reading frame with oligos GGG CAT ATG GCA ACA TCA GGT GAC TGT CCC AG (forward) and GGG GGA TCC TCA TCT CCA GAT GGG GAA GAC TTT TTC (reverse). Products were cloned into the NdeI and BamHI sites of the vector pGM10 by standard procedures (Martin & Keller, 1996).

The anti-symmetric (sym10) and anti-asymmetric arginine (asym24) antibodies were from Upstate and the anti-monomethyl arginine antibody (ab414) was from Abcam. Anti-pTyr antibody sc-7020 was from Santa Cruz, anti-CPSF-100 and -160 antibodies were described (Jenny & Keller, 1995). The anti-CF I_m antibody was described (Rüegsegger et al., 1998).

In vitro protein methylation assay. In a typical methylation assay, the reaction mixture contained between 400 and 1000 ng protein substrate, 1 μCi S-adenosyl-L-[methyl-^{3}H]-methionine (82 Ci/mmol; GE Healthcare) and methyl transferase preparations. The reaction was completed to 10 μl with MT reaction buffer (20 mM Tris-HCl pH 7.9, 40 mM KCl, 0.1 mg/ml BSA, 0.02% NP-40, 10% glycerol, 0.5 mM EDTA, 0.5 mM DTT) and incubated at 30°C for 2 hrs or over night. 10 μl of 2x SDS PAGE loading buffer was added and the products were separated on 10 or 12% SDS PAGE gels depending on the size of the substrate. The proteins were then blotted by a semi-dry blotting system onto nitrocellulose (Protran BA85, Whatman) and the filters were exposed to a PhosphorImager Tritium screen (GE Healthcare) for quantification or to Kodak BioMax MS film with a BioMax TRANSCREEN LE intensifying screen (from Kodak) usually over night.

ES(+/+) and ES(-/-) PRMT1 knockout cell extracts and methylation assays. Methylation assays with ES cell extracts were done as described (Ostareck-Lederer et al., 2006).

Purification of protein methyl transferase activity. Buffers: Buffer A contained 25 mM Tris-HCl pH 7.9, 10% glycerol, 0.2 mM EDTA, pH 8.0, 0.02% NP-40, 0.5 mM dithiotreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin hemisulfate, 0.7 μg/ml pepstatin. Buffer B was buffer A with 1 M KCl and buffer C was buffer A with 0.1 M KCl. All columns were run on a FPLC system (GE Healthcare) inside a cooling cabinet.

Hela cytoplasmic extract was prepared as described (Dignam et al., 1983), except that in all buffers, Hepes-KOH was replaced by Tris-HCl pH 7.9. Cytoplasmic extract corresponds to the S100 fraction of Dignam (Dignam et al., 1983). Typically, 20 ml of dialyzed cytoplasmic extract was spun to remove insoluble materials and applied to three 5 ml HiTrap-DEAE-FF columns connected in series and equilibrated in buffer C. The column was washed with 1.5 column volumes of buffer C und protein was eluted with a linear gradient of buffers A and B from 0.1 to 1 M KCl. Fractions were assayed with 500 ng of GST-68-GRP and ^{3}H-SAM as substrates. Active fractions were pooled, dialyzed against buffer A with Hepes, pH 8.0 and 50 mM KCl and loaded onto a 5 ml HiTrap SP column (GE Healthcare) and eluted with a 20-1000 mM gradient; activity was in the flow through. The flow through was directly loaded on a 1ml HiTrap heparin column (GE Healthcare) and eluted with a gradient of 50-1000 mM KCl (protease inhibitors were omitted from here on). Active fractions were concentrated in a Centricon-30 device and separated on a Superdex-200 gel filtration column (GE Healthcare) in buffer A with 150 mM KCl. Active fractions were subjected to ammonium sulfate precipitation (35%) by direct addition of solid ammonium sulfate. The precipitate was collected by centrifugation at 10'000 rpm for 30 min and the pellet was resuspended in buffer A with 500 mM ammonium sulfate and loaded on a 1 ml Phenyl-Superose column (GE

Healthcare). A decreasing gradient of 500 to 20 mM ammonium sulfate was applied. Active fractions were dialyzed against buffer A containing 20 mM KCl and loaded on a MonoQ column (GE Healthcare) and eluted with a 20-800 mM KCl gradient.

Yeast two-hybrid analysis of CF I_m59. The coding sequence for the full-length FLJ12529 protein (GenBank accession number gi: 22450805) was PCR-amplified and cloned into pB27 as a C-terminal fusion to LexA (N-LexA-FLJ12529-C) and sequenced. Yeast two-hybrid screening was performed by Hybrigenics, S.A., Paris, France (http://www.hybrigenics.com). The insert was used as a bait to screen a random-primed human placenta cDNA library constructed into pP6. pB27 and pP6 derive from the original pBTM116 (Vojtek & Hollenberg, 1995) and pGADGH (Bartel et al., 1993) plasmids, respectively.

68 million clones (6.8 -fold the complexity of the library) were screened with a mating approach with Y187 (matα) and L40ΔGal4 (mata) yeast strains as previously described (Fromont-Racine et al., 1997). 153 His^+ colonies were selected on a medium lacking tryptophan, leucine and histidine, and supplemented with 5 mM 3-aminotriazole to handle bait autoactivation. The prey fragments of the positive clones were amplified by PCR and sequenced at their 5' and 3' junctions. The resulting sequences were used to identify the corresponding interacting proteins in the GenBank database (NCBI) with a fully automated procedure.

GST pull-down experiments. Proteins were in vitro translated with the TNT rabbit reticulocyte lysate system (Promega) and labeled with ^{35}S-methionine (GE). 500 ng purified GST (control) or GST fusion-proteins were incubated with *in vitro* translated proteins in 50 µl total reaction volume in the presence of 0.5 mM PMSF, 0.4 µg/ml leupeptin, 0.7 µg/ml pepstatin, 50 µg/ml BSA (Biolabs) and optional 1U/µl DNase I and 10 ng/µl RNase A (see text or figure legends). Volumes were adjusted with PBS/0.01% NP-40 and reactions

incubated for 60 min at RT. Glutathione-Sepharose (20 μl of 50% suspension; GE Healthcare) was added and the volume was adjusted to 500 μl with PBS/0.01% NP-40. Incubation was continued for 90 min on a rotor arm and the beads were washed four times with PBS/0.01% NP-40. Proteins were eluted by the addition of 2×SDS-loading buffer, separated by 10 or 12% SDS-PAGE, and visualized by autoradiography or exposure to a Phosphorimager Tritium screen (GE Healthcare) for quantification.

Mass spectral analysis. Protein fractions enriched for methyltransferase activity were ethanol precipitated befor tryptic digest. For determination of methylarginine in the GAR motif protein GST-68_GR (4 μg) was methylated with 1 μg PRMT5/WD45/pICln methylosome, 0.1 mM SAM and a buffer consisting of 40 mM Tris-HCl, pH 7.9, 200 mM NaCl and 5 mM DTT over night at 30°C. Dimethylated products were immunoprecipitated with sym10 antibody coupled to protein A Sepharose beads, 25 μl non-reducing SDS loading buffer was added to washed beads, loaded on 10% SDS gel, stained with colloidal coomassie blue (Novex) and gel slices containing the expected protein were isolated and proteins digested in gel with trypsin. Peptides were subjected to electrospray mass spectral analysis.

RNAi. Double-stranded siRNA oligos PRMT5-1 (sense sequence: 5' GUA UGA GUG GGC UGU GAC A[dT][dT], PRMT5-2 (sense sequence: 5' CAC AGU ACU ACA UGG CUU U[dT][dT] and CF I_m68 (sense sequence: 5' GAC CGA GAU UAC AUG GAU A[dT][dT]) were from SIGMA. For reverse transfection according to the Lipofectamine RNAiMAX manual (Invitrogen), 10^4 HEK293 cells per well were seeded in 24-well plates with 20 nM siRNA and Lipofectamine-RNAiMAX (Invitrogen) in 0.6 ml OptiMEM medium (Invitrogen) and left to grow for 3 to 5 days at 37°C with 5% CO_2.

Steady state kinetics. Parameters for the sum of mono- and di-methylation were determined by titration of the substrate GST-68_GR in reactions with ^{3}H-SAM and either the

PRMT5/WD45 or PRMT5/WD45/pICln complex. After incubation reactions were precipitated with 10% TCA and counted in a scintillation counter.

Results of kinetics tests for dimethylation of GST-68_GR mutants by the PRMT5/WD45/pICln1 complex are presented as Western blot. Mutant proteins were titrated at 0.0, 0.1, 0.2, 0.6 and 1.8 μM (from left to right) concentrations together with ^{3}H-SAM and 1 μg of the recombinant complex for 2.5 hrs at 30°C and the reaction was separated on a 12% SDS gel and blotted onto a nitrocellulose membrane (ns is no substrate). The blot was probed with the sym10 antibody and further processed with a goat anti-rabbit HRP conjugate followed by ECL.

FIGURE LEGENDS

Figure 1. Graphic map of CF I_m59 and CF I_m68 proteins. (A) Domains are indicated by different patterns (code see under C). RRM indicates RNA recognition motif, Pro-rich is proline–rich domain, GAR is glycine-arginine motif (box of potential or cryptic GAR is white in CF I_m59 because no methylation was identified), RS/RD signifies RS-like domain and Phos is phosphotyrosine motif (see text). Also indicated as inserts are included and skipped exons in alternatively spliced forms of CF I_m. (B) Protein sequence of the CF I_m68 fragment in the GST-68_GR clone and with the exon included (in parentheses) in 68L1_GR. (C) Fragments of CF I_m59 and CF I_m68 used in the experiments. Protein segments are not drawn to scale.

Figure 2. PRMTs generate different types of arginine methylation. Arginine is modified to monomethylarginine by both type I and type II enzymes. Type I PRMTs then generate asymmetric dimethylarginine and type II enzymes catalyze symmetric dimethylarginine.

Figure 3. Methylation assay with cell extracts lacking PRMT1. Cell extracts from ES (+/+) and ES (-/-) cells were incubated with [^{14}C]SAM and recombinant GST-68_GRP, PABPN1 or hnRNP K (10 pmol each). Reactions were resolved on 12% SDS PAGE and methylation was detected by autoradiography.

Figure 4. Methylation assay with [H]$_6$-PABPN1, GST-CF I$_m$25 and GST-68_GR. Proteins (0.5 μg) were incubated with ^{3}H-SAM and either nuclear extract (NE), cytoplasmic extract (CE) or 0.5 μg of the PRMT5/pICln1/WD45 methylosome complex (Meto). ">" indicates expected size of GST-CF Im25.

Figure 5. CF I$_m$59 interaction with the SH3 domain in PRMT2 in two-hybrid screen and GST pulldown. (A) The minimal interaction domain (underlined) in the SH3 domain (in bold) of PRMT2 found in the two-hybrid screen. (B) Interaction between PRMT2 and the fragments of CF I$_m$59 and CF I$_m$68 demonstrated by GST pulldown. For details see text.

Figure 6. The PRMT5/WD45 complex is needed for methylation of the CF I$_m$68 GAR motif. The control substrate Sm proteins D1/D2 (1 μg) and GST-68_GR (800 ng) were incubated in standard methylation reactions with 1 μl nuclear (NE) or cytoplasmic extract (CE) or with increasing amounts of PRMT5 complexes as indicated. Amounts of enzyme present in the reactions were (from left to right) PRMT5/pICln: 10, 100 and 1000 ng; PRMT5/WD45 and PRMT5/WD45/pICln: 1, 10 and 100 ng.

Figure 7. Only arginines within the GGRGRGRF motif are methylated in CF I$_m$68. Site directed point mutations within the GAR motif in the GST-68_GR construct and fragments of CF I$_m$68 were subjected to methylation in vitro by recombinant pRMT5/WD45 complex. (A) 1 μg of recombinant substrate protein was incubated with 100 ng of PRMT5/WD45 and [^{3}H]SAM. (B) The same as A except that 1 μg of PRMT5/WD45 was used. (C) Same as B

except that the reaction contained cold SAM and the blot was probed with sym10 polyclonal antibody against SDM arginines. (D) Blot from C stained with Ponceau-S.

Figure 8. Kinetics of dimethyl transfer to GST-68_GR mutants in the GAR motif. Substrates were titrated as 0.0, 0.1, 0.2, 0.6 and 1.8 μM (from left to right) and reactions included ^{3}H-SAM and 1 μg PRMT5/WD45/pICln complex, ns is no substrate. Products were separated on SDS gels, blotted onto nitrocellulose membrane and probed with α-sym10 antibody. For details see Materials and Methods.

Figure 9. Mass spectral analysis of methylated GAR motif. Sum of spectra from LC/MS run of di-, tri-, and quadruply-methylated AAFPQGGRGRGRFPGAVPGGDR-peptide with triple charge.

Figure 10. CF I_m59 and CF I_m68 subunits are differentially dimethylated. (A) Nuclear (NE) or cytoplasmic (CE) extract, purified (CF I_m purif) recombinant *P. pastoris* expressed CF I_m (rec CF I_m) or partially purified CF II_m (CF II_m part) fractions separated on SDS gels and blotted to NC-filters were probed with α-CF I_m antibody (reacts with both CF I_m59 and CF I_m68 subunits) or α-monomethyl, sym10 or asym24 antibodies.

Figure 11. Summary of possible interactors of CF I_m59 (upper panel) and CF I_m68 (lower panel). Interaction with question marks are hypothetical; see text for details.

Supplementary Figures

Suppl. Figure S1. Alignment of the CF I_m59 and CF I_m68 protein sequences. Homologies are marked by light blue background except in the proline-rich domain, prolines are marked by yellow background and in the RS-like domain, arginines/lysines are indicated by red background, glutamines/aspartates by orange background and serines by green background. The RRM at the N-terminus is underlined in blue color and the GAR motif is boxed in red. Alternatively spliced exons are indicated by stippled boxes.

Suppl. Figure S2. (A) Western blot of purified CF I_m and Hela nuclear extract probed with α-CF I_m antibody. The antibody is reacting with both CF I_m59 and CF I_m68. (B) Western blot of GST pull-down with GST-CF I_m25 from nuclear extract. Both, longer and shorter forms of CF I_m59 and CF I_m68 are part of the complex and are precipitated together with CF I_m25.

Suppl. Figure S3. (A-C) Results of a first approach to determine methylation activity of nuclear extract upon recombinant cleavage and polyadenylation factors. Arrowheads indicate location of corresponding proteins on coomassie stained SDS gels (0.5 to 1 μg were loaded). NE or N is Hela nuclear extract and CE or C is cytoplasmic Hela extract. (D) Methylation assays with Superdex-200 fraction 7 (Sup-200, fr 7) of PRMT5/WD45 protein purified from Hela cytoplasmic extract. Addition of SV40 and poly(A) RNAs prevents methylation of recombinant PABPN1 (H6_PABPN1). Purified PABPN1 (PABPN1 purif) most likely contains methylated arginines and is not further methylated.

Suppl. Figure S4. Alignment of phosphotyrosine-containing motifs in CF I_m59 and CF I_m68 and homologous motif in Src kinase bound by Nck1/2.

Suppl. Figure S5. Knock down of PRMT5 and CF I_m68 by RNAi. Si-P5-1 (siRNA PRMT5-1), siP5-2 (siRNA PRMT5-2) and si-CF I_m68 (siRNA CF I_m68) correspond to RNAi minus or plus (-/+) these siRNAs in HEK293 cells (see Materials and Methods). Either the lysate or an

immunoprecipitate with the α-CF I_m antibody were probed in Western blots with the antibodies indicated at the left of the figure. Whereas the knockdown of the PRMT5 and CF I_m68 proteins was efficient, dimethylated CF I_m68 only decreased after CF I_m68 knockdown but dimethylation of CF I_m68 was not influenced by knockdown of PRMT5 (lane 6, with α -sym10).

Suppl. Figure S6. Symmetric and asymmetric dimethylation in active or inactive CPSF preparations after purification from Hela extracts. In the upper panel, in Western blots with inactive fractions of CPSF, CPSF-160 is asymmetrically (asym24) and symmetrically (sym10) dimethylated but not phosphorylated at tyrosines (α-P-Tyr). In the lower panel, active fractions of CPSF were analysed where no symmetric dimethylation was detected but tyrosine phosphorylation in both CPSF-100 and CPSF-160.

Acknowledgements

We thank Christiane Rammelt for help with RNAi and Suzette Moes and Paul Jenö for mass spectral analysis. We thank Isabelle Kaufmann, Henk de Vries and Elmar Wahle for protein fractions. W. K. was supported by the University of Basel and the Swiss National Science Fund (Grant No. 3100A0-102132/2).

References

Bartel P, Chien CT, Sternglanz R, Fields S. 1993. Elimination of false positives that arise in using the two-hybrid system. *Biotechniques 14*:920-924.

Bedford MT, Frankel A, Yaffe MB, Clarke S, Leder P, Richard S. 2000. Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J Biol Chem 275*:16030-16036.

Bedford MT, Richard S. 2005. Arginine methylation an emerging regulator of protein function. *Mol Cell 18*:263-272.

Berger I, Fitzgerald DJ, Richmond TJ. 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotechnol 22*:1583-1587.

Bienroth S, Wahle E, Suter-Crazzolara C, Keller W. 1991. Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J Biol Chem 266*:19768-19776.

Boisvert FM, Cote J, Boulanger MC, Richard S. 2003. A proteomic analysis of arginine-methylated protein complexes. *Mol Cell Proteomics 2*:1319-1330.

Branscombe TL, Frankel A, Lee JH, Cook JR, Yang Z, Pestka S, Clarke S. 2001. PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. *J Biol Chem 276*:32971-32976.

Brown KM, Gilmartin GM. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor I_m. *Mol Cell 12*:1467-1476.

Buday L, Wunderlich L, Tamas P. 2002. The Nck family of adapter proteins: regulators of actin cytoskeleton. *Cell Signal 14*:723-731.

Cardinale S, Cisterna B, Bonetti P, Aringhieri C, Biggiogera M, Barabino SM. 2007. Subnuclear localization and dynamics of the Pre-mRNA 3' end processing factor mammalian cleavage factor I 68-kDa subunit. *Mol Biol Cell 18*:1282-1292.

Chari A, Golas MM, Klingenhager M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H, Fischer U. 2008. An assembly chaperone collaborates with the SMN complex to generate spliceosomal snRNPs. *Cell 135*:497-509.

Chen Y, Zhou X, Liu N, Wang C, Zhang L, Mo W, Hu G. 2008. Arginine methylation of hnRNP K enhances p53 transcriptional activity. *FEBS Lett 582*:1761-1765.

Colgan DF, Murthy KG, Prives C, Manley JL. 1996. Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature 384*:282-285.

Coseno M, Martin G, Berger C, Gilmartin G, Keller W, Doublie S. 2008. Crystal structure of the 25 kDa subunit of human cleavage factor Im. *Nucleic Acids Res 36*:3474-3483.

de Vries H, Rüegsegger U, Hübner W, Friedlein A, Langen H, Keller W. 2000. Human pre-mRNA cleavage factor II_m contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J 19*:5895-5904.

Denman RB. 2008. Protein methyltransferase activities in commercial in vitro translation systems. *J Biochem 144*:223-233.

Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SM. 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor I_m mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem 279*:35788-35797.

Dichtl B, Keller W. 2001. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J 20*:3197-3209.

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res 11*:1475-1489.

Frankel A, Yadav N, Lee J, Branscombe TL, Clarke S, Bedford MT. 2002. The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity. *J Biol Chem 277*:3537-3543.

Frese S, Schubert WD, Findeis AC, Marquardt T, Roske YS, Stradal TE, Heinz DW. 2006. The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains. *J Biol Chem 281*:18236-18245.

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

Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet 16*:277-282.

Fronz K, Otto S, Kolbel K, Kuhn U, Friedrich H, Schierhorn A, Beck-Sickinger AG, Ostareck-Lederer A, Wahle E. 2008. Promiscuous modification of the nuclear poly(A)-binding protein by multiple protein-arginine methyltransferases does not affect the aggregation behavior. *J Biol Chem 283*:20408-20420.

Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA 6*:1197-1211.

Gunderson SI, Beyer K, Martin G, Keller W, Boelens WC, Mattaj LW. 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell 76*:531-541.

Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell 1*:255-264.

Herrmann F, Pably P, Eckerich C, Bedford MT, Fackelmayer FO. 2009. Human protein arginine methyltransferases in vivo - distinct properties of eight canonical members of the PRMT family. *J Cell Sci 122*:667-677.

Ingham RJ, Colwill K, Howard C et al. 2005. WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol 25*:7092-7106.

Jenny A, Keller W. 1995. Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and polyadenylation specificity factor. *Nucleic Acids Res 23*:2629-2635.

Kühn U, Nemeth A, Meyer S, Wahle E. 2003. The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem 278*:16916-16925.

Kyburz A, Friedlein A, Langen H, Keller W. 2006. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell 23*:195-205.

Latreille M, Larose L. 2006. Nck in a complex containing the catalytic subunit of protein phosphatase 1 regulates eukaryotic initiation factor 2alpha signaling and cell survival to endoplasmic reticulum stress. *J Biol Chem 281*:26633-26644.

Lee J, Sayegh J, Daniel J, Clarke S, Bedford MT. 2005a. PRMT8, a new membrane-bound tissue-specific member of the protein arginine methyltransferase family. *J Biol Chem 280*:32890-32896.

Lee JH, Cook JR, Yang ZH, Mirochnitchenko O, Gunderson SI, Felix AM, Herth N, Hoffmann R, Pestka S. 2005b. PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *J Biol Chem 280*:3656-3664.

Lee YH, Stallcup MR. 2009. Minireview: protein arginine methylation of nonhistone proteins in transcriptional regulation. *Mol Endocrinol 23*:425-433.

Lin WJ, Gary JD, Yang MC, Clarke S, Herschman HR. 1996. The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J Biol Chem 271*:15034-15044.

Martin G, Keller W. 1996. Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J 15*:2593-2603.

Martin G, Keller W. 2007. RNA-specific ribonucleotidyl transferases. *RNA 13*:1834-1849.

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

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

Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. 2006. An interaction between U2AF 65 and CF $I_{(m)}$ links the splicing and 3' end processing machineries. *EMBO J 25*:4854-4864.

Miranda TB, Miranda M, Frankel A, Clarke S. 2004. PRMT7 is a member of the protein arginine methyltransferase family with a distinct substrate specificity. *J Biol Chem 279*:22902-22907.

Neuenkirchen N, Chari A, Fischer U. 2008. Deciphering the assembly pathway of Sm-class U snRNPs. *FEBS Lett 582*:1997-2003.

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell 127*:635-648.

Ostareck-Lederer A, Ostareck DH, Rucknagel KP, Schierhorn A, Moritz B, Huttelmaier S, Flach N, Handoko L, Wahle E. 2006. Asymmetric arginine dimethylation of heterogeneous nuclear ribonucleoprotein K by protein-arginine methyltransferase 1 inhibits its interaction with c-Src. *J Biol Chem 281*:11115-11125.

Pahlich S, Zakaryan RP, Gehring H. 2006. Protein arginine methylation: Cellular functions and methods of analysis. *Biochim Biophys Acta 1764*:1890-1903.

Qi C, Chang J, Zhu Y, Yeldandi AV, Rao SM, Zhu YJ. 2002. Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor alpha. *J Biol Chem 277*:28624-28630.

Rual JF, Venkatesan K, Hao T et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature 437*:1173-1178.

Rüegsegger U, Beyer K, Keller W. 1996. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem 271*:6107-6113.

Rüegsegger U, Blank D, Keller W. 1998. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell 1*:243-253.

Ryan K, Bauer DL. 2008. Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation. *Int J Biochem Cell Biol 40*:2384-2396.

Ryan K, Khleborodova A, Pan J, Ryan XP. 2009. Small molecule activators of pre-mRNA 3' cleavage. *RNA 15*:483-492.

Schurter BT, Koh SS, Chen D, Bunick GJ, Harp JM, Hanson BL, Henschen-Edman A, Mackay DR, Stallcup MR, Aswad DW. 2001. Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry 40*:5747-5756.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell 33*:365-376.

Shimazu T, Horinouchi S, Yoshida M. 2007. Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3'-end processing. *J Biol Chem 282*:4470-4478.

Tang J, Gary JD, Clarke S, Herschman HR. 1998. PRMT3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *J Biol Chem 273*:16935-16945.

Vagner S, Vagner C, Mattaj IW. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev 14*:403-413.

Vethantham V, Rao N, Manley JL. 2008. Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. *Genes Dev 22*:499-511.

Vojtek AB, Hollenberg SM. 1995. Ras-Raf interaction: two-hybrid analysis. *Methods Enzymol 255*:331-342.

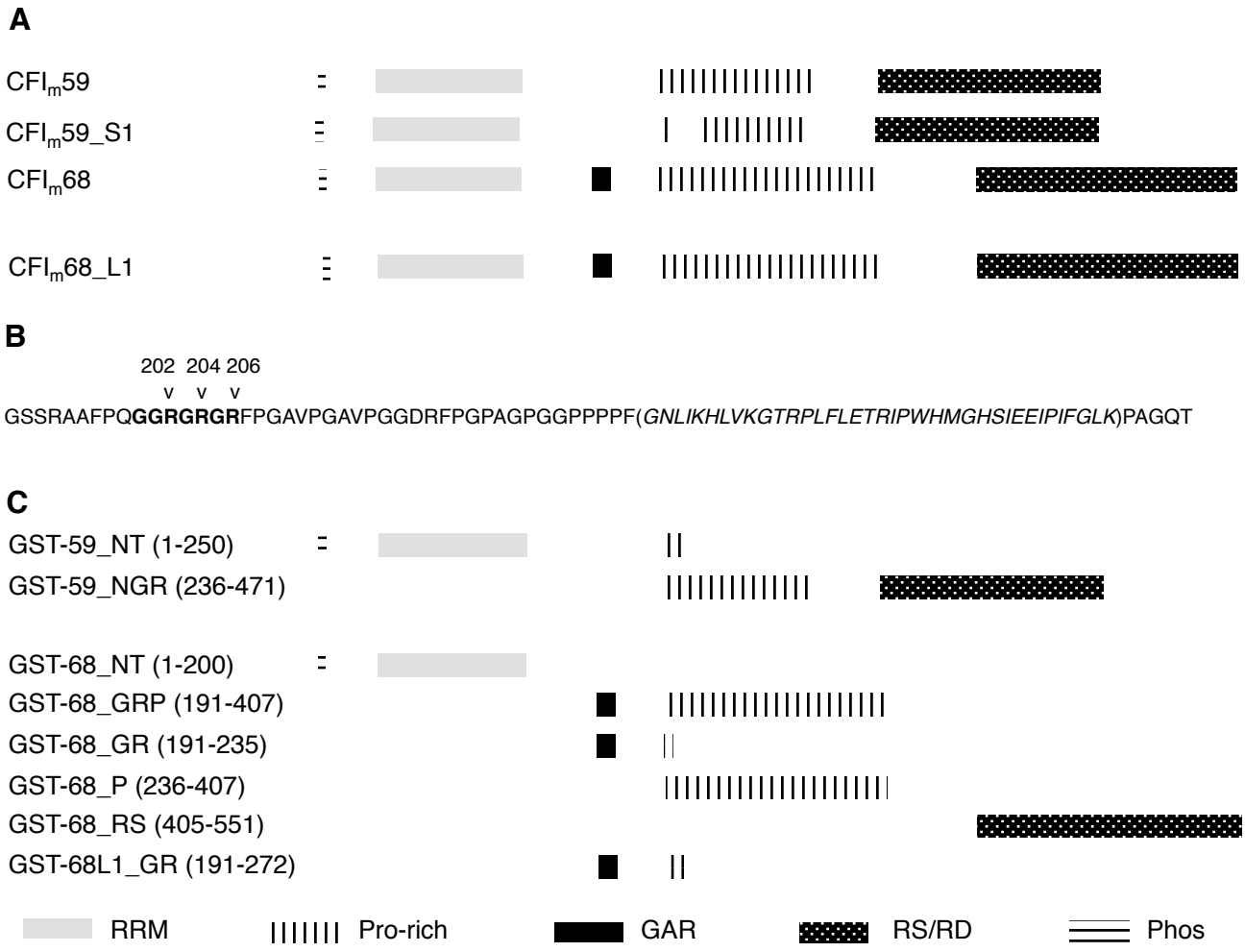
Wahle E. 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell 66*:759-768.

Wahle E, Keller W. 1996. The biochemistry of polyadenylation. *Trends Biochem Sci 21*:247-250.

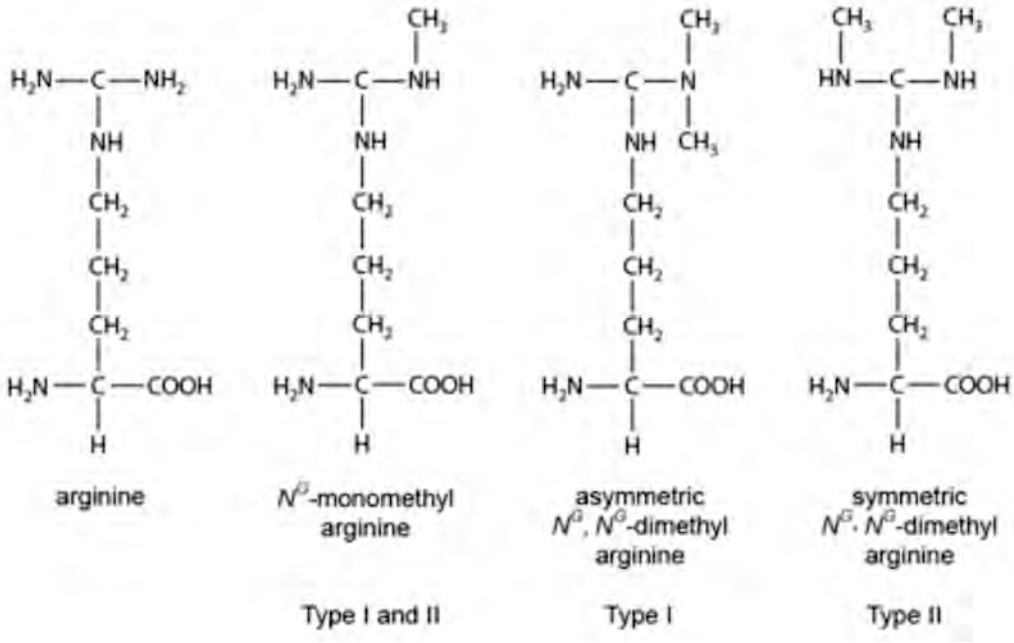
Zhang X, Cheng X. 2003. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure 11*:509-520.

Zhao J, Hyman L, Moore C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev 63*:405-445.

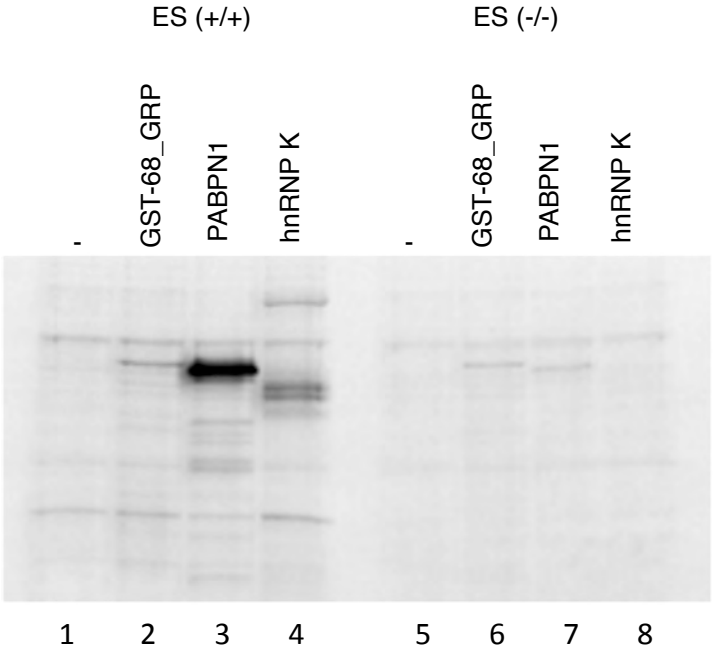
Martin et al., Figure 1



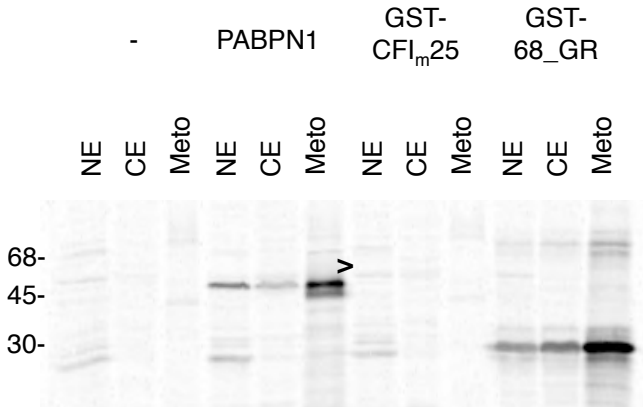
Martin et al., Figure 2



Martin et al., Figure 3



Martin et al., Figure 4

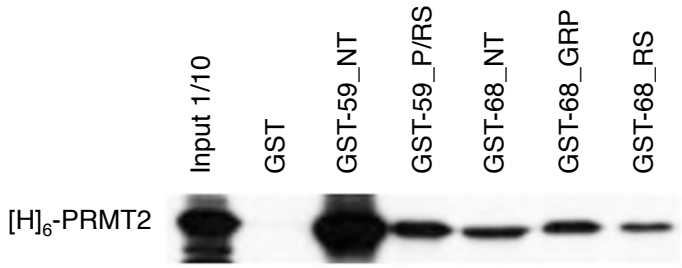


Martin et al., Figure 5

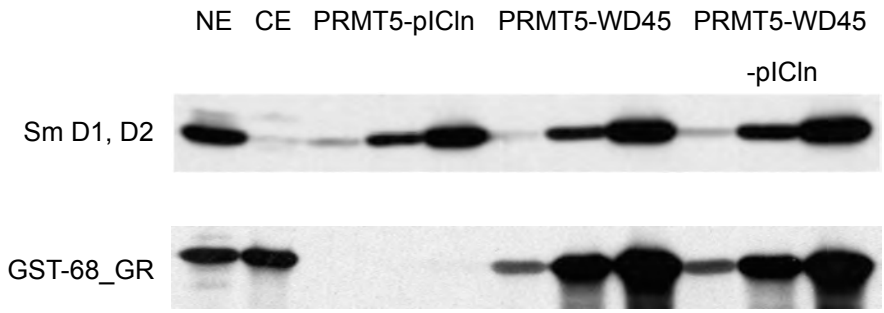
A

21 SEAGLLQEG**VQPEEFVAIADYAATDETQLSFLRGEKILIL**
61 **RQTTADWWWGERAGCCGYIPANHVGKHVD**EYDPEDT

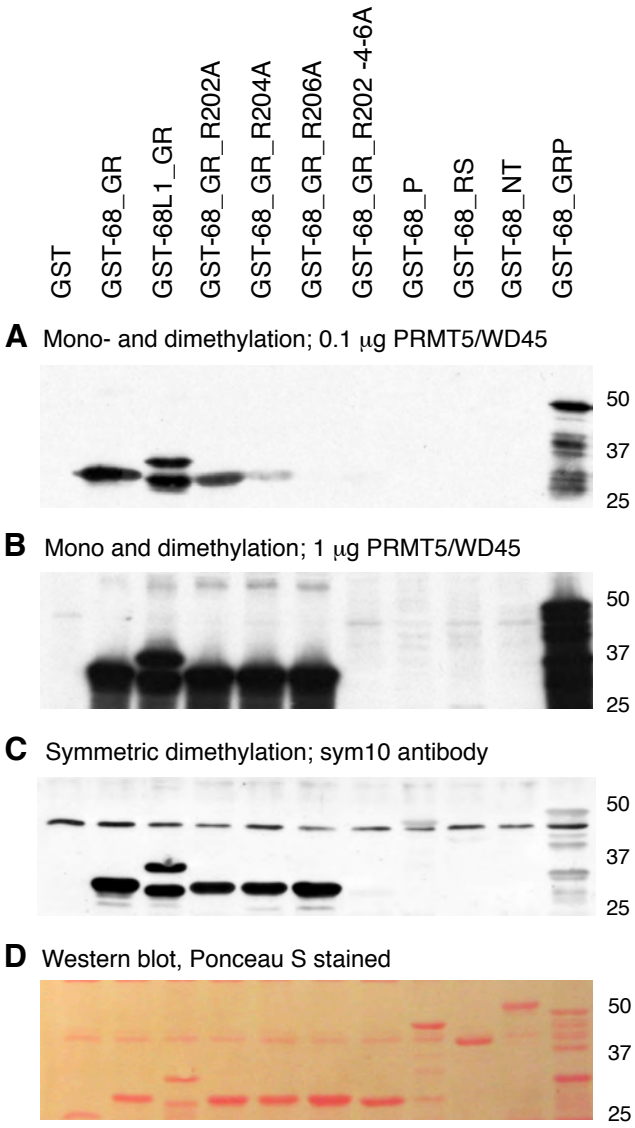
B



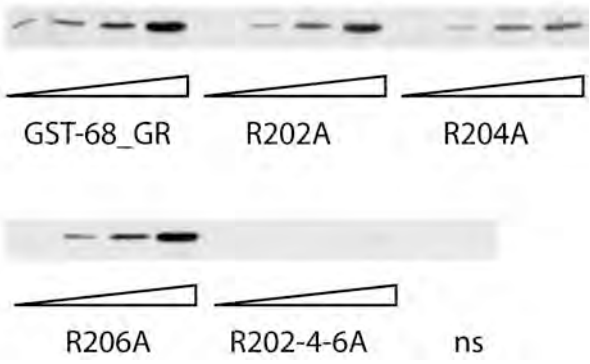
Martin et al., Figure 6



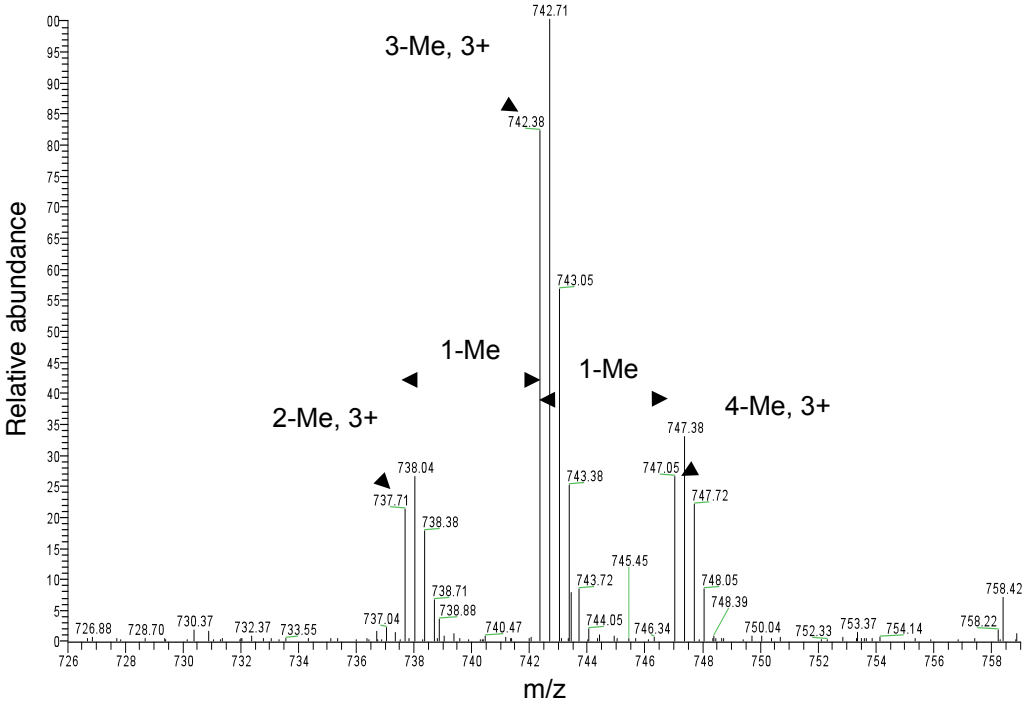
Martin et al., Figure 7



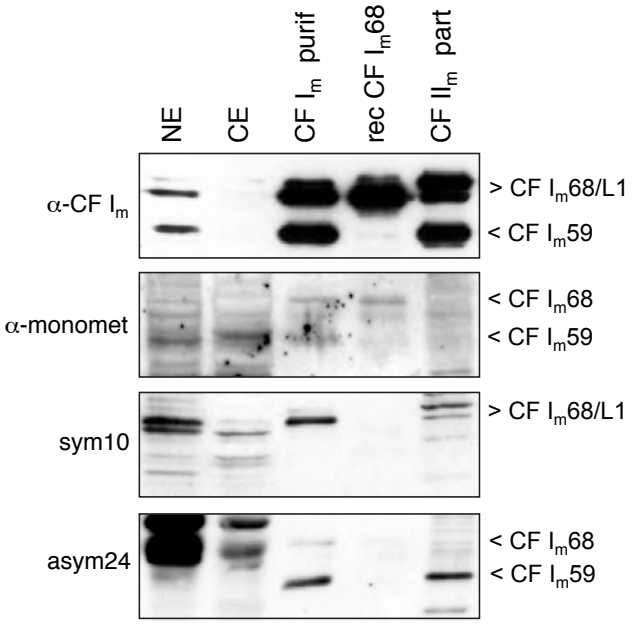
Martin et al., Figure 8



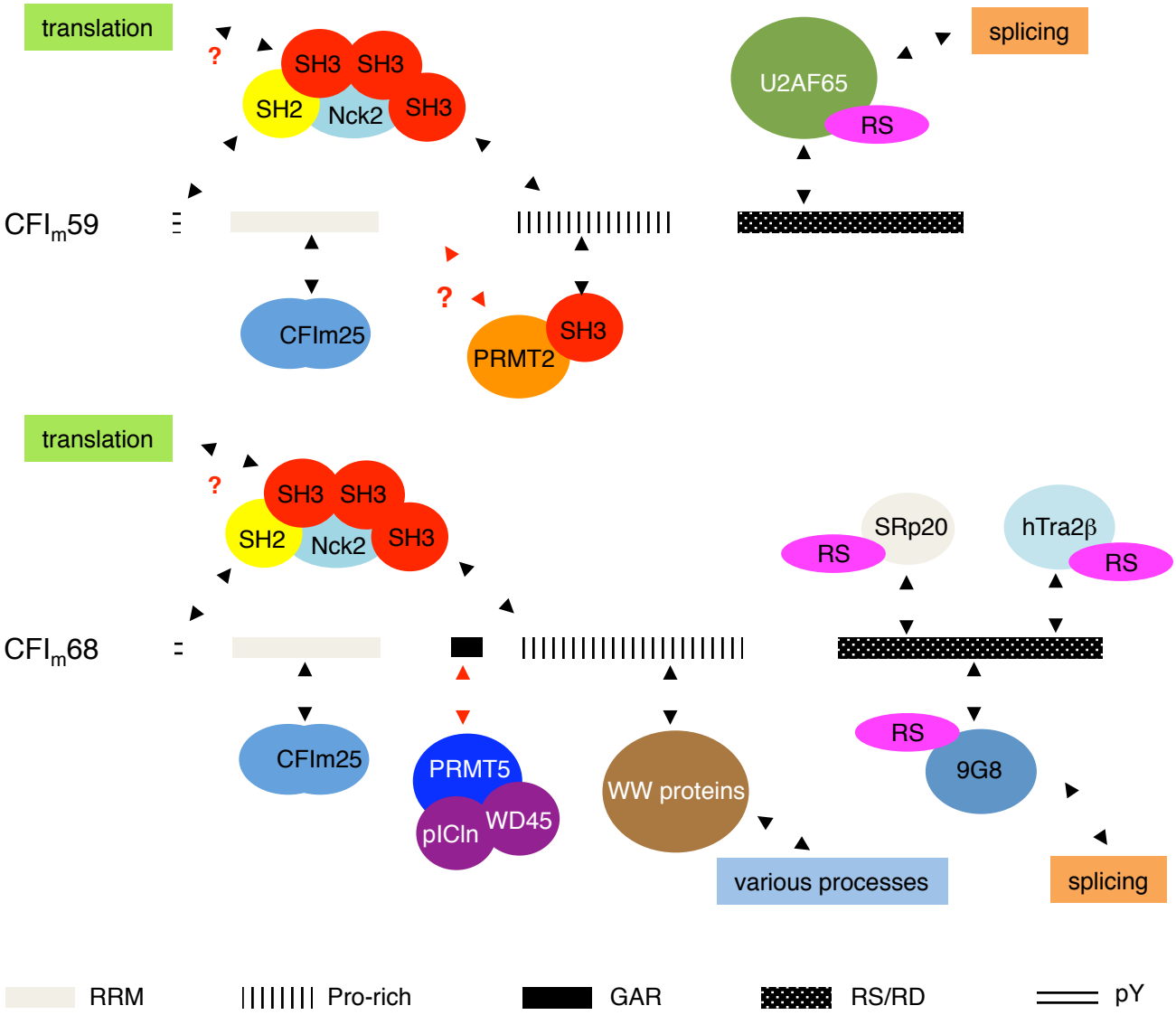
Martin et al., Figure 9



Martin et al., Figure 10

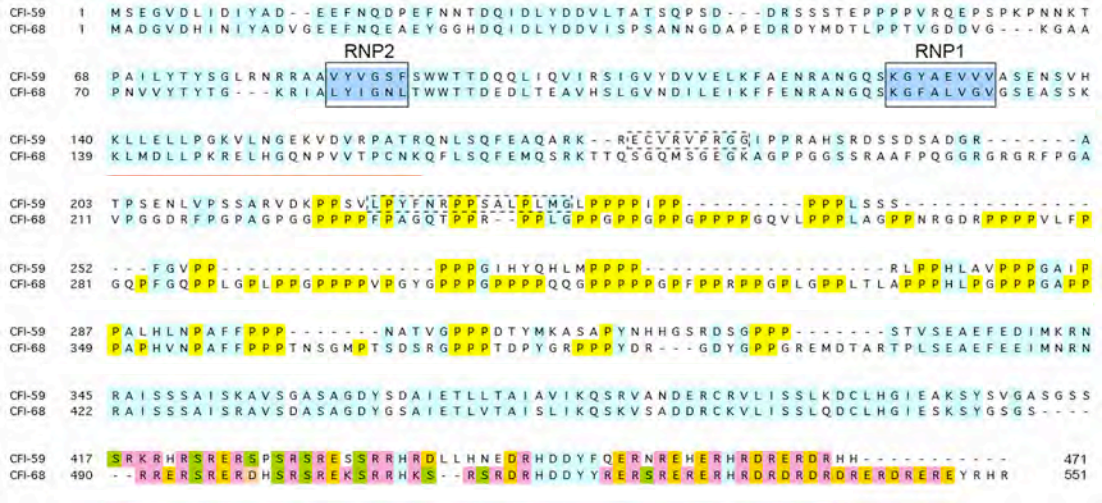


Martin et al., Figure 11

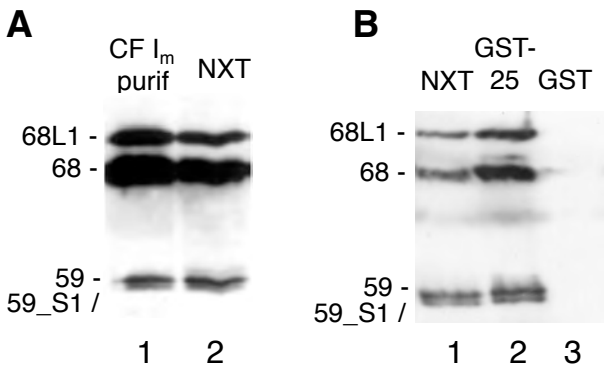


Supplementary Figures

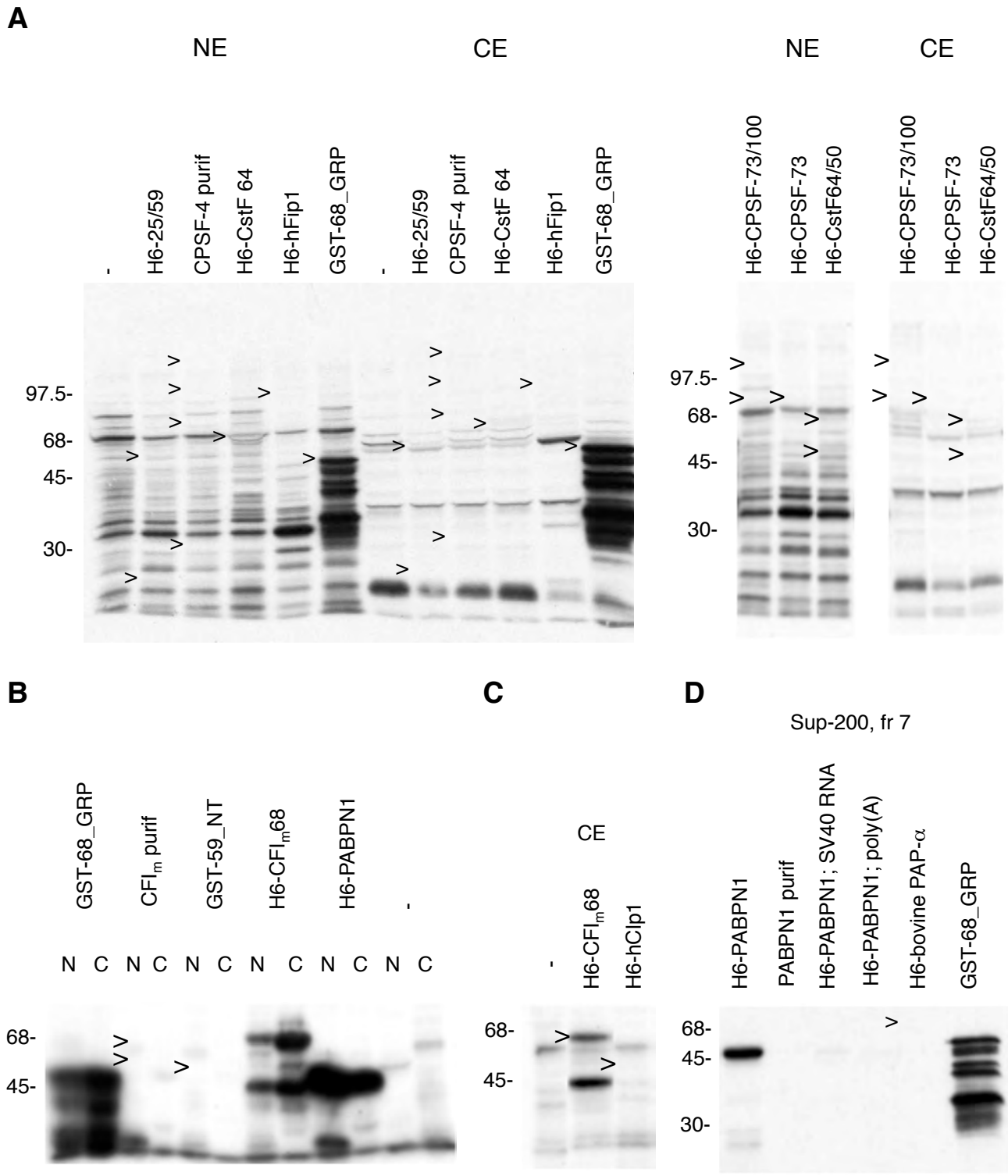
Martin et al; Suppl. Fig. S1



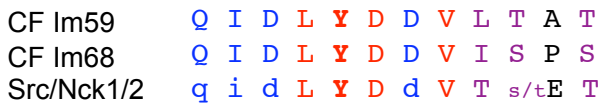
Martin et al; Suppl. Fig. S2



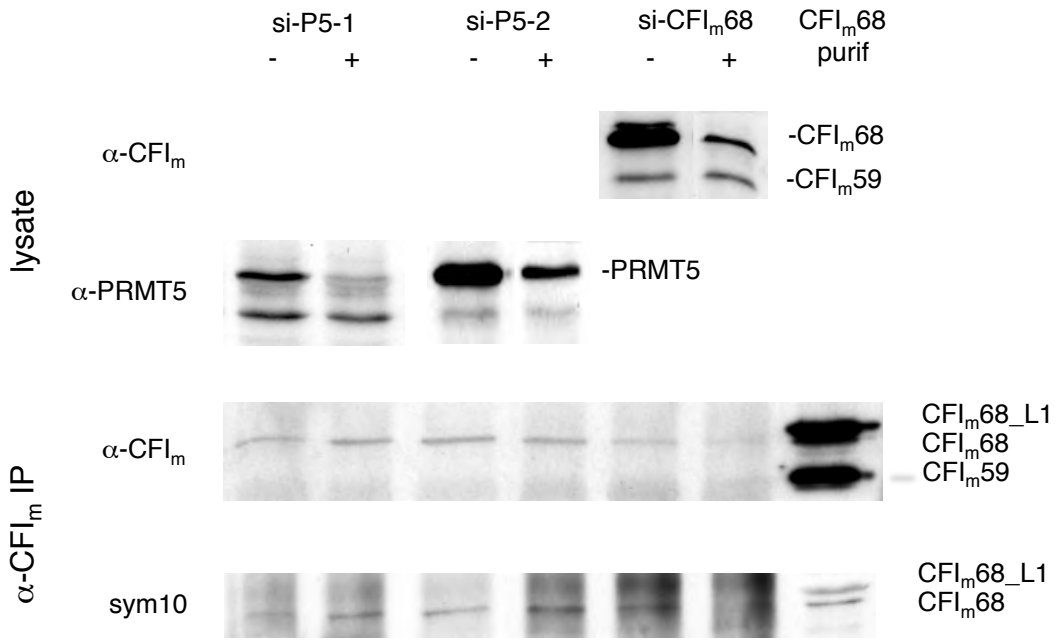
Martin et al; Suppl. Fig. S3



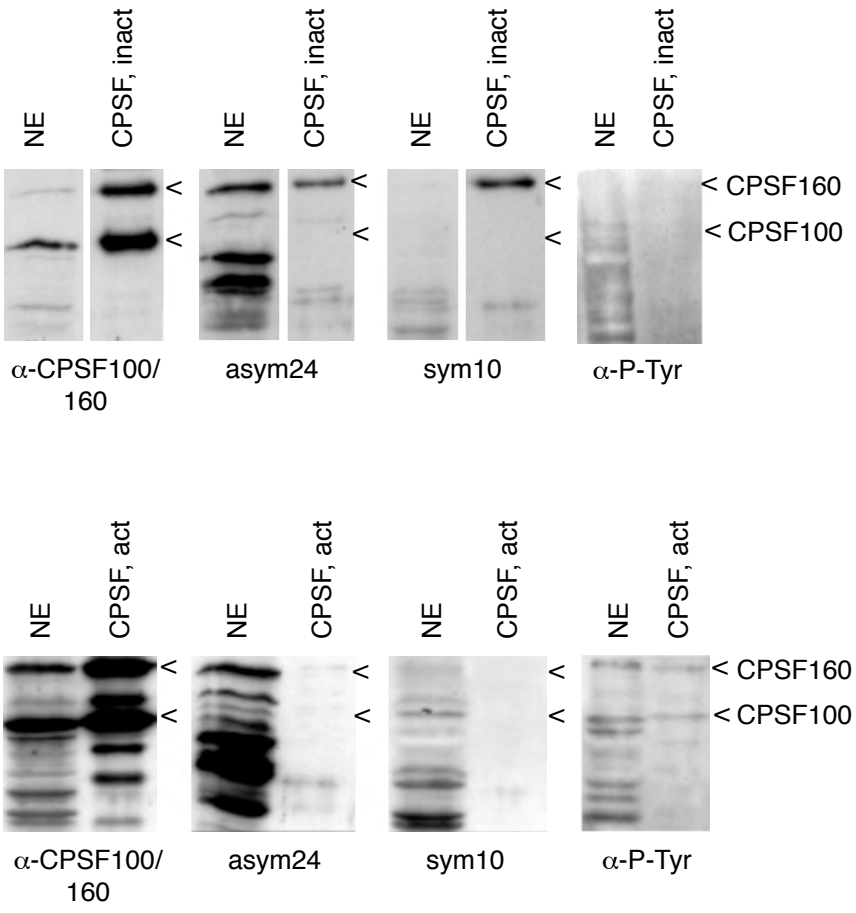
Martin et al; Suppl. Fig. S4



Martin et al; Suppl. Fig. S5



Martin et al; Suppl. Fig. S6



Martin et al; Suppl. Table 1

Supplementary Table 1. Factors tested for methylation by nuclear and cytoplasmic extracts, as depicted in Suppl. Figure 3

Factor/s	Labels in Figure S3	Comment	Source	Methylation
CF I_m25/59	A: H6-25/59	coexpressed	Insect/baculo	negative
CF I_m68	A,B,D: GST-68_GRP	GST tag	rec *E. coli*	**positive**
CF I_m25/59/68/68L	B: CFI_m purif	MonoQ fr. 40	Hela purified	negative
CF I_m59	B: GST-59_NT	GST tag	rec *E. coli*	negative
CF I_m68	B, C: H6-CFI_m68	His-6 tag	Insect/baculo	**positive**
hClp1	C: H6-hClp1	His-6 tag	rec *E. coli*	negative
CPSF	A: CPSF-4 purif	fr. 4; inactive	Hela purified	negative
CPSF-73	A: H6-CPSF-73	His-6 tag	Insect/baculo	negative
CPSF-73/100	A: H6-CPSF-73/100	CPSF73 has His-6 tag	Insect/baculo	negative
CstF-64	A: H6-CstF-64	His-6 tag	Insect/baculo	negative
CstF-50/64	A: H6-CstF64/50	CstF64 has His-6 tag	Insect/baculo	negative
hFip1	A: H6-hFip1	His-6 tag	Insect/baculo	negative
PABPN1	B, D: H6-PABPN1	His-6 tag	rec *E. coli*	**positive**
PABPN1	D: PABPN1 purif	From E. Wahle	Calf thymus purif.	negative
PAP-α	D: H6-bovine PAP-α	His-6 tag	rec *E. coli*	negative

Factors not tested: CstF-77, Pcf11, symplekin. "rec *E. coli*" is protein, expressed in *E. coli*.

8.3 IGHMBP2 is a Ribosome-Associated Helicase Inactive in the Neuromuscular Disorder Distal SMA Type 1

Ighmbp2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1

Günther UP, Handoko L, Laggerbauer B, Chari A, Sickmann A, Gehring N, von Au K, Schülke D, Fischer U: **Hum Mol Genet 2009**, 18: 1288-1300.

Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formlation of results:	20 %

Human Molecular Genetics, 2009, Vol. 18, No. 7 **1288–1300**
doi:10.1093/hmg/ddp028
Advance Access published on January 20, 2009

IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1)

Ulf-Peter Guenther[1,2,†], Lusy Handoko[3,†], Bernhard Laggerbauer[3], Sibylle Jablonka[4], Ashwin Chari[3], Mona Alzheimer[4], Jürgen Ohmer[3], Oliver Plöttner[3], Niels Gehring[6], Albert Sickmann[5], Katja von Au[1], Markus Schuelke[1,7] and Utz Fischer[3,*]

[1]Department of Neuropediatrics, Charité University Hospital, D-13353 Berlin, Germany, [2]Department of Biology, Chemistry and Pharmacy, Free University Berlin, D-14195 Berlin, Germany, [3]Theodor-Boveri-Institute and [4]Institute for Clinical Neurobiology, University of Würzburg, D-97074 Würzburg, Germany, [5]Rudolf-Virchow-Center for Experimental Biomedicine, University of Würzburg, D-97078 Würzburg, Germany, [6]University of Heidelberg, Im Neuheimer Feld 156, D-69120 Heidelberg, Germany and [7]NeuroCure Clinical Research Center, Charité University Hospital, D-10117 Berlin, Germany

Received November 22, 2008; Revised and Accepted January 13, 2009

Distal spinal muscular atrophy type 1 (DSMA1) is an autosomal recessive disease that is clinically characterized by distal limb weakness and respiratory distress. In this disease, the degeneration of α-motoneurons is caused by mutations in the immunoglobulin μ-binding protein 2 (IGHMBP2). This protein has been implicated in DNA replication, pre-mRNA splicing and transcription, but its precise function in all these processes has remained elusive. We have purified catalytically active recombinant IGHMBP2, which has enabled us to assess its enzymatic properties and to identify its cellular targets. Our data reveal that IGHMBP2 is an ATP-dependent 5′→3′ helicase, which unwinds RNA and DNA duplices *in vitro*. Importantly, this helicase localizes predominantly to the cytoplasm of neuronal and non-neuronal cells and associates with ribosomes. DSMA1-causing amino acid substitutions in IGHMBP2 do not affect ribosome binding yet severely impair ATPase and helicase activity. We propose that IGHMBP2 is functionally linked to translation, and that mutations in its helicase domain interfere with this function in DSMA1 patients.

INTRODUCTION

Mutations in the immunoglobulin μ-binding protein 2 *(IGHMBP2)*-gene cause distal spinal muscular atrophy type 1 (DSMA1; OMIM#604320), formerly known as spinal muscular atrophy with respiratory distress type 1 (SMARD1) (1). This autosomal recessive disorder manifests in early childhood as a result of progressive degeneration of α-motoneurons in the spinal cord. Characteristic clinical symptoms comprise respiratory distress due to diaphragmatic paralysis and muscle weakness that is more prominent in the distal limbs (2,3). Despite the nearly ubiquitous expression of the gene product IGHMBP2 (OMIM*600502), the α-motoneurons are predominantly affected (4,5). DSMA1, like the genetically distinct spinal muscular atrophy (SMA), thus appears to be caused by malfunction of a 'housekeeping' protein (6–8).

Based on sequence homology, IGHMBP2 has been classified as a member of the UPF1-like group within the helicase superfamily 1. The signature of this group is a DExxQ-type helicase/ATPase domain consisting of seven to nine conserved sequence motifs, including the Walker A and B motifs that are involved in ATP binding and hydrolysis (9,10). Consistent with these predictions, immunoprecipitates of IGHMBP2 were shown to unwind DNA duplices in an ATP-dependent fashion *in vitro* (11). This finding, together with the originally reported predominant localization of IGHMBP2 in the nucleus, has nourished the idea that IGHMBP2 contributes to the replication or transcription of DNA (12–15).

*To whom correspondence should be addressed. Tel: +49 931 888 4029; Fax: +49 931 4026; Email: utz.fischer@biozentrum.uni-wuerzburg.de
†These authors contributed equally to this work.

© *The Author 2009. Published by Oxford University Press. All rights reserved.*
For Permissions, please email: journals.permissions@oxfordjournals.org

This hypothesis still awaits proof by functional assays or, to the least, the identification of interacting partners. Based on co-localization of IGHMBP2 with splicing factors and an association with pre-mRNA authors have suggested a role in the processing of pre-mRNAs (11). However, this appears unlikely, given that IGHMBP2 has so far never been identified in proteomic profiles of spliceosomes, although their isolation has made major progress in the past years (16–19).

To reconcile the finding of IGHMBP2 immunoreactivity mainly in the cytosol and in the axons of neuronal cells (5), we hypothesized that its main substrate might not be nuclear DNA but rather cytoplasmic RNA. As helicases are thought to process nucleic acids irrespective of their sequences (20), we further sought cellular interaction partners that would confer substrate specificity to IGHMBP2 and/or modulate its enzymatic activity. The elucidation of the cellular processes to which IGHMBP2 contributes, might further highlight reasons for the predominant degeneration of α-motoneurons in DSMA1.

One of the obstacles in the characterization of IGHMBP2 so far may simply have been the difficulty to prepare recombinant protein of this helicase. We tackled this problem by establishing an efficient two-step purification strategy. This approach enabled us to characterize the enzymatic activity of IGHMBP2, to identify its associated macromolecules in its predominantly cytoplasmic cellular environment, and to analyze DSMA1-related mutants of the protein *in vitro* and *in vivo*. These IGHMBP2 mutants exhibit specific loss of enzymatic activity, which seems to be the primary biochemical defect underlying DSMA1. Our data point to a role of IGHMBP2 in translation and further suggest that DSMA1 is caused by alterations in the translational output of cells.

RESULTS

IGHMBP2 is an ATPase and ATP-dependent 5′→3′ helicase

As a first step towards a detailed functional characterization of IGHMBP2, we devised a reproducible and fast procedure to isolate full-length human IGHMBP2 protein expressed in *Escherichia coli*. For this, we constructed an expression plasmid containing the coding sequence of human IGHMBP2 preceded by an in-frame N-terminal GST-tag and flanked by a C-terminal hexahistidine tag (6×His-tag). Taking advantage of both tags, we purified IGHMBP2 to apparent homogeneity using a two-step purification protocol (Supplementary Material, Fig. S1). We also constructed plasmids for proteins with mutated Walker A and B motifs predicted to abolish ATPase activity of IGHMBP2 protein variants (Fig. 1A). These variants were purified in parallel to wild-type protein and utilized as controls in our assays and additionally to rule out the presence of potential contaminating NTPases and helicases.

Having established a purification procedure for IGHMBP2, we first tested whether this protein was able to hydrolyze ATP. Hydrolysis activity of IGHMBP2 was determined in the absence and presence of nucleic acids as it had been proposed that binding of ATP and of substrate nucleic acids is cooperatively coupled (21). The recombinant protein was incubated with α[^{32}P]-ATP and products of hydrolysis were analyzed after 1 h by thin layer chromatography. As shown in Figure 1B (lanes 2, 3, 4 and 6) recombinant IGHMBP2 catalyzed ATP hydrolysis in the presence of either DNA or homopolymeric RNAs such as poly(A), poly(C) and poly(U), which became evident by the appearance of labeled ADP and the simultaneous disappearance of ATP. Only poly(G) (Fig. 1B, lane 5) was less efficient in this assay. The absence of nucleic acids (Fig. 1B, lane 1) as well as mutations in the Walker A or B motifs (Fig. 1B, lanes 7 and 8) reduced ATPase activity to background levels (Fig. 1B, lanes 9–14), the latter verifying that our protein preparations were free of contaminating ATPases. Thus, recombinant IGHMBP2 exhibited catalytic activity and hence was used for further functional testing. These data are further corroborated by the capability of IGHMBP2 to form protein-DNA and protein-RNA complexes (Supplementary Material, Fig. S2).

In the next step, we determined the specificity of nucleotide binding to IGHMBP2. For this, we added a 50 molar excess of various unlabeled nucleotides as competitors to the ATPase assay. Under these conditions, only ATP or dATP efficiently inhibited hydrolysis of α[^{32}P]-labeled ATP, while the other nucleotides had no measurable effect (Fig. 1C). We conclude that the nucleotide cofactor requirement of IGHMBP2 is identical to other known members of the UPF1-like group of helicases (21–23).

Given our observations that ATPase activity of IGHMBP2 is strongly stimulated by the presence of nucleic acids, we assessed if ATP hydrolysis would provide the free energy for unwinding the partially double stranded region of RNA and/or DNA duplices. For strand displacement assays, an RNA-duplex was generated from a radioactively labeled artificial RNA of ∼200 nucleotides, annealed to a smaller partially complementary un-labeled RNA. This duplex contained single strand overhangs at both the 3′- and 5′-ends (Fig. 1D). The RNA-hybrid was then incubated with increasing amounts of IGHMBP2 and analyzed by native gel electrophoresis. As shown in Figure 1D (lanes 3–7), IGHMBP2 unwound the duplex in a dose-dependent manner, as evident by the appearance of the labeled single-stranded RNA at the expense of the RNA duplex substrate. Strikingly, the omission of ATP or the introduction of a mutation in the Walker B motif entirely abolished the helicase activity of IGHMBP2 (Fig. 1D, lanes 8 and 9), illustrating the specificity of the reaction.

The enzymatic activity of IGHMBP2 was further characterized with duplices that contained single-stranded overhangs at either 5′ or 3′ termini only. As shown in Figure 1E (lanes 3 and 7), only the duplex with a 5′ single-stranded overhang was unwound, whereas the substrate containing a 3′ single-stranded overhang remained stable in the presence of IGHMBP2. Similar results were obtained with DNA-duplices (Supplementary Material, Fig. S3). Taken together, our data identify IGHMBP2 as an enzyme that unwinds RNA and DNA duplices with 5′ overhangs in an ATP-dependent reaction, which classifies IGHMBP2 as a 5′→3′ helicase.

IGHMBP2 is part of large RNP complexes in cellular extracts

In order to elucidate the cellular context of IGHMBP2 function, we re-investigated its subcellular localization by confocal laser microscopy using different monospecific α-IGHMBP2 antibodies.

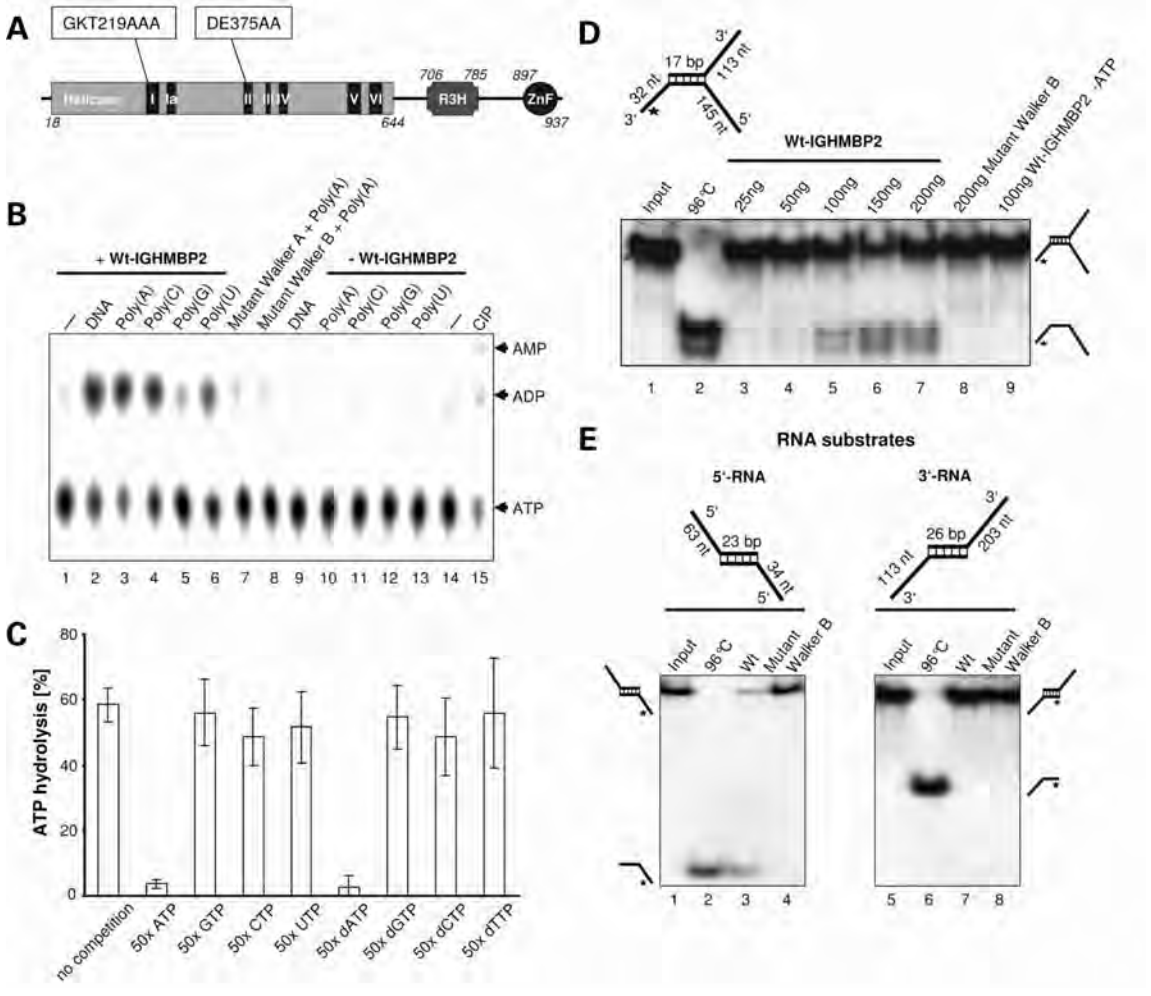


Figure 1. Recombinant IGHMBP2 is an active 5′→3′ RNA helicase. (**A**) Primary structure of IGHMBP2 showing the motifs of the helicase superfamily 1 domain, a nucleic acid binding domain of the R3H type, as well as an AN1-type zinc finger motif. The positions of each IGHMBP2 domain are numbered according to Interpro database (http://www.ebi.ac.uk/interpro) and are given in italics. The localization of the control mutations in the Walker A (GKT219AAA) and Walker B (DE375AA) motif of the helicase domain are indicated. (**B**) IGHMBP2 is a nucleic acid-dependent ATPase. Purified IGHMBP2 protein was assayed in the absence or presence of dsDNA and various RNA homopolymers. A control reaction with Calf Intestine Phosphatase (CIP) was performed to identify the positions of hydrolysis products. (**C**) Wild-type (Wt)-IGHMBP2 hydrolyzes selectively ATP and dATP. Hydrolysis products were quantified by densitometry in three independent experiments. (**D**) ATP-dependent RNA unwinding activity by increasing amounts of wild-type and mutant IGHMBP2 protein. (**E**) IGHMBP2 is a 5′→3′ helicase. The directionality of IGHMBP2 helicase activity was assayed with RNA duplices containing either a 5′ or a 3′ overhang. The position of the radio-label in the RNA is indicated at the right side with a star.

For the sake of clarity, we describe here only the results obtained with an α-IGHMBP2 antibody raised against the C-terminal part of the protein but very similar results were also obtained with other antisera (see also Supplementary Material, Fig. S6 for characterization of these antibodies).

We initially analyzed primary embryonic mouse motoneurons as this cell type is primarily affected in DSMA1 (Fig. 2A and B). Contrary to previous studies that have suggested functions of IGHMBP2 on DNA and pre-mRNA in the nucleus (11–15), we observed an almost exclusive localization within the cytoplasm of motoneurons. Ighmbp2 was detected in the perinuclear cytoplasm as well as in the axons and growth cones (Fig. 2; see also 5) and its expression was significantly reduced in all three compartments of motoneurons obtained from *nmd (*neuromuscular degeneration) mice, an animal model of DSMA1 (4; see also Supplementary Material, Fig. S7). The granular pattern of cells stained with different α-IGHMBP2 antibodies was reminiscent of the distribution of components from the translation machinery, and indeed, we also observed co-localization of Ighmbp2 with the eukaryotic translation initiation factor eIF4G2 (DAP-5 antibody) and with ribosomal RNA (Y10B antibody) (Fig. 2). These findings were corrobated by localization studies in different cell-types, or by transfection of mouse N1E115 neuroblastoma cells with FLAG-tagged variants of IGHMBP2 (Fig. 2C and Supplementary Material, Fig. S8).

We next determined whether IGHMBP2 forms stable complexes with other cellular components. For this we prepared a cytoplasmic extract from a murine fibroblast cell line (FM3A) under physiological conditions. The extract was analyzed by size fractionation in a linear 5–30% (v/v) glycerol gradient (Fig. 3). The presence of IGHMBP2 in individual fractions was determined by Western blotting with α-IGHMBP2 antibodies. As shown in the upper panel of Figure 3, IGHMBP2 sedimented

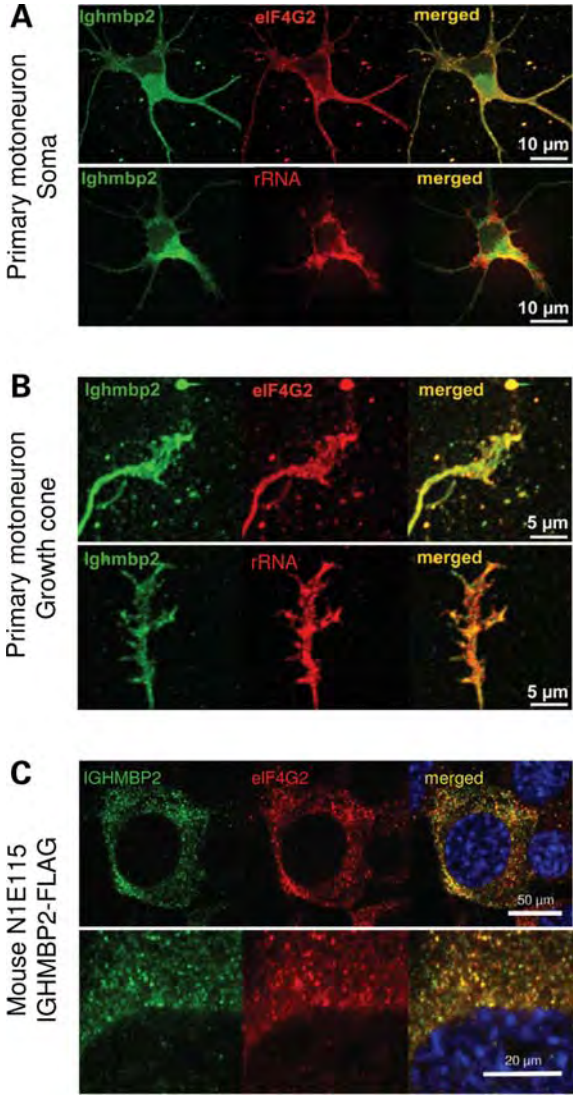


Figure 2. IGHMBP2 is a cytoplasmic protein and part of large ribonucleoprotein complexes. (**A**) Ighmbp2 co-localizes with eIF4G2 (DAP-5) and ribosomal RNA (Y10B) at the cell membrane in the soma compartment of isolated motoneurons from control embryos (Ighmbp2$^{+/+}$). Ighmbp2 is evenly distributed throughout the cytoplasm of the cell body, whereas eIF4G2 is more restricted to the cell membrane. The rabbit antibody used to detect Ighmbp2 was raised against the C-terminal fragment of Ighmbp2 (**B**). In the growth cone, co-localization of Ighmbp2 with eIF4G2 and rRNA spans the whole compartment. (**C**) Upper panel: N1E115 mouse neuroblastoma cells were transfected with a full-length wild-type IGHMBP2 construct containing a C-terminal FLAG-tag. The overlay shows co-localization of immunoreactivity against the ribosome (α-eIF4G2, DAP-5 antibody) and against the IGHMBP2-FLAG fusion protein (α-FLAG antibody) to a large extent. DNA is stained by DAPI (blue). Lower panel: Higher resolution confocal micrograph of the same cells highlighting the border area between nucleus and cytosol demonstrating predominant co-localization of IGHMBP2 and eIF4G2 in the cytosol.

in two major peaks with S-values above 30S (indicated by bars), whereas recombinant IGHMBP2 alone was found on top of the gradient (data not shown, see also Supplementary Material, Fig. S9). Increasing the salt concentration or RNase A treatment of the extract resulted in the quantitative dissociation of IGHMBP2 from these complexes, whereas RNase-free DNase had only marginal effects (Fig. 3). Similar results were obtained

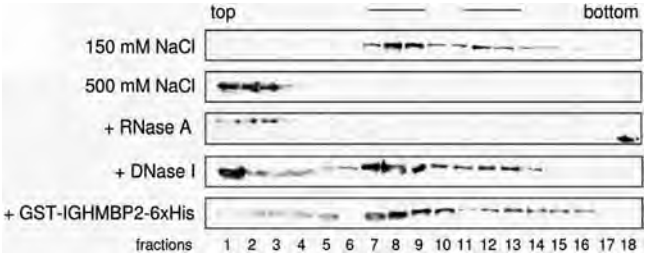


Figure 3. Endogenous Ighmbp2 is present in large ribonucleoprotein complexes in cellular extracts. Mouse FM3A cytoplasmic extract was fractionated by glycerol gradient centrifugation and analyzed by western blotting with α-IGHMBP2 antibody raised against the full-length protein. Ighmbp2 was found in two major peaks, which disappeared under high salt conditions (500 mM NaCl) or when the extract was treated with RNase A, but was resistant to DNase I. The moderate shift of the protein to the top of the gradient after DNase treatment was only occasionally observed and was most likely due to the dissociation of Ighmbp2 during the incubation period of the extract. The bottom panel shows integration of recombinant Ighmbp2 into similar complexes as the endogenous protein.

with extracts derived from other cell lines (mouse and rat) or from mouse brain tissue (data not shown). Taken together, these data are consistent with the idea that IGHMBP2 is part of large cytosolic ribonucleoprotein complexes (RNPs) *in vivo*.

IGHMBP2 is a ribosome-associated helicase

For the identification of components within the IGHMBP2 ribonucleoprotein complex, we devised a purification strategy from cellular extracts. This strategy was based on the finding that our recombinant, catalytically active IGHMBP2 forms complexes upon addition to cellular extract that are very similar in size and stability as the endogenous protein (Fig. 3, lower panel). Hence, we reasoned that the recombinant IGHMBP2 protein engages in similar interactions and would enable the isolation of interacting components by biochemical means. For purification of interactors, we immobilized recombinant IGHMBP2 on glutathione-Sepharose and incubated this affinity matrix with FM3A cytoplasmic extracts. Following elution of IGHMBP2 with glutathione, co-eluting proteins were identified by gel electrophoresis and mass spectrometry (Fig. 4A). As a control, GST alone was immobilized and treated identically. Strikingly, most proteins specifically eluting together with IGHMBP2 either belong to the large or small ribosomal subunit or were components of the translational apparatus (Table 1). Likewise, 5S, 18S and 28S rRNAs were abundantly present in the IGHMBP2 eluate (Fig. 4A, right panel). Notably, the degradation of ribosomal RNA by treatment with RNase A drastically reduced the amount of ribosomal proteins that co-eluted with IGHMBP2 (Fig. 4B). These results suggest that ribosomal subunits, rather than individual components (proteins or RNA) had bound to the IGHMBP2 column.

To confirm this, we subjected the eluate to sedimentation analyses on glycerol gradients. Strikingly, two distinct complexes were separated on these gradients corresponding to the small (40S) and large (60S) ribosomal subunits, as evidenced by the identification of the corresponding ribosomal proteins and rRNAs, in the same fractions (Fig. 4C, marked with bars). Furthermore, under conditions that stabilized 80S ribosomes, IGHMBP2 and its interactors sedimented only in

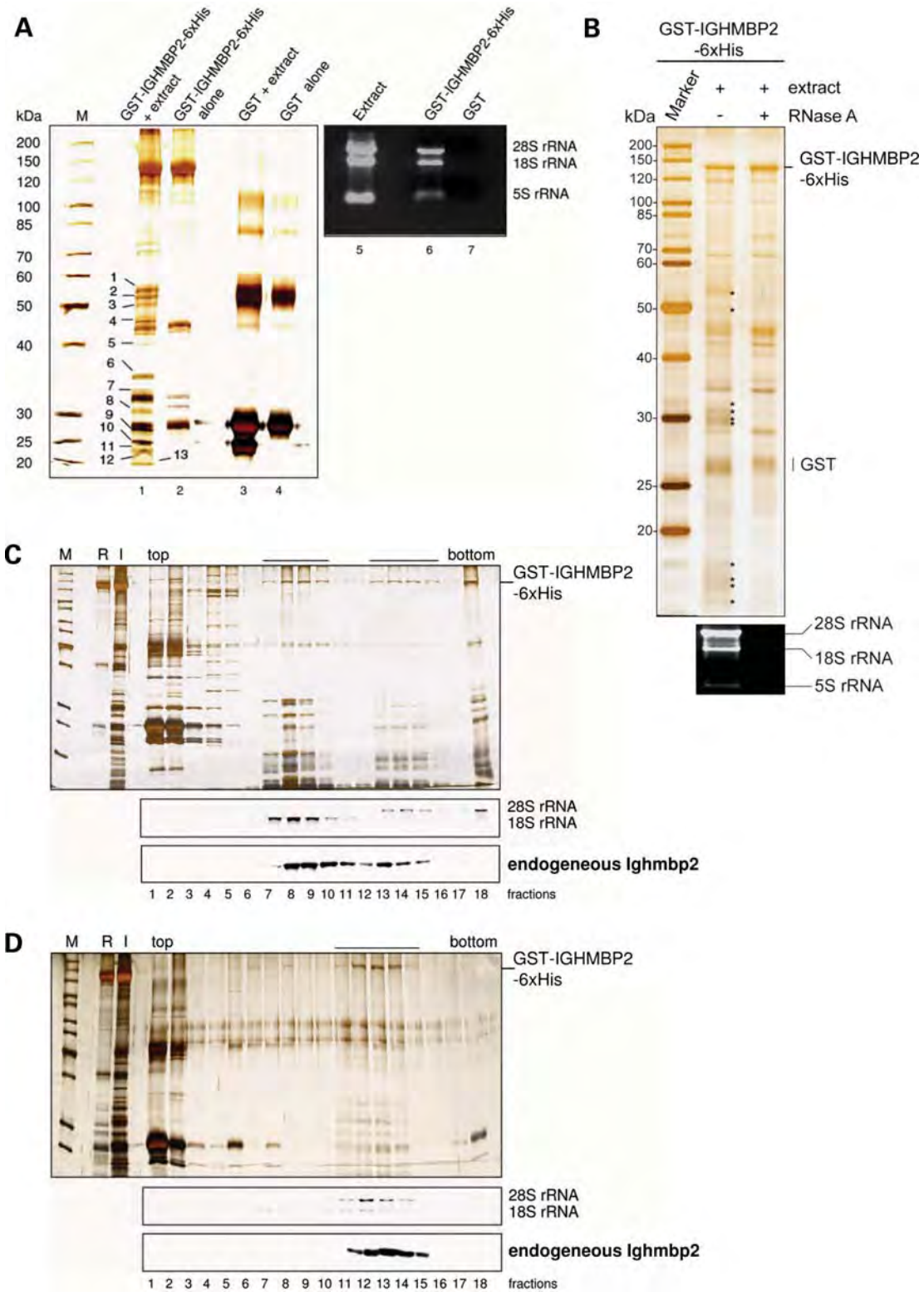


Figure 4. IGHMBP2 associates with ribosomes *in vitro*. (**A**) Isolation of IGHMBP2 interactors using immobilized recombinant IGHMBP2 as an affinity matrix. IGHMBP2-interactors were identified as ribosomal proteins, translation factors and other translation-related proteins (see also Table 1). RNA analysis of the eluate revealed the presence of 5S, 18S and 28S ribosomal RNAs in the IGHMBP2-complex (right panel). (**B**) Treatment of cytosolic extract with RNase A prevents association of IGHMBP2 with ribosomes. Recombinant IGHMBP2 immobilized on Glutathione–Sepharose beads was incubated with cytosolic extract, which was either mock-treated or incubated with RNase A. Bound proteins were analyzed by SDS–PAGE and visualized by silver staining. The lower gel shows ribosomal RNA in the mock-treated and RNase A treated extract. Ribosomal proteins specifically bound to IGHMBP2 are marked by asterisks. (**C**) IGHMBP2 associates with 40S and 60S ribosomal subunits. Upper panel: The IGHMBP2 complex was analyzed by a glycerol gradient centrifugation and the

a single peak (Fig. 4D). Of note, the sedimentation profiles of the purified, eluted complexes corresponded to those of endogenous Ighmbp2 in cell extracts under the same experimental conditions (Fig. 4C and D, lower panels).

The experiments described earlier suggest an involvement of Ighmbp2 in translation. We therefore analyzed the sedimentation of the endogenous protein under conditions that stall translation at specific stages. Ighmbp2 sedimented at 80S under conditions that stabilized monosomes, but split into 40S and 60S complexes when the subunits were dissociated by the addition of EDTA (Fig. 5A and B). Binding of Ighmbp2 to actively translating ribosomes was investigated in extracts of cycloheximide-treated cells. The ability of cycloheximide to lock polysomes in their translation process was monitored by the detection of fragile X mental retardation protein (FMRP), a protein known to preferentially associate with polysomes (24). In agreement with these findings, FMRP was found predominantly in polysomal fractions. Likewise, some Ighmbp2 co-sedimented with polysomes but the majority was present as 80S monosomes (Fig. 5C). These data indicate that Ighmbp2 either transiently associates with polysomes or that the transition to elongation causes its dissociation from the ribosome.

Pathogenic IGHMBP2 variants remain ribosome associated but display loss of enzymatic activity

The identification of Ighmbp2 as a ribosome-associated RNA-helicase allowed us next to investigate the influence of DSMA1 missense mutations on its activity and ribosome binding. All disease-causing missense-mutations in *IGHMBP2* known thus far (1–3) are located within or adjacent to the helicase domain and target highly conserved amino acid residues, suggesting the unwinding activity of IGHMBP2 to be affected in DSMA1. Out of 28 known missense mutations, most of which are localized within or directly adjacent to motifs described to be essential for distinct steps of catalysis by helicases, we subjected nine representative mutant IGHMBP2 proteins to functional analysis (Fig. 6A). The lower panel of Figure 6A shows that mutant proteins were obtained in comparable quantity and purity when compared with wild-type protein. Initially, these mutants were tested for their ATPase activity. Seven out of nine mutants (p.Q196R, p.T221A, p.C241R, p.E382K, p.H445P, p.N583I and p.R603H) exhibited only background level activity when compared with wild-type IGHMBP2 (Fig. 6B). As ATP hydrolysis and binding of nucleic acids are cooperatively coupled, we determined the capacity of the mutant proteins to bind to either single-stranded RNA or DNA in electrophoretic mobility shift assays (EMSA; Supplementary Material, Fig. S4). DSMA1-related mutant proteins displayed variable reduction of nucleic acid binding consistent with the reported contribution of the respective conserved helicase motifs to substrate nucleic acid binding (reviewed in Hall and Matson). Furthermore, all ATPase deficient mutants were also defective in RNA and DNA unwinding, consistent with the fact that ATP hydrolysis is a prerequisite for helicase activity of the enzyme (Fig. 1B and D; Supplementary Material, Fig. S5). However, two pathogenic IGHMBP2 variants (p.T493I and p.D565N) showed ATPase activities indistinguishable from wild-type protein (Fig. 6B). Nonetheless, the p.D565N mutant was completely inactive in the helicase assay (Fig. 6C), but still retained its full nucleic acid binding capacity (Supplementary Material, Fig. S4). This indicates that pathogenic mutations in the helicase motifs of IGHMBP2 reduce its enzymatic activities, either both ATPase and helicase activities or helicase activity alone as in the case of the p.D565N variant (Table 2).

We also tested a DSMA1-causing mutation, p.T493I, which is not located in close proximity to conserved helicase motifs and found that it indeed did not interfere with enzymatic activity (Fig. 6B and C). However, we have recently shown that the p.T493I mutation reduces the intracellular steady-state levels of IGHMBP2, hence providing an explanation for its pathogenicity (25).

Finally, we investigated whether pathogenic mutations might also affect IGHMBP2 binding to ribosomes. This was exemplified by the incubation of three mutant proteins (p.D565N, p.R603H and p.T221A) in cytoplasmic extracts and subsequent sedimentation analysis on glycerol gradients in comparison to wild-type IGHMBP2 (Fig. 7). Neither mutants p.T221A and p.R603H (both helicase and ATPase deficient and p.R603H additionally impaired in nucleic acid binding) nor the mutant p.D565N (ATPase active, but helicase deficient) exhibited any difference in ribosomal association. These data are in good agreement with our finding that wild-type and p.R603H mutant IGHMBP2 protein co-localize indistinguishably from each other with a ribosomal marker (elF4G2) in N1E115 mouse neuroblastoma cells (Supplementary Material, Fig. S8). Thus, dissociation of IGHMBP2 from ribosomes is not the major defect in cells of DSMA1-patients. Collectively, our studies suggest that most DSMA1 missense mutations affect the catalytic activity of IGHMBP2 suggesting that DSMA1 results primary from impaired helicase activity of IGHMBP2.

DISCUSSION

Although substantial improvements in understanding the genetics and the clinical spectrum of DSMA1 have been made, only little is known about the function of the disease gene product IGHMBP2. Here, we provide evidence that IGHMBP2 acts as an ATP-dependent RNA helicase with a direct association with

fractions were run on a PAGE gel and visualized by silver staining. Lower panel: Control experiment with FM3A cytoplasmic extract that was conducted strictly in parallel to the glycerol gradient centrifugation of the upper panel. Due to the lower abundance of endogenous IGHMBP2, the protein had to be identified by western blot in the respective fractions with antibodies raised against the full-length protein. The presence of 18S and 28S rRNAs in the small and large complex (middle panel) identifies these forms as 40S and 60S ribosomal subunits. Note: The slight differences in the migration behavior of GST-IGHMBP2 in different fractions are due to the different amounts of proteins loaded. The identity of IGHMBP2 in both, the binding experiment as well as in the gradient has been confirmed by mass spectrometry. (**D**) IGHMBP2 interacts with 80S ribosomes. Upper panel: The IGHMBP2 complex isolated in the presence of Mg^{2+} as well as FM3A cytoplasmic extract prepared strictly in parallel under the same conditions was separated by sucrose gradient centrifugation. Under these conditions, recombinant as well as endogenous IGHMBP2 was detected in a single 80S peak (compare upper and lower panels) containing 18S and 28S rRNAs (middle panel). M, molecular weight marker; R, control binding reaction on GST-protein; I, IGHMBP2-bound components prior to gradient centrifugation.

Table 1. Proteins associated with IGHMBP2

Number	Protein	M (kDa)	Mascot score
1	RpL4	49.9	155.7
	Serpine 1 mRNA binding protein 1	44.7	356.1
2	EF1A	50.1	353.3
	Y box binding protein 1	35.7	446.1
	EF1G	49.9	1426.5
3	EF1G	49.9	632.4
	elF4 subunit 6 (elF3-p48)	52.2	690.6
	EF1A	50.1	124.4
	RpL3	46	177.3
	Y box binding protein 1	35.7	374.8
4	elF3 subunit 4 (elF3-p44)	35.6	272.0
	Y box binding protein 3	38.8	278.0
5	hnRNP E1	37.5	533.1
6	elF3 subunit 2, elF3-p36	36.4	498.2
	RpL6	33.4	506.5
7	RpL5	34.2	306.8
	RpL0 (ribosomal protein p0)	34.2	131.3
8	RpL7	29.8	273.2
	RpL8	27.9	240.9
	RpS3	26.7	552.9
9	GST	25.8	864.5
10	RpL17	21.3	229.1
	RpL18	20.1	237.9
	RpL23	17.7	182.5
11	RpL11	20.1	148.6
	RpL23	17.7	173.1
12	RpS11	18.4	219.2
	RpL26	17.2	345.5
13	RpS18	17.7	400.5
	RpS15	16.9	503.2
	RpS13	17.1	235.2
	RpL32	15.7	228.8
	RpL28	15.6	196.4

Putative IGHMBP2 interactors identified through an affinity matrix of immobilized recombinant IGHMBP2 and subsequent identification by nano LS MS/MS analysis of peptides derived from tryptic in gel digests (43). The Mascot Score (MS) is a measure of significance of protein identification and is given as $MS = -10^{*}\log(p)$, where p is the probability that the observed match is a random event. Three or more matching peptides and a significant probability score ($P < 0.05$ corresponds to a Mascot Score >35, SWISS-PROT database October 2007) were required to secure identity assignment.

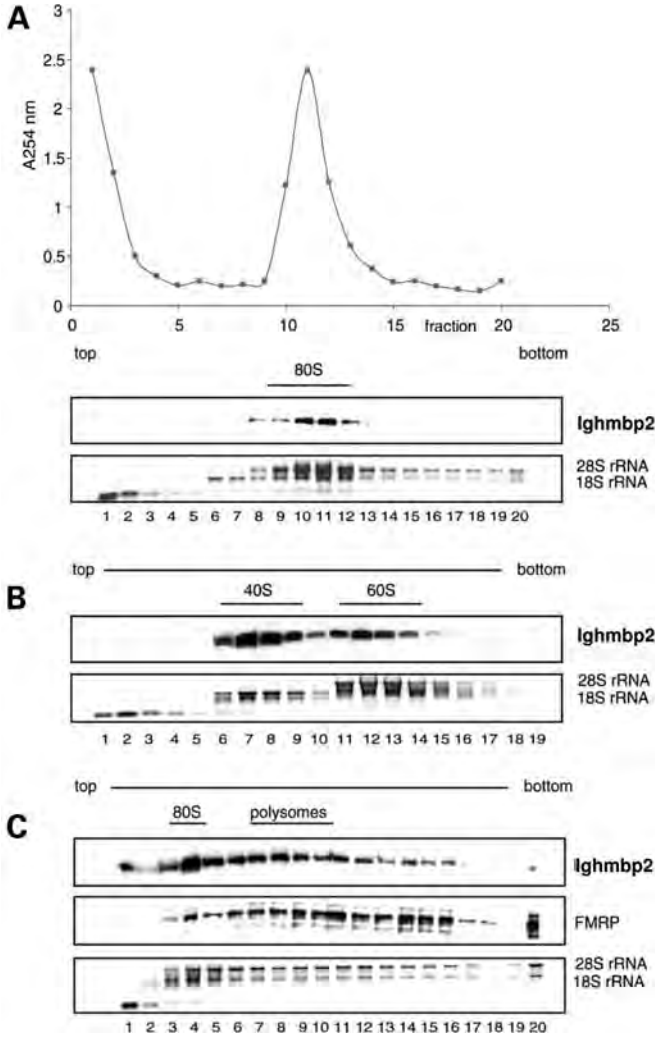


Figure 5. Ighmbp2 associates with ribosomes *in vivo*. Cytoplasmic extracts were prepared from mouse FM3A cells, fractionated on linear sucrose gradients and Ighmbp2 was detected with an antiserum raised against the full-length protein. (**A**) Ighmbp2 associates with 80S ribosomes in the presence of Mg^{2+}. (**B**) Ighmbp2 associates with 40S and 60S ribosomal subunits in the presence of 15 mM EDTA (upper panel), as indicated by co-sedimentation of 18S and 28S rRNAs (lower panel). (**C**) Ighmbp2 loosely associates with polysomes in the presence of cycloheximide. Ighmbp2 was detected primarily in the 80S peak and to a lesser extent also in the polysomal fractions.

ribosomes predominantly in the cytoplasm. Our experimental approach is different from previous studies since it utilizes recombinant protein, purified in its active state by two affinity steps. This strategy not only allowed for a comprehensive analysis of pathogenic DSMA1 mutations, but has also provided reliable controls against spuriously contaminating activities.

We show that IGHMBP2 complies with the characteristics of canonical helicases, (i) it hydrolyzes preferably ATP and discriminates among NTP co-factors, (ii) its ATPase activity is stimulated by RNA or DNA and (iii) it is capable of unwinding nucleic acid duplices *in vitro*. Further, IGHMBP2 not only shares sequence and structural similarities with other members of the UPF1-like helicases within superfamily 1, such as senataxin (SETX) or Up-frameshift mutation 1 (hUPF1/RENT1), but also with the enzymatic features reported for these helicases (9,21,26). Our recombinant IGHMBP2 not only displayed specific ATPase activity but also unwound RNA or DNA duplices in $5' \rightarrow 3'$ direction. This is in accord with a previous report (12), but in conflict with a $3' \rightarrow 5'$ unwindase activity observed in another study (11). However, with our experimental approach that uses Walker A and B mutants of IGHMBP2 in parallel, we can unambiguously rule out that the activity observed with wild-type IGHMBP2 originates from adventitious contaminants. It is not surprising that IGHMBP2 unwinds arbitrary duplices chosen on behalf of the unknown natural substrates. As observed with other helicases *in vitro*, it is reasonable to assume that IGHMBP2 would only exhibit sequence specificity upon interactions with other proteins (20,21,27).

As another approach to elucidate the malfunction of IGHMBP2 in DSMA1 patients, we have tested pathogenic mutants of IGHMBP2 for their enzymatic activities and their ability to associate with ribosomes. Among these, only the mutation p.T493I was not located within known motifs of

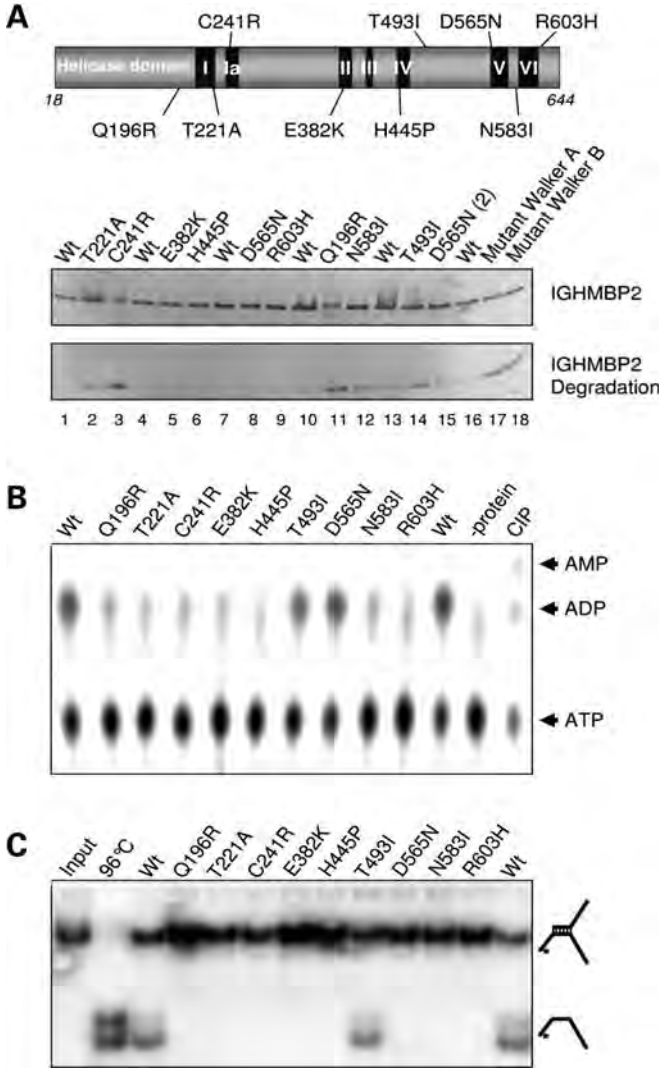


Figure 6. Pathogenic variants of IGHMBP2 show impaired enzymatic activity. (**A**) DSMA1-related mutants of IGHMBP2. Upper panel: Position of all analyzed DSMA1-mutations within the helicase domain. Lower panel: Expression and purification of IGHMBP2 variants. Silver-staining of an SDS–PAGE gel verified that similar amounts of wild-type and mutant protein were used in the assays. (**B**) ATPase activity of DSMA1-related IGHMBP2 variants. ATPase activity of wild-type and mutant IGHMBP2 protein was stimulated with poly(A) RNA homopolymers under standard assay conditions. All pathogenic IGHMBP2 variants except p.T493I and p.D565N displayed severely reduced ATPase activity. (**C**) RNA helicase activity of DSMA1-related IGHMBP2 variants. Except for the p.T493I mutant, all DSMA1-related IGHMBP2 variants did not display any unwinding activity on RNA duplices.

the helicase domain. Indeed, the protein showed unaltered enzymatic activity. Instead, a plausible reason for the pathogenicity of this mutant was recently attributed to its apparent lower steady-state levels in DSMA1 patient cell lines (25). Clearly, further analyses are needed to illuminate the etiology of DSMA1-causing missense mutations, which are not located in the proximity of the helicase motifs of IGHMBP2. All other DSMA1 mutations tested, however, reside in or close to conserved motifs within the helicase domains (representing the majority of IGHMBP2 mutants known), and, indeed, impair enzymatic activity. This is not a foregone conclusion for all tested IGHMBP2 mutant proteins based on structural data of the helicase domain of hUPF1, which shares high sequence identity with the helicase domain of IGHMBP2 (9). Most IGHMBP2 mutations analyzed by us (p.Q196R, p.T221A, p.C241R, p.E382K, p.H445P, p.N583I, p.R603H), however, resulted in loss of ATPase activity, which is a prerequisite for helicase activity. This makes the loss of ATPase activity the primary defect in these mutants. The p.N583I and p.R603H mutants additionally exhibited inefficient binding to nucleic acids, suggesting that different mechanisms cause the loss of ATPase activity (Supplementary Material, Fig. S4). The p.D565 N mutation, in contrast, uncouples ATPase activity from RNA unwinding. A similar effect has previously been reported for mutations in the UL5 and PcrA paralogues (28–30). From our mutational analysis we conclude that DSMA1 mutations primarily affect either (i) the ATPase activity, (ii) the unwinding activity or (iii) lead to severely reduced steady-state protein levels *in vivo* (Table 1). Of note, all disease-causing mutations that are found outside of the helicase domain result in the generation of mRNAs with premature stop codons. These mRNAs are likely to be eliminated by nonsense mediated mRNA decay (NMD) and hence are not translated into mutant protein. These findings further support the conclusion that the loss of the helicase activity of IGHMBP2 is the primary pathological basis of DSMA1. The next logical step for the understanding of the DSMA1 pathogenesis would thus be the analysis of the cellular pathways, which critically depend on the helicase function of IGHMBP2.

In view of the various cellular functions currently proposed for IGHMBP2, we have investigated its subcellular localization and the interactions it engages in. We have employed monospecific antibodies against endogenous IGHMBP2, as well as tag-specific antibodies against recombinant IGHMBP2, and find in each case predominant staining in the cytoplasm. In order to identify interacting partners of IGHMBP2, we performed a series of sedimentation and interaction experiments, which revealed an association of IGHMBP2 with ribosomes.

Several lines of experimental evidence support this conclusion: First, by immunofluorescence, IGHMBP2 was found to co-localize to a large extent—though not exclusively—with a translation factor (eIF4G2) and ribosomal RNA (rRNA) in the cytoplasm. Second, centrifugation of extracts in density gradients under physiological conditions revealed that IGHMBP2 sediments in an RNase A sensitive complex with a sedimentation coefficient of 80S, which corresponds to that of initiating ribosomes. Third, and most importantly, immobilized recombinant IGHMBP2 specifically bound to ribosomes and formed complexes indistinguishable from those observed in cellular extracts. Of note, the p.R603H mutant, which is incapable of binding to artificial nucleic acids still associates with ribosomes. IGHMBP2 may hence contact the ribosome primarily via protein-protein interactions. However, we cannot exclude at present that IGHMBP2 associates with secondary structures in rRNA not present in our artificial substrates or contacts rRNA by means of sequence motifs outside the helicase domain.

Our data raise the possibility that IGHMBP2 controls the translation of specific mRNA molecules. The finding that the protein interacts with ribosomal subunits and 80S ribosomes

Table 2. Summary of the functional analysis of IGHMBP2 mutants

IGHMBP2 mutant	Wt	Q196R	T221A	C241R	E382K	H445P	T493I	D565N	N583I	R603H	Walker A	Walker B
Helicase motif	−	−	I	Ia	II	IV	−	V	−	VI	I	II
Degradation	−	++	+	++	+	−	−	−	+	−	++	++
RNA binding	++	+	++	+	+	+	+	++	+	−	n.d.	n.d.
DNA binding	++	−	++	−	+	+	+	++	−	−	n.d.	n.d.
ATPase activity	++	−	−	−	−	−	++	++	−	−	−	−
RNA/DNA unwinding	++	−	−	−	−	−	++	−	−	−	−	−

Effect of DSMA1-related mutations on IGHMBP2 enzymatic activities *in vitro*. −, complete loss of activity; +, moderate reduction of activity; ++, full activity; n.d., not determined.

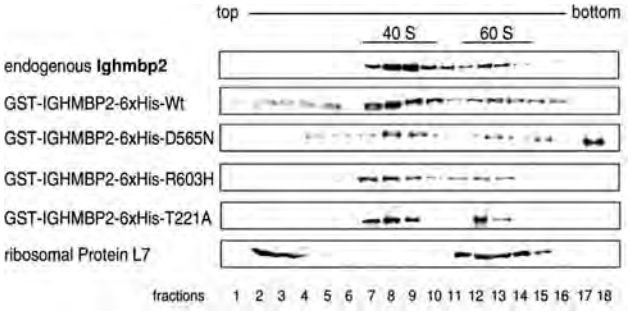


Figure 7. IGHMBP2 mutants are still able to bind ribosomal subunits. All mutants tested (p.D565N, p.R603H and p.T221A) had a similar sedimentation profile on a glycerol gradient when compared with wild-type and endogenous Ighmbp2. The sedimentation profile of the 60S ribosomal subunit was confirmed by western blotting with an α-RPL7 antibody (bottom panel).

rather than with polysomes further strengthens this view. Such a scenario would be expected for a protein that acts early in translation, possibly at the level of translation initiation. Other functions for IGHMBP2 related to translation such as regulation of mRNA stability cannot be excluded at present.

Previous studies had indicated that a significant fraction of cellular Ighmbp2 localizes to the nuclear compartment where it was suggested to act on DNA and in pre-mRNA splicing. In contrast, we have detected Ighmbp2 primarily in the cytoplasm of different cell types (including motoneurons) and in association with ribosomes. These data raise the possibility that Ighmbp2 is a multifunctional factor acting in several different pathways. Clearly, further detailed functional analysis of this protein is required to address this issue.

Our results support the idea that the impaired helicase activity of ribosome-associated IGHMBP2 is the cause for the motoneuron degeneration in DSMA1. Interestingly, a dysregulation of RNA metabolism is assumed to underlie other motoneuron diseases, such as amyotrophic lateral sclerosis type 4 (ALS4) as well as ataxia with ocular apraxia 2 (31–33) and has recently been proposed for SMA (6,34,35). Comparable to the situation SMA, it remains puzzling why in DSMA1-patients motoneurons are specifically vulnerable to the loss of IGHMBP2 function. However, our results pinpoint the assumption that dysregulation of neuronal mRNA metabolism might be the common denominator between SMA and DSMA1. We speculate that the neuron-specific phenotype seen in DSMA1 patients might be attributed to a particular requirement of neurons for IGHMBP2 to efficiently translate certain mRNAs (e.g. RNAs with excessive secondary and/or tertiary structures). This hypothesis appears to be attractive as it could also explain other tissue specific effects observed in the *nmd* mouse model of DSMA1 (5,36). The dilated cardiomyopathy (DCM) and diaphragmatic weakness in this model might be due to less efficient translation of myocyte-specific mRNAs. Identification of mRNAs whose expression depends on IGHMBP2 in motoneurons may hence give insight into the cell-type specific features of the disease.

As an alternative to the aforementioned scenario, the phenotype of DSMA1 could result from a neuron-specific function of IGHMBP2. Motoneuron integrity might thus rely on the availability of ribosome-associated IGHMBP2 involved in processes such as mRNA localization or localized protein synthesis.

MATERIAL AND METHODS

Cloning and *in vitro* mutagenesis of IGHMBP2

The full-length open-reading frame of human *IGHMBP2* cDNA sequence was cloned between *Sal* I and *Not* I restriction sites into the pBluescript KSII(−)-vector (Stratagene, Amsterdam, The Netherlands). Subsequently, a linker sequence containing a 6×His-tag was introduced at the C-terminus by dual asymmetric PCR (37). This plasmid (pBS-*IGHMBP2*-6×His) was used as template for *in vitro* mutagenesis following the protocol of the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The absence of fortuitous mutations was verified by automatic DNA sequencing. Subsequently, all open-reading frames were subcloned into the pGEX-6P1 vector (GE Healthcare, Munich, Germany) thus creating the different expression plasmids (pGEX-*IGHMBP2*-wt/mut-6×His).

The mutant constructs contained either a missense mutation originally found in DSMA1 patients or substitutions of highly conserved residues in the Walker A (GKT219AAA) and B-motif (DE375AA) of the ATPase domain. DSMA1-related mutations were chosen because of their position within or adjacent to different helicase motifs (p.Q196R, p.T221A, p.C241R, p.E382K, p.H445P, p.D565N, p.N583I, p.R603H). One mutation (p.T493I) was not in close proximity to any helicase motif but was chosen because it had been identified in two DSMA1 patients with a variable onset of disease.

For expression of wild-type and p.R603H-*IGHMBP2*-FLAG in eukaryotic cells, we used the pCMV-Tag vector system (Stratagene). In brief, full-length human *IGHMBP2* cDNA without the stop codon was amplified with tailed primers and

cloned in-frame between the *Not* I and *Sal* I restriction sites of the pCMV-Tag4b vector. Correct cloning was verified through automatic sequencing. Subsequently, the c.1808G>A mutation (p.R603H) was introduced into pCMV-Tag4b-wt-*IGHMBP2* using oligonucleotide-induced site-directed mutagenesis and subsequent selection of mutants with *Dpn* I (38). The derived pCMV-Tag4b-R603H-*IGHMBP2* clone was also controlled by automatic sequencing.

Expression of IGHMBP2 protein

After transformation of the constructs into *E. coli* Rosetta-gami™ (DE3)pLysS (Novagen, Darmstadt, Germany), expression was induced by adding IPTG to 1 mM at 16°C overnight. Purification of wild-type and mutant IGHMBP2 proteins was achieved by two sequential affinity-based purification steps performed at 4°C. In brief, crude bacterial extracts were prepared by sonification in lysis buffer [20 mM HEPES-NaOH pH 7.0, 1 M NaCl, 5 mM $MgCl_2$, 20% (v/v) glycerol, 0.01% NP-40, 5 mM β-mercaptoethanol and the following protease inhibitors 0.1 mM AEBSF, 0.5 mg/l leupeptin, 0.7 mg/l pepstatin A and 2 mg/l aprotinin]. The extracts were ultracentrifuged (48 400*g*, 1 h, 4°C) and the supernatants were applied on GST-Sepharose 4B-columns (GE Healthcare), pre-equilibrated with lysis buffer for 2 h. Columns were then washed twice with wash buffer A (same as lysis buffer but without protease inhibitors) and wash buffer B (same as buffer A but with 0.5 M NaCl). The IGHMBP2 part of the fusion proteins was then cleaved off the bound GST-tags by PreScission™ protease (GE Healthcare) in 4 ml wash buffer B at 4°C overnight. The resulting supernatants were applied onto Ni-NTA agarose columns (Qiagen, Hilden, Germany) which were then washed with one volume wash buffer C (same as buffer B but with 20 mM imidazole at pH 7.0) and two volumes wash buffer D (same as buffer C but with 50% (v/v) glycerol). Bound proteins were eluted with 1 ml elution buffer (same as buffer D but with 300 mM imidazole at pH 7.0) and fractions of 250 µl were subsequently analyzed by SDS–PAGE and Coomassie blue staining. Fractions containing the highest concentrations of IGHMBP2 were pooled and stored at −20°C in 50% (v/v) glycerol. For high-level expression of full-length wild-type 6×His-IGHMBP2 see Supplementary Material.

ATPase assays

0.9 pmol (=100 ng) of purified IGHMBP2 was incubated in 50 µl of a standard ATPase reaction mixture using α[^{32}P]-ATP (Hartmann Diagnostics, Braunschweig, Germany) for 1 h at 37°C (39). Single-stranded homopolymeric RNA (Sigma-Aldrich, Munich, Germany) or double-stranded DNA (GE Healthcare) was added to a final concentration of 0.1 mg/ml. All reactions were supplemented with 10% (v/v) glycerol. Reactions were stopped by the addition of one volume of 0.5 M Na_2EDTA pH 8.0, and 0.8 µl were analyzed on Poly(ethylenimine)-cellulose (PEI) thin-layer chromatography plates (Merck, Darmstadt, Germany). Hydrolysis products were visualized by autoradiography and on a FujiBas-2000 PhosphoImager (Fujifilm Europe, Düsseldorf, Germany) and quantified with TINA v2.0 software (Raytest, Berlin, Germany). To identify hydrolysis products, a standard reaction was carried out with 0.1 U Calf Intestine Phosphatase (CIP) (New England Biolabs, Frankfurt, Germany) for 6 min at 37°C.

Unwinding assays

The helicase activity was measured in 40 µl of standard assay mixture (modified after Laggerbauer *et al*. 40) containing 25 fmol radioactively labeled nucleic acid, 0. 9 pmol of purified protein (=100 ng full length protein), 50 mM NaCl, 30 mM Tris–HCl pH 7. 0, 1.2 mM $MgCl_2$, 1.5 mM DTT, 0.5 mM EDTA, 20 units of RNasin™ Plus (Promega, Mannheim, Germany), 200 µM ATP and 15% glycerol. RNasin™ Plus was omitted from the DNA unwinding reaction. The mixture was incubated at 30°C for 1 h and the reactions were then stopped by the addition of 150 mM Na_2EDTA pH 8.0, 50% (v/v) glycerol, 2% (w/v) SDS and 0.25% (w/v) xylene cyanol. Reaction products were subjected to gel electrophoresis through non-denaturing 8% PAGE (19:1) in 0.5× TBE at 4°C. Visualization was achieved by autoradiography.

Antibodies

Antibodies against IGHMBP2 were generated by immunization of rabbits with two different bacterially expressed mouse Ighmbp2 proteins: full-length Ighmbp2, and an N-terminal fragment (1–296 amino acid) (ImmunoGlobe, Himmelstadt, Germany). The antisera were affinity-purified on their antigen covalently linked to CnBr-activated Sepharose (GE Healthcare). Both antibodies (α-full length-IGHMBP2 and α -N-terminal-IGHMBP2) were used for western blot analyses and recognize two bands of 120 and 75 kDa (Supplementary Material, Fig. S6). The bands were both identified as Ighmbp2 by means of protein mass spectrometry of immunoprecipitated samples.

The antiserum used for motoneuron staining (α-C terminal-IGHMBP2, Fig. 2A and B) had been generated in rabbit against a peptide of the most C-terminal 19 amino acids of Ighmbp2 and had been characterized previously (5). A polyclonal antibody against goat eIF4G2 (DAP-5, G-20) and the mouse monoclonal antibody against rRNA (Y10b) were purchased from Santa Cruz (Heidelberg, Germany) and the α-FLAG M2 antibody from Stratagene. Cy2-coupled α-goat IgG antibody and Cy3-coupled antibody against rabbit IgG were obtained from Dianova (Hamburg, Germany). A polyclonal antibody against ribosomal protein 7 (RPL7) was a kind gift of Dr Dabauvalle (University of Wuerzburg). Secondary Alexa555-coupled α-mouse IgG and Alexa488-coupled α-goat IgG antibodies were purchased from Invitrogen (Karlsruhe, Germany).

Cell culture, transfection, immunofluorescence and microscopy

Mouse N1E115 neuroblastoma cells. Cells were grown on 10 mm diameter glass cover slips in 5% CO_2 atmosphere, and transfected with Transfectin™ Lipid Reagent (Biorad, Munich, Germany) as published (41). After 24 h the cells were fixed with 4% PFA and permeabilized with 0.1%

Triton X-100/0. 05% SDS and blocked with 20 mM glycine pH 8.5. Incubation with the DAP-5 antibody (1:100) was done overnight and subsequently with the mouse α-FLAG M2 antibody (1:1000) for 30 min. Secondary antibodies (Alexa555-coupled α-mouse IgG, 1:200) and (Alexa488-coupled α-goat IgG, 1:200) were each successively incubated for 30 min. Finally the nuclei were stained with DAPI (Roche, Mannheim, Germany) at 0.5 mg/l final concentration. The cells were then mounted with Mowiol (Sigma-Aldrich) and visualized under a Leica laser confocal microscope (Leica Microsystems, Wetzlar, Germany).

Primary mouse motoneuron culture. The ventro-lateral part of the lumbar spinal cord of E14 embryos was dissected and transferred to Hank's balanced salt solution. After treatment with 0.05% trypsin (w/v) for 15 min, cells were triturated and cultured after enrichment by panning with antibodies against the mouse p75 neurotrophin receptor (Abcam, Cambridge, MA). Cells were plated at a density of 2000 cells/cm^2 in four-well plates for 5 days in Neurobasal Medium (Invitrogen) with 2% horse serum, 500 μM GlutaMAX^{TM}-I supplement (Invitrogen) and B27 supplement (1:50; Invitrogen) at 37°C in a 5% CO_2 atmosphere. GDNF (Promega), CNTF (M. Sendtner, Institute for Clinical Neurobiology, Germany) and BDNF (Promega) were added to the medium at a final concentration of 10 ng/ml each. The medium was replaced every second day. After a culture of 7 days on glass cover slips, cells were fixed with 1.5 ml 4% PFA for 10 min. After blocking with 10% BSA cells were incubated overnight with the following antibodies: Rabbit polyclonal antiserum against Ighmbp2 (α-C terminal IGHMBP2), DAP-5 (G-20, 1:50) and Y10B antibodies (1:1000). Cells were then washed three times with 1× TBS-T (200 mM TRIS, pH 8.0, 8% NaCl, 1% Tween-20 (Sigma-Aldrich) and incubated for 1 h at room temperature with the following secondary antibodies Cy2-conjugated α-rabbit (1:200, Dianova) or a Cy3-conjugated α-mouse, α-goat (1:300, Dianova), respectively. After washing with 1× TBS-T the cover slips were embedded in Mowiol (Sigma-Aldrich) and digital photos were taken with a Leica SP2 with a HCX PL APO CS 63× 1.4 oil-immersion objective, with identical settings for pinhole and voltage in control and Ighmbp2-deficient motoneurons.

For the quantification of Ighmbp2 within the different cellular compartments the staining intensity in the cell body, the axon and the growth cone were analyzed with the AIDA software (Raytest). Background intensity was measured for every single image. The intensity of Ighmbp2 staining was depicted as arbitrary units per area, based on quantum levels per pixel, according to the protocol of the manufacturer.

Preparation of cell extracts and glycerol gradients

Mouse FM3A cells from a suspension culture were harvested by centrifugation at 200*g* for 5 min at 4°C. Cells were washed twice with ice-cold PBS, and then suspended in cell lysis buffer (10 mM HEPES-KOH pH 7. 5, 150 mM NaCl, 0.01% NP-40, 1 mM DTT and protease inhibitor). After 15 min incubation on ice, the cells were homogenized with a Dounce homogenizer. Homogenization was evaluated under a phase-contrast microscope. The homogenate was then centrifuged for 15 min at 2000*g* and the supernatant, now referred to as cytoplasmic fraction, was centrifuged at 25 000*g* for 30 min. Linear gradients of 5–30% (v/v) glycerol were prepared in 10 mM HEPES-KOH pH 7.5, 150 mM NaCl using a gradient mixer. After equilibrating the gradients to 4°C, 500 μl of FM3A cytoplasmic extract were layered on top, followed by centrifugation for 5 h at 180 000*g* at 4°C. Fractions of 700 μl were collected from the top of the tube. Protein samples from each fraction were precipitated by trichloracetate (TCA) and analyzed by western blotting with α-IGHMBP2 (and α-RPL7) antibodies. To examine the sedimentation of recombinant IGHMBP2 in cellular extracts (Fig. 4C and D) and the ribosome association of pathogenic IGHMBP2 variants (Fig. 6D), 100 ng recombinant wild-type or mutant IGHMBP2 were incubated with 500 μl FM3A cytosolic extract (10 μg/μl) for at least 10 min on ice prior to glycerol gradient centrifugation, and analyzed as described earlier.

Ribosome/polysome profile analysis

For the analysis of the Ighmbp2-ribosome/polysome association, mouse FM3A cell extract was prepared in cell extraction buffer [20 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.3% Igepal and protease inhibitor]. To dissociate ribosomes or polysomes, cells were lysed in the same buffer but without $MgCl_2$. The cytoplasmic fraction was adjusted to 15 mM of EDTA and incubated for 20 min on ice prior to the sedimentation on a sucrose gradient. For Ighmbp2-polysome association, the FM3A cells were incubated for 15 min with 100 μg/ml cycloheximide prior to the cell harvest to accumulate polysomes, stalled in translocation. Sucrose gradients of 5–30 and 20–50% (w/v) were prepared in 20 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM $MgCl_2$ or 15 mM EDTA using a gradient mixer. 0.5 ml of the extract was loaded onto the corresponding gradient followed by centrifugation at 198 000*g* for 135 min or for 4 h (for the gradient with EDTA) at 4°C. Fractions of 700 μl were prepared from each gradient and 500 μl of each fraction was directed to TCA precipitation and analyzed by western blotting with α-IGHMBP2 antibody. One hundred microliters of each fraction were further used to measure the RNA content at $\lambda = 254$ nm, and 100 μl were used for RNA extraction. Ribosomal profiles were obtained by plotting the absorbance A_{254} in each fraction against the fraction numbers. RNA was extracted with phenol, precipitated with ethanol, resolved on a 1% agarose gel containing ethidiumbromide and visualized under UV-illumination.

Purification of the IGHMBP2 complex

Cytoplasmic extract from mouse FM3A cells was prepared in lysis buffer [10 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.01% Igepal and 5% glycerol]. GST-IGHMBP2-6×His was expressed and purified as described earlier. The fact that human and mouse IGHMBP2 are 77% identical and 84% similar at the protein level allowed us to use recombinant human IGHMBP2 for immunoprecipitation of the IGHMBP2 complex from mouse cell extracts.

Before incubation with cytoplasmic extract ~100 μg GST-IGHMBP2-6×His and 100 μg GST were immobilized on Glutathione–Sepharose beads (GE Healthcare) and washed with lysis buffer. To reduce non-specific binding of proteins, cytoplasmic extract (15 ml) was pre-adsorbed on GST beads for 1 h at 4°C, before the flow through from this column was incubated with a GST-IGHMBP2-6×His-column for 1 h at 4°C. The beads were washed twice with 10 ml lysis buffer, transferred into a new 1.5 ml tube followed by washing with 5 × 1 ml lysis buffer. IGHMBP2 or GST and its associating proteins were eluted with 2 × 300 μl elution buffer [50 mM glutathione-SH in 10 mM HEPES-KOH pH 8.0, 150 mM NaCl, 1 mM DTT and protease inhibitor] at 4°C. The protein eluates were TCA-precipitated, resolved in 2× SDS loading buffer [125 mM Tris–HCl pH 6.8, 17% (v/v) glycerol, 4.1% (w/v) SDS, 0.001% (w/v) bromphenol blue, 17% (v/v) b-mercaptoethanol], loaded on a 10% SDS–polyacrylamide gel and visualized by silver staining. RNA was isolated from 100 μl of each eluate as described earlier.

For further purification of the isolated IGHMBP2 complex, we carried out affinity purification as mentioned earlier and 0.5 ml of the eluate was loaded on a linear 5–30% (v/v) glycerol gradient prepared as described earlier. To analyze the *in vitro* IGHMBP2 association with 80S ribosomes, FM3A cytoplasmic extract used for the affinity purification of the complex was prepared in a buffer containing $MgCl_2$ [20 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.3% Igepal and protease inhibitors]. The IGHMBP2 complex was affinity-purified from the extract as described above and was sedimented on a 7–47% (w/v) sucrose gradient containing 20 mM HEPES-KOH pH 7.5, 100 mM KCl and 5 mM $MgCl_2$. The sucrose and glycerol gradients were centrifuged for 4 h at 180 000*g*. The gradients were fractionated into 17 fractions of 700 μl. One hundred microliters from each fraction was subjected to RNA isolation and 600 μl of each fraction was directed to TCA precipitation followed by SDS–PAGE and silver staining (42).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We would like to thank Susanne Lützkendorf and Barbara Lucke for excellent technical assistance and Michael Sendtner and Christoph Hübner for their help, suggestions and critically reading of the manuscript.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants from the Deutsche Forschungsgemeinschaft to K.v.A. (GR 1837/1-1 and HU 408/3-3), to M.S. (HU 408/3-3), to A.S. (FZT 82) and to U.F. (SFB 581 TP18) and by grants from the Charité (Rahel Hirsch fellowship for K.v.A.).

REFERENCES

1. Grohmann, K., Schuelke, M., Diers, A., Hoffmann, K., Lucke, B., Adams, C., Bertini, E., Leonhardt-Horti, H., Muntoni, F., Ouvrier, R. *et al.* (2001) Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nat. Genet.*, **29**, 75–77.
2. Grohmann, K., Varon, R., Stolz, P., Schuelke, M., Janetzki, C., Bertini, E., Bushby, K., Muntoni, F., Ouvrier, R., van Maldergem, L. *et al.* (2003) Infantile spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Ann. Neurol.*, **54**, 719–724.
3. Guenther, U.P., Varon, R., Schlicke, M., Dutrannoy, V., Volk, A., Hubner, C., von Au, K. and Schuelke, M. (2007) Clinical and mutational profile in spinal muscular atrophy with respiratory distress (SMARD): defining novel phenotypes through hierarchical cluster analysis. *Hum. Mutat.*, **28**, 808–815.
4. Cook, S.A., Johnson, K.R., Bronson, R.T. and Davisson, M.T. (1995) Neuromuscular degeneration (nmd): a mutation on mouse chromosome 19 that causes motor neuron degeneration. *Mamm. Genome.*, **6**, 187–191.
5. Grohmann, K., Rossoll, W., Kobsar, I., Holtmann, B., Jablonka, S., Wessig, C., Stoltenburg-Didinger, G., Fischer, U., Hubner, C., Martini, R. and Sendtner, M. (2004) Characterization of Ighmbp2 in motor neurons and implications for the pathomechanism in a mouse model of human spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Hum. Mol. Genet.*, **13**, 2031–2042.
6. Eggert, C., Chari, A., Laggerbauer, B. and Fischer, U. (2006) Spinal muscular atrophy: the RNP connection. *Trends Mol. Med.*, **12**, 113–121.
7. Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.*, **17**, 419–437.
8. Monani, U.R. (2005) Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease. *Neuron.*, **48**, 885–896.
9. Cheng, Z., Muhlrad, D., Lim, M.K., Parker, R. and Song, H. (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J.*, **26**, 253–264.
10. Gorbalenya, A.E. and Koonin, E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.
11. Molnar, G.M., Crozat, A., Kraeft, S.K., Dou, Q.P., Chen, L.B. and Pardee, A.B. (1997) Association of the mammalian helicase MAH with the pre-mRNA splicing complex. *Proc. Natl Acad. Sci. U.S.A.*, **94**, 7831–7836.
12. Biswas, E.E., Nagele, R.G. and Biswas, S. (2001) A novel human hexameric DNA helicase: expression purification and characterization. *Nucleic Acids Res.*, **29**, 1733–1740.
13. Mizuta, T.R., Fukita, Y., Miyoshi, T., Shimizu, A. and Honjo, T. (1993) Isolation of cDNA encoding a binding protein specific to 5′-phosphorylated single-stranded DNA with G-rich sequences. *Nucleic Acids Res.*, **21**, 1761–1766.
14. Shieh, S.Y., Stellrecht, C.M. and Tsai, M.J. (1995) Molecular characterization of the rat insulin enhancer-binding complex 3b2. Cloning of a binding factor with putative helicase motifs. *J. Biol. Chem.*, **270**, 21503–21508.
15. Zhang, Q., Wang, Y.C. and Montalvo, E.A. (1999) Smubp-2 represses the Epstein-Barr virus lytic switch promoter. *Virology*, **255**, 160–170.
16. Chen, Y.I., Moore, R.E., Ge, H.Y., Young, M.K., Lee, T.D. and Stevens, S.W. (2007) Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.*, **35**, 3928–3944.
17. Moore, M.J., Schwartzfarb, E.M., Silver, P.A. and Yu, M.C. (2006) Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. *Mol. Cell.*, **24**, 903–915.
18. Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.*, **12**, 1231–1245.
19. Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419**, 182–185.
20. Silverman, E., Edwalds-Gilbert, G. and Lin, R.J. (2003) DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene*, **312**, 1–16.
21. Czaplinski, K., Weng, Y., Hagan, K.W. and Peltz, S.W. (1995) Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. *RNA*, **1**, 610–623.

22. Weng, Y., Czaplinski, K. and Peltz, S.W. (1996) Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol. Cell Biol.*, **16**, 5477–5490.
23. Weng, Y., Czaplinski, K. and Peltz, S.W. (1998) ATP is a cofactor of the Upf1 protein that modulates its translation termination and RNA binding activities. *RNA*, **4**, 205–214.
24. Feng, Y., Absher, D., Eberhart, D.E., Brown, V., Malter, H.E. and Warren, S.T. (1997) FMRP associates with polyribosomes as an mRNP and the I304 N mutation of severe fragile X syndrome abolishes this association. *Mol. Cell.*, **1**, 109–118.
25. Guenther, U.P., Handoko, L., Varon, R., Stephani, U., Tsao, C.Y., Mendell, J.R., Lutzkendorf, S., Hubner, C., von Au, A.K., Jablonka, S. *et al.* (2008) Clinical variability in distal spinal muscular atrophy type 1 (DSMA1): determination of steady-state IGHMBP2 protein levels in five patients with infantile and juvenile disease. *J. Mol. Med*. Epub ahead of print: doi:10.1007/s00109-008-0402-7.
26. Kim, H.D., Choe, J. and Seo, Y.S. (1999) The sen1(+) gene of Schizosaccharomyces pombe a homologue of budding yeast SEN1 encodes an RNA and DNA helicase. *Biochemistry*, **38**, 14697–14710.
27. Rocak, S. and Linder, P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. *Nat. Rev. Mol. Cell Biol.*, **5**, 232–241.
28. Dillingham, M.S., Soultanas, P. and Wigley, D.B. (1999) Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic Acids Res.*, **27**, 3310–3317.
29. Graves-Woodward, K.L., Gottlieb, J., Challberg, M.D. and Weller, S.K. (1997) Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis DNA unwinding and coupling between hydrolysis and unwinding. *J. Biol. Chem.*, **272**, 4623–4630.
30. Soultanas, P., Dillingham, M.S., Wiley, P., Webb, M.R. and Wigley, D.B. (2000) Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J.*, **19**, 3799–3810.
31. Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A. *et al.* (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.*, **74**, 1128–1135.
32. Duquette, A., Roddier, K., Nabb-Baltar, J., Gosselin, I., St-Denis, A., Dicaire, M.J., Loisel, L., Labuda, D., Marchand, L., Mathieu, J., Bouchard, J.P. and Brais, B. (2005) Mutations in senataxin responsible for Quebec cluster of ataxia with neuropathy. *Ann. Neurol.*, **57**, 408–414.
33. Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le, B., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M. *et al.* (2004) Senataxin the ortholog of a yeast RNA helicase is mutant in ataxia-ocular apraxia 2. *Nat. Genet.*, **36**, 225–227.
34. Pellizzoni, L. (2007) Chaperoning ribonucleoprotein biogenesis in health and disease. *EMBO Rep.*, **8**, 340–345.
35. Rossoll, W., Jablonka, S., Andreassi, C., Kroning, A.K., Karle, K., Monani, U.R. and Sendtner, M. (2003) Smn the spinal muscular atrophy-determining gene product modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. *J. Cell Biol.*, **163**, 801–812.
36. Maddatu, T.P., Garvey, S.M., Schroeder, D.G., Hampton, T.G. and Cox, G.A. (2004) Transgenic rescue of neurogenic atrophy in the nmd mouse reveals a role for Ighmbp2 in dilated cardiomyopathy. *Hum. Mol. Genet.*, **13**, 1105–1115.
37. Young, L. and Dong, Q. (2004) Two-step total gene synthesis method. *Nucleic Acids Res.*, **32**, e59.
38. Sambrook, J. and Russell, D.W. (2006) *In vitro* mutagenesis using double-stranded DNA templates: selection of mutants with Dpn I. *CSH Protocols*, doi:10.1101/pdb.prot3813.
39. Laggerbauer, B., Lauber, J. and Luhrmann, R. (1996) Identification of an RNA-dependent ATPase activity in mammalian U5 snRNPs. *Nucleic Acids Res.*, **24**, 868–875.
40. Laggerbauer, B., Achsel, T. and Luhrmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices *in vitro*. *Proc. Natl Acad. Sci. U.S.A.*, **95**, 4188–4192.
41. Chong, K.W., Lee, A.Y., Koay, E.S., Seet, S.J. and Cheung, N.S. (2006) pH dependent high transfection efficiency of mouse neuroblastomas using TransFectin. *J. Neurosci. Methods*, **158**, 56–63.
42. Merril, C.R., Goldman, D., Sedman, S.A. and Ebert, M.H. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science*, **211**, 1437–1438.
43. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H.E., Schonfisch, B., Perschil, I., Chacinska, A., Guiard, B. *et al.* (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl Acad. Sci. U.S.A.*, **100**, 13207–13212.

Supplementary data

Material and Methods

Preparation of RNA an DNA duplices

RNA Duplices. Plasmids used to obtain run-off RNA-transcripts were pGEM-MoI and pGEM-C4 (1), pCDNA3 (Invitrogen) and pBluescript II KS(+)-Δ*Sac* I / *Kpn* I (2). One strand of the RNA duplex was labeled during transcription by incorporation of α[^{32}P]-CTP (3000 Ci/mmol, Hartmann Diagnostics). Unlabeled RNA was also obtained by run-off transcription and all transcripts were subsequently purified by denaturing gel electrophoresis (3). Annealing of the duplices was performed as previously described (2). T7 and SP6 RNA polymerases and all restriction endonucleases were obtained from New England Biolabs and T3 RNA polymerase from Promega. The RNA substrate C4/MoI was generated by annealing two RNA-transcripts (C4, pGEM-C4 linearized by *Pvu* II, 197nt, SP6 RNA polymerase; MoI, pGEM-MoI linearized by *Hind* III, 86nt, T7 RNA polymerase). The 5'-RNA substrate consisted of two templates (RNA-A1, pCDNA3 linearized by *Eco* RI, 57nt, SP6 RNA polymerase; RNA-A2, pCDNA3 linearized by *Not* I, 86nt, T7 RNA polymerase). The 3'-RNA substrate was prepared from two transcripts of the same plasmid (RNA-A3 and RNA-A4, pBluescript II KS(+)-Δ*Sac* I/*Kpn* I linearized by *Pvu* II, 139nt, T3 RNA polymerase and 229 nt, T7 RNA polymerase).

DNA duplices. To produce duplex DNA substrates, synthetic oligonucleotides in HPLC quality (Biomers, Ulm, Germany) were annealed as described above. Oligonucleotides were radioactively labeled with $\gamma[^{32}P]$-ATP (3000 Ci/mmol, Hartmann Diagnostics) using T4 polynucleotide kinase (New England Biolabs). The labeled DNA was purified by phenol/chloroform extraction with subsequent ethanol precipitation and gel filtration chromatography using G-25 columns (GE Healthcare). The 5‘-DNA duplex contained the radioactively labeled DNA-A1 oligonucleotide (5’-GAA TAG GGC CCT CTA GAT GCA TGC TCG AGC GGC CGC CAG TGT GAT GGA TAT CTG CAG-3’) and the DNA-A2 oligonucleotide (5’-GGG AGA CCC AAG CTT GGT ACC GAG CTC GGA TCC ACT AGT AAC GGC CGC CAG TGT GCT GGA ATT CTG CAG ATA TCC ATC ACA CTG GC-3’) whose sequence corresponded to the 5’-RNA substrate. The 3‘-DNA duplex was made by annealing the radioactively labeled DNA-A1rev oligonucleotide (5’-GAC GTC TAT AGG TAG TGT GAC CGC CGG CGA GCT CGT ACG TAG ATC TCC CGG GAT AAG-3’) and the DNA-A2rev oligonucleotide (5’- CGG TCA CAC TAC CTA TAG ACG TCT TAA GGT CGT GTG ACC GCC GGC AAT GAT CAC CTA GGC TCG AGC CAT GGT TCG AAC CCA GAG GG-3’). Both oligonucleotides contained the same sequence as their counterparts of the 5’-DNA substrate but in reverse order.

Electrophoresis Mobility Shift Assay

The complexes formed between IGHMBP2 and single-stranded (ss) RNA were analyzed in 40 µl of standard assay mixture used in the strand displacement assays but without ATP and containing 25 fmol single-stranded radioactively labeled nucleic acid instead of 25 fmol duplices. After 1 hour incubation at room temperature, the

IGHMBP2: RNA complexes were cross-linked with 0.2% glutardialdehyde to prevent potential dissociation during electrophoresis. At the pH 8.3 used in our electrophoretic system IGHMBP2 (pI = 9.1) was predicted to carry a net positive charge that would have forced the protein into the opposite direction as the nucleic acid. Hence, cross-linking was carried out for 10 min on ice and the reaction mixtures were then supplemented with 2x loading buffer (89 mM TRIS-borate; 2 mM EDTA pH 8.0, 20% glycerol, 10 µg/µl heparin, and 0.1% (w/v) xylenecyanol, 0.1% (w/v) bromphenol blue). Electrophoretic separation was performed at 25 mA through a 4.5% polyacrylamide gel (80:1) containing 4% glycerol in 0.5x TBE at 4°C followed by visualization through autoradiography.

Recombinant expression of wild-type IGHMBP2 for oligomerization studies

The full open reading frame of human *IGHMBP2* including start- and stop-codon was RT-PCR amplified from RNA extracts of EBV immortalized lymphoblastoid cells with tailed primers containing a *Sal* I (5') and *Not* I (3') restriction site and the proof-reading Phusion™ DNA polymerase (Finnzymes, Espo, Finland). The PCR product of 2991 bp was cloned into the *Sal* I / *Not* I sites of the multiple cloning site of pET-28a (Novagen) which appends a 6xHis-Tag to the N-terminus. The correct sequence of the clone (pET-28a-wt-*IGHMBP2*) was verified by automatic sequencing. The plasmid was electroporated into *Escherichia coli* Rosetta II cells (Novagen). Recombinant protein production was performed in 2.5 liter culture bottles that were grown at continuous shaking (225 RPM) at 37°C for around 4 hours until an OD_{600}=0.8 was reached. After adjustment for one hour at room temperature, protein expression was induced with 1 mM IPTG and the cells were then grown at 16°C overnight. The pellet was prepared at 7,000 x *g* centrifugation and subsequently disrupted by sonification (Ultrasonic disintegrator UP100H, Hielscher, Teltow, Germany) with 5 x 30 seconds at 100 W, 100% amplitude and intermittent cooling on ice for 30 seconds in lysis buffer [1 M NaCl, 20 mM HEPES pH 7.0, 20% Glycerol, 0.01% NP40, 5 mM β-mercaptoethanol, 5 mM $MgCl_2$ and the proteinase inhibitors: 0.7 µg/ml Pepstatin, 0.1 mM AEBSF, 2 µg/ml Apronitin, 0.5 µg/ml Leupeptin] at 4°C. The ultracentrifugation supernatant (30 minutes at 50,000 x *g*) with the biochemically active protein was then used to load a HiTrap™ Chelating (nickel) 5 ml column (GE Healthcare) for affinity purification with the ÄKTA Purifier™ system (GE Healthcare). Elution was done with a linear 25-500 mM imidazole gradient in 20 mM HEPES pH 7.0, 500 mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.1 mM AEBSF. The eluate containing the IGHMBP2 portion was pooled, dialyzed against 20 mM HEPES pH 7.0, 500

mM NaCl, 10% glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol + 0.1 mM AEBSF and concentrated to 10 mg/ml with ultracentrifugation concentrators (Vivaspin™ 4 ml, 50,000 MWCO, Sartorius, Göttingen, Germany). 1 mg protein were subsequently separated by size exclusion chromatography on a Superdex™ 200 10/300 GL column (GE Healthcare) at a flow rate of 0.5 ml/min and collected in 0.5 ml fractions. The column had been calibrated with proteins of different sizes (HMW Kit, GE Healthcare) before.

All the elution fractions were investigated on a Coomassie gel for the presence of the full-length IGHMBP2 band at around 120 kDa. In order to remove aggregates and degradation products, subsequently all fractions with a predominance of the full-length band were pooled again with the exception of those fractions (<9.0 ml elution volume) that were close to the exclusion volume of the FPLC column (8.21 ml) thus potentially representing protein aggregates. After concentration with 50,000 MWCO Vivaspin™ ultracentrifugation concentrators to 100 µl, the samples were rerun thrice on the Superdex™ 200 10/300 GL column (Supplementary Figure 9) with 150, 250 and 500 mM NaCl in the running buffer. Equal amounts from each 0.5 ml elution fraction were again analyzed on a Coomassie gel (lower panel, Supplementary Figure 9).

Figure legends

Supplementary Figure 1 Expression and purification of recombinant IGHMBP2. IGHMBP2 was expressed in *Escherichia coli* with an N-terminal GST- and a C-terminal 6xHIS-tag fusion tag. To purify IGHMBP2 we have developed a strategy which involved two affinity-based purification procedures.

Supplementary Figure 2 IGHMBP2 binds to nucleic acids. Electrophoretic mobility shift assays (EMSA) were performed with increasing amounts of wild-type (Wt)-IGHMBP2 in the presence of single stranded RNA or DNA. The position of the single stranded nucleic acid and the complexes are indicated. The multiple bands observed in the RNA binding assay at higher concentrations of Wt-IGHMBP2 may indicate that more than one molecule of IGHMBP2 can bind to the single stranded RNA.

Supplementary Figure 3 IGHMBP2 unwinds DNA-duplices in 5´→3´direction. Increasing amounts of Wt-IGHMBP2 were incubated with DNA-duplices with either 5' single-stranded overhangs (5'-DNA) or with 3' single-stranded overhangs (3'-DNA). Control reactions with mutant IGHMBP2 protein (DE375AA) displayed no unwinding activity. The position of the single and double stranded DNA strands is shown at the sides and the radio-labeled strands are marked with a star. Wt-IGHMBP2 unwound only DNA-duplices with 5' single-stranded tails corroborating the 5'→3' translocation polarity observed with the 5'-RNA substrates (see Figure 1E).

Supplementary Figure 4 IGHMBP2 mutants show variably reduced nucleic acid binding capacity. Electrophoretic mobility shift experiments were conducted with wild-type and mutant IGHMBP2 protein in the presence of either single stranded RNA or single stranded DNA under standard assay conditions (see Supplementary Figure 2). As for the wild-type protein, the multiple bands observed in the RNA binding assay may reflect the binding of more than one IGHMBP2 molecule to the single stranded RNA.

Supplementary Figure 5 DSMA1-related IGHMBP2 variants display impaired DNA unwinding activity. DNA unwinding activity of wild-type and mutant IGHMBP2 protein was assayed using a DNA duplex with a single stranded 5' overhang under standard assay conditions. As on RNA duplices, all DSMA1-related IGHMBP2 variants, with the exception of the T493I mutant, display no unwinding activity on DNA duplices.

Supplementary Figure 6 Western blots of cell extracts with two different IGHMBP2 antibodies demonstrating a main band at 120 kDa and, as verified by mass spectrometry, most probably a degradation band at 75 kDa. The α-C terminal IGHMBP2 antibody has been evaluated previously (4).

Supplementary Figure 7 Full length Ighmbp2 is reduced in cell body, axon and growth cone of Ighmbp2-deficient isolated motoneurons from the *nmd* mouse. *Upper panel*: Antisera against the C-terminus of Ighmbp2 were used for immunocytochemical analysis to depict the amount of full length IGHMBP2 and its distribution in soma,

axon and growth cone of Ighmbp2 deficient and control neurons. *Lower panel*: Ighmbp2 deficient cells show a significantly reduced full length Ighmbp2 presence throughout the entire motoneuron, including cell body, axon and growth cone. **Note:** Due to the intronic *Ighmbp2* splice site mutation in the *nmd* mouse, Ighmbp2 expression is not completely abrogated but only reduced to approximately 30% which, however, does not suffice to avert the disease phenotype.

Supplementary Figure 8: *Upper two* panels: N1E115 mouse neuroblastoma cells were transfected with a wildtype and a p.R603H-mutant IGHMBP2 construct containing the human protein with a C-terminal FLAG-tag. The overlay shows co-localization of immunoreactivity against the ribosome (eIF4G2, DAP-5) and against the IGHMBP2-FLAG fusion protein to a large extent. *Lower two panels*: Higher resolution confocal micrograph of the same cells highlighting the border area between nucleus and cytosol. Co-localization was not altered after using a p.R603H-mutant IGHMBP2-FLAG construct. As depicted in Table 1 and Supplementary Figure 4, the p.R603H variant is unable to bind artificial RNA substrates thus indicating that co-localization does not depend on the nucleic acid binding capacity of IGHMBP2s helicase domain.

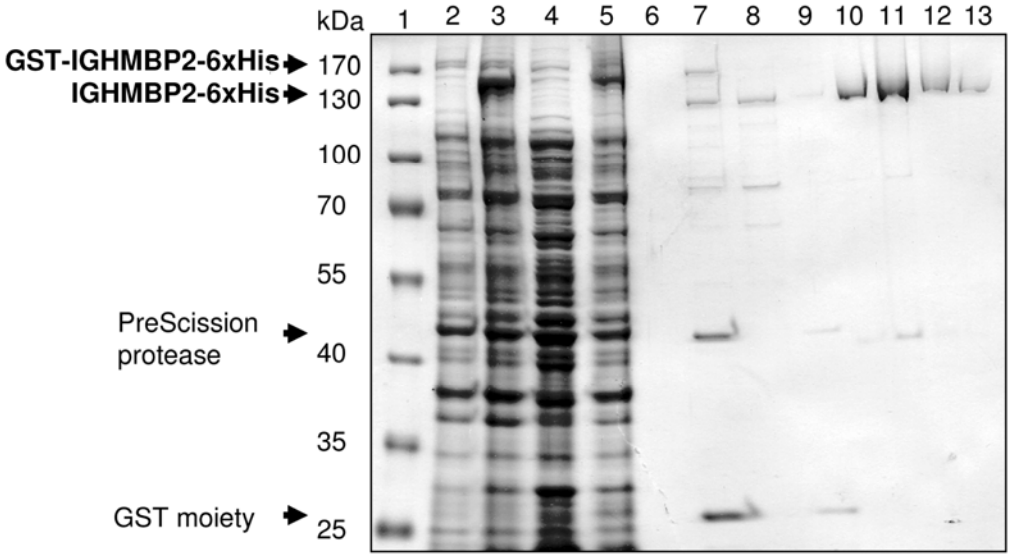
Supplementary Figure 9: FPLC elution curve from a high resolution Superdex^{TM} 200 10/300 GL column (GE Healthcare) of recombinant 6xHis-IGHMBP2 produced in *Escherichia coli*. The molecular sizes of the proteins and protein complexes within the elution fractions hint towards homo-oligomers with a maximum size of octamers (=856 kDa) and a predominant tetramer (=425 kDa). The right inset shows the calibration curve with proteins of known molecular weight. The red dots indicate the posi-

tion of the K_{AV} values of the 9.6 and 11.1 ml elution fractions and their respective molecular weights (left inset). The gel image below the elution curve depicts a Coomassie-stained PAGE-gel of equal volumes from the different elution fractions. The oligomerization of IGHMBP2 was not salt depended between 150 und 500 mM. The results are in good agreement with the elution profiles of IGHMBP2 homomers (syn. hHcsA) performed by another group (5).

References

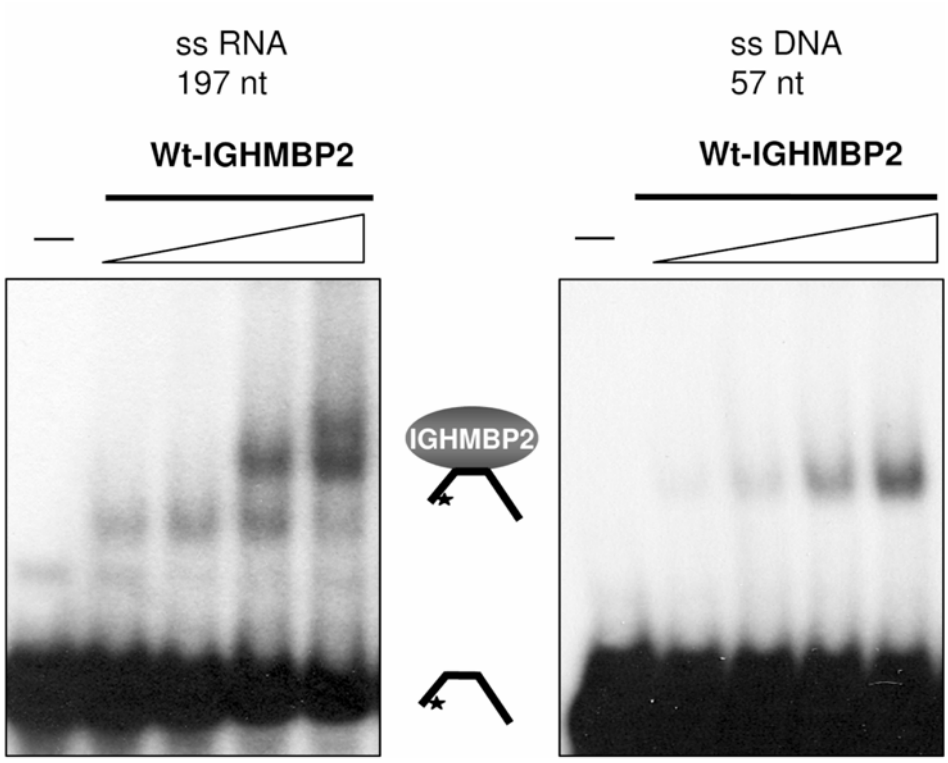
1. Scheffner, M., Knippers, R., and Stahl, H. 1989. RNA unwinding activity of SV40 large T antigen. *Cell* **57**:955-963.
2. Seybert, A., Hegyi, A., Siddell, S.G., and Ziebuhr, J. 2000. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA.* **6**:1056-1068.
3. Laggerbauer, B., Achsel, T., and Luhrmann, R. 1998. The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **95**:4188-4192.
4. Grohmann, K., Rossoll, W., Kobsar, I., Holtmann, B., Jablonka, S., Wessig, C., Stoltenburg-Didinger, G., Fischer, U., Hubner, C., Martini, R. et al. 2004. Characterization of Ighmbp2 in motor neurons and implications for the pathomechanism in a mouse model of human spinal muscular atrophy with respiratory distress type 1 (SMARD1). *Hum. Mol. Genet.* **13**:2031-2042.

5. Biswas, E.E., Nagele, R.G., and Biswas, S. 2001. A novel human hexameric DNA helicase: expression, purification and characterization. *Nucleic Acids Res.* **29**:1733-1740.

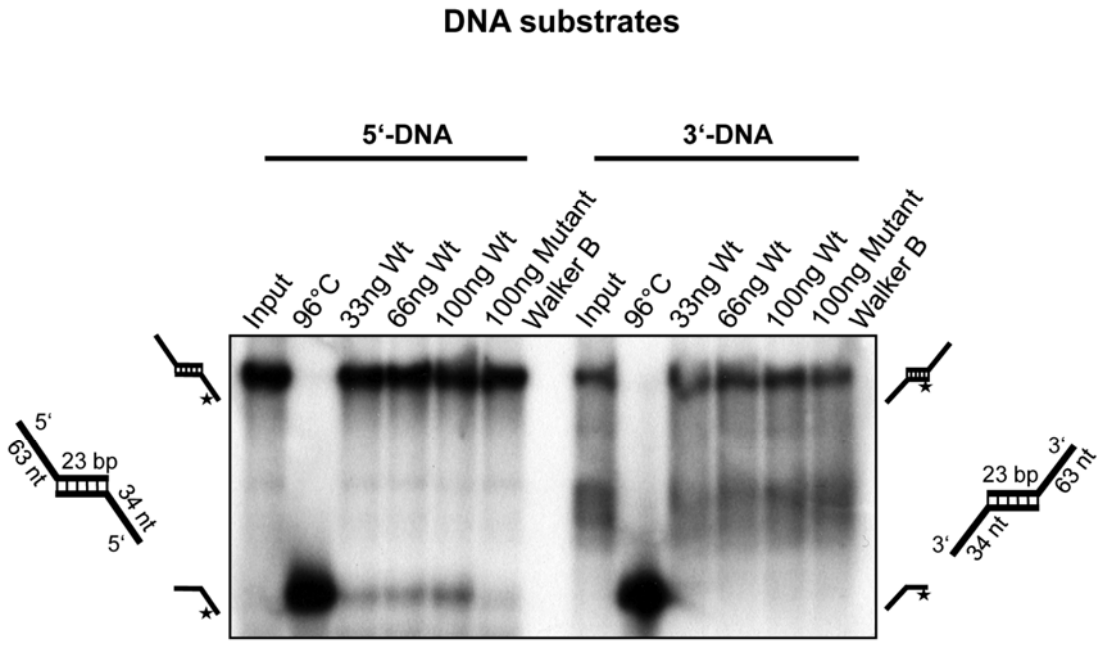


Lanes
1= marker
2= before induction
3= after induction
4= supernatant after centrifugation
5= pellet after centrifugation
6= supernatant after digestion with PreScission protease
7= GST-Sepharose after digestion with PreScission protease
8= flow-through on Ni-NTA beads
9= Ni-NTA beads after elution
10-13= fractions of elution

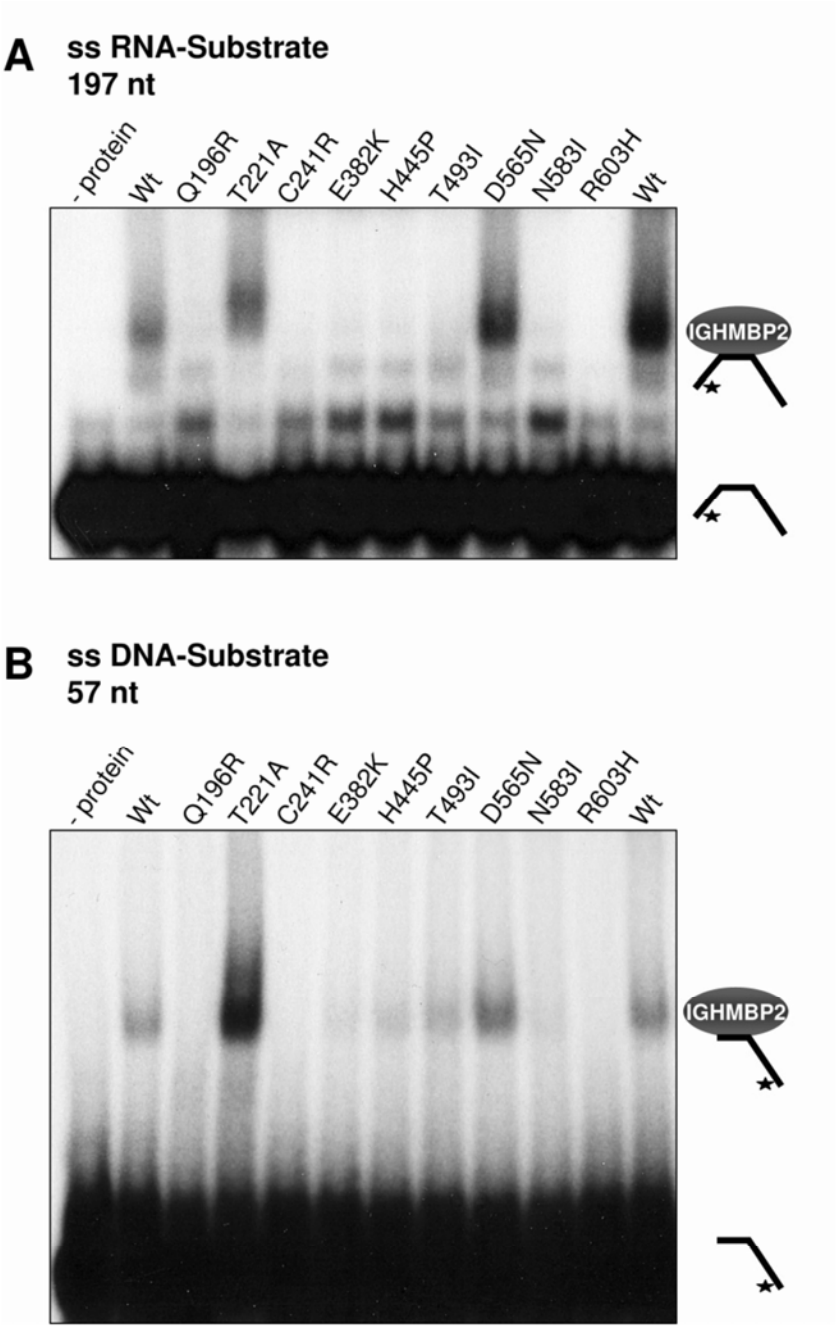
Guenther et al., Supplementary Figure 1



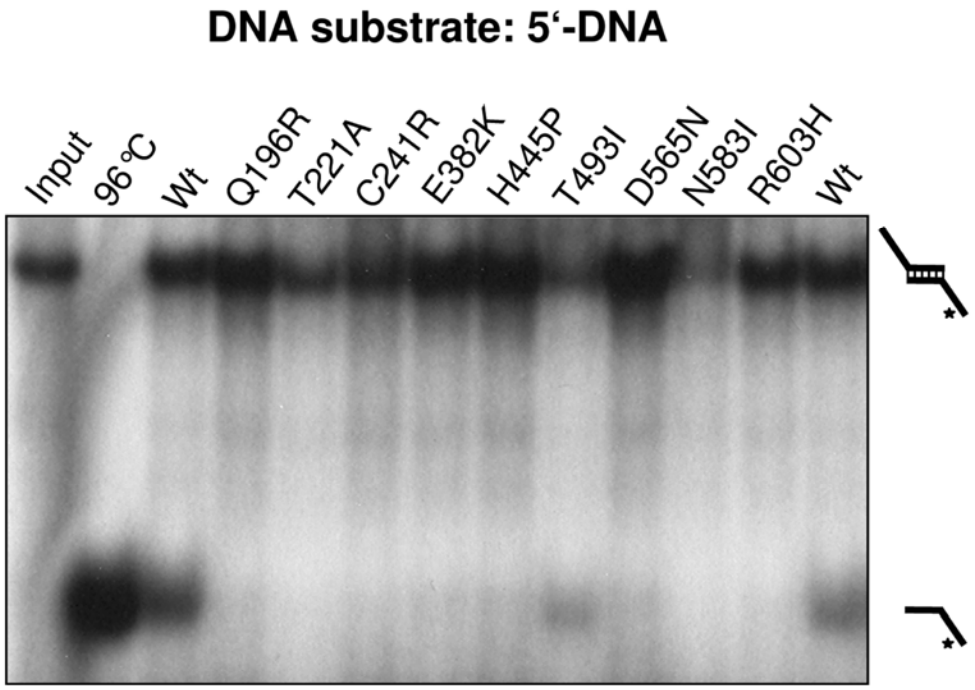
Guenther et al., Supplementary Figure 2



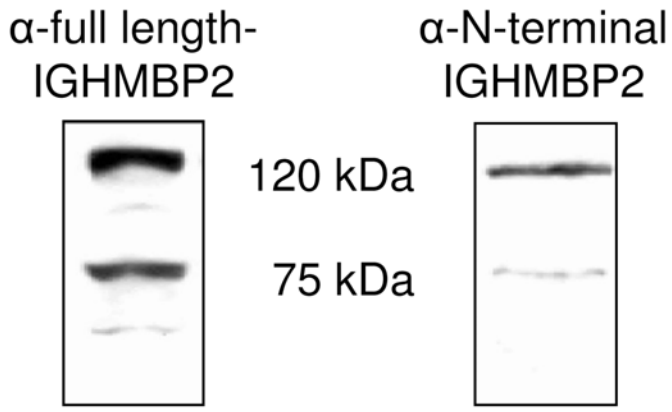
Guenther et al., Supplementary Figure 3



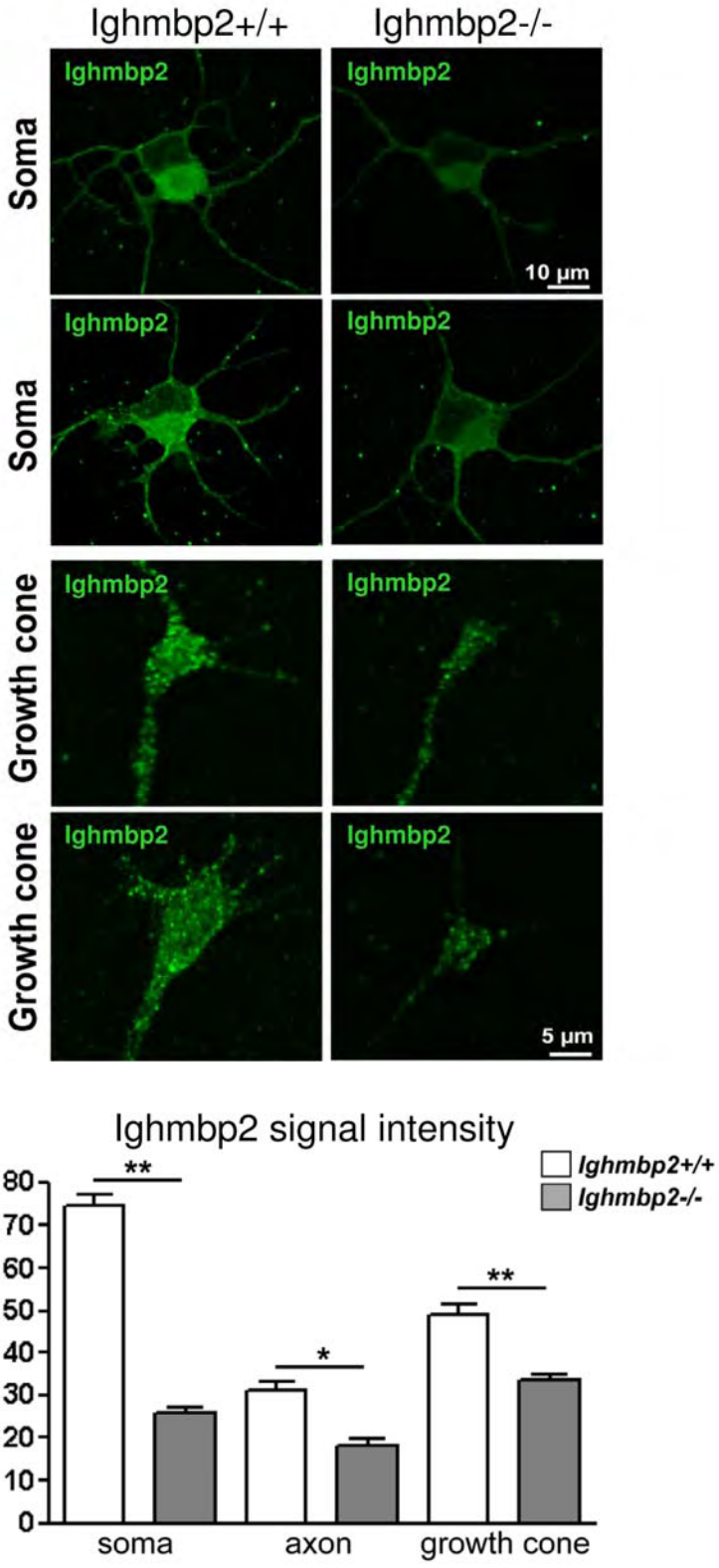
Guenther et al., Supplementary Figure 4



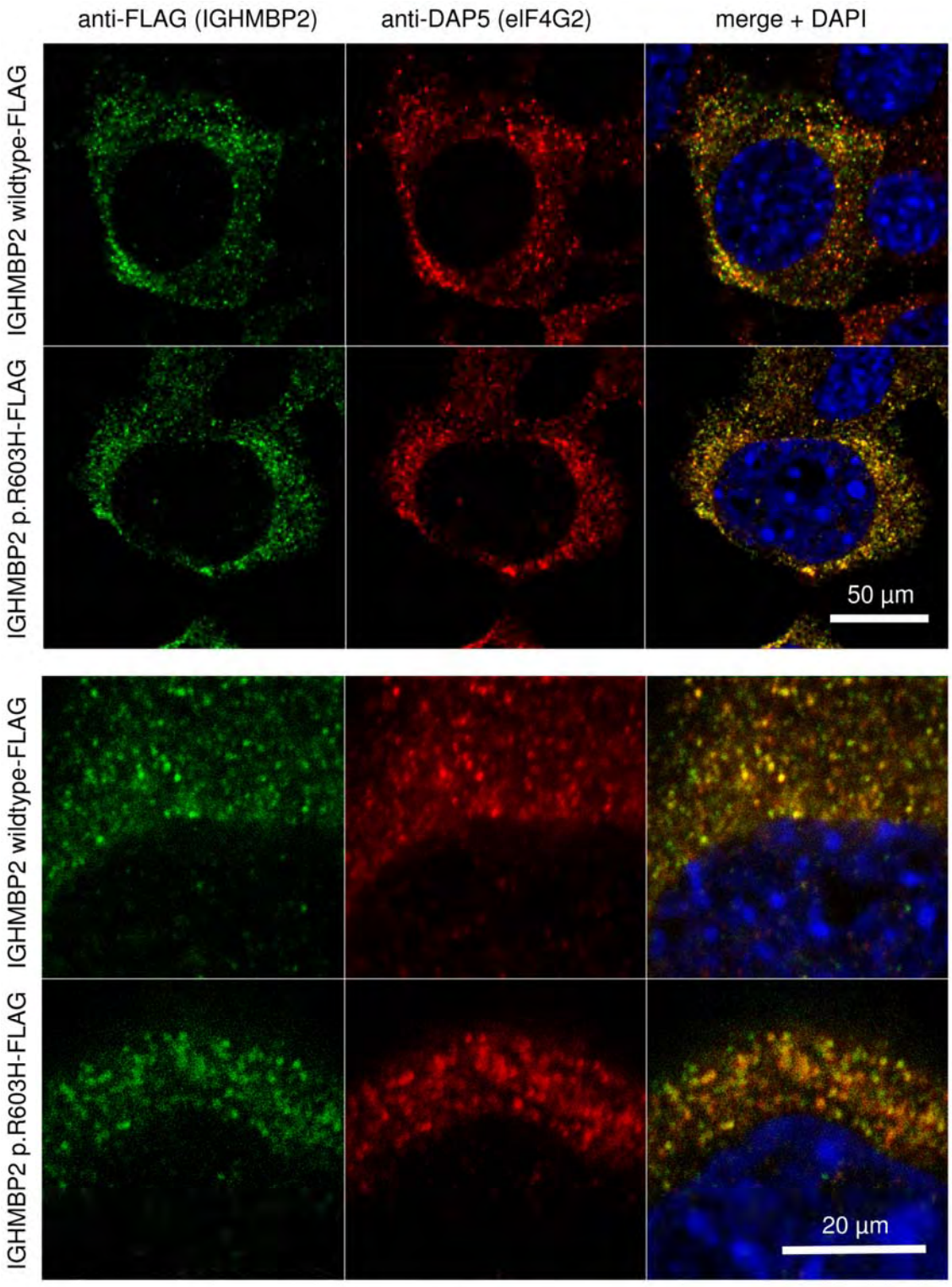
Guenther et al., Supplementary Figure 5



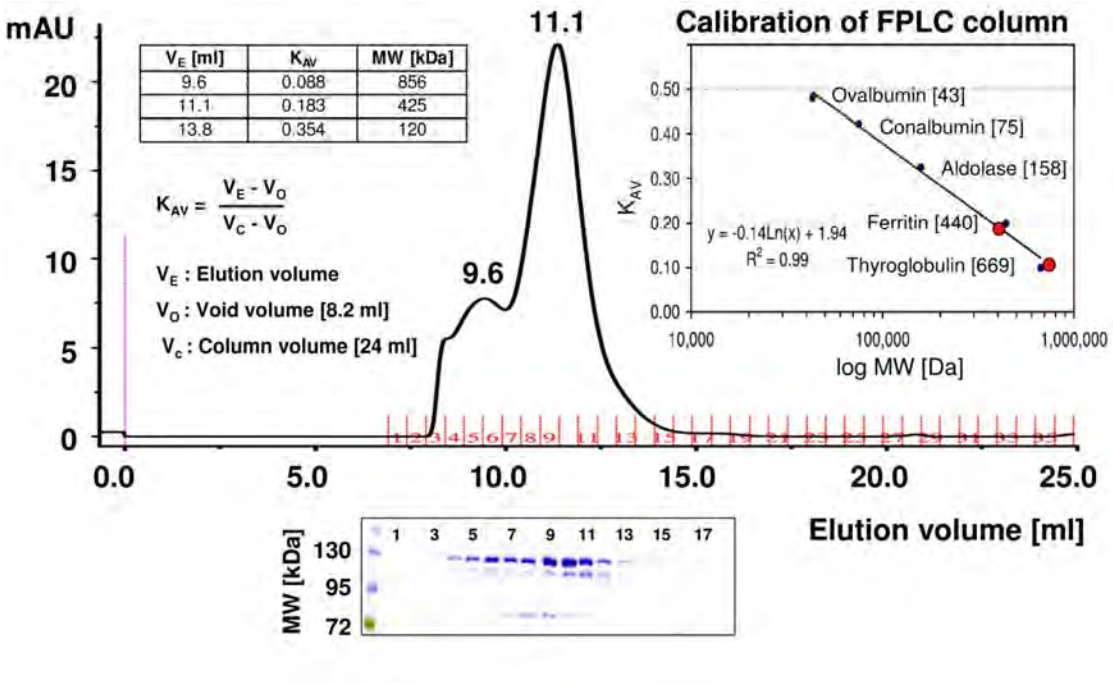
Guenther et al., Supplementary Figure 6



Guenther et al., Supplementary Figure 7



Guenther et al., Supplementary Figure 8



Guenther et al., Supplementary Figure 9

8.4 A 5´-Fluorosulfonylbenzoyladenosine-Based Method to Identify Physiological Substrates of a Drosophila p21-Activated Kinase

A 5´-fluorosulfonylbenzoyladenosine-based method to identify physiological substrates of a Drosophila p21-activated kinase.

Menzel N, Chari A, Fischer U, Linder M, Raabe T: **Anal Biochem 2007**, **368**:178-184.

Thesis author´s contribution:

Conception:	20 %
Experimental contribution:	20 %
Formulation of Results:	20 %

Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 368 (2007) 178–184
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

A 5′-fluorosulfonylbenzoyladenosine-based method to identify physiological substrates of a *Drosophila* p21-activated kinase

Nicolas Menzel [a], Ashwin Chari [b], Utz Fischer [b], Monica Linder [c], Thomas Raabe [a,*]

[a] *Institute of Medical Radiation and Cell Research, University of Würzburg, D-97078 Würzburg, Germany*
[b] *Institute of Biochemistry, D-97074 Würzburg, Germany*
[c] *Institute of Biochemistry, Protein Analytics, Justus-Liebig-University Giessen, D-35392 Giessen, Germany*

Received 12 March 2007
Available online 26 May 2007

Abstract

Nearly all processes in cells are regulated by the coordinated interplay between reversible protein phosphorylation and dephosphorylation. Therefore, it is a great challenge to identify all phosphorylation substrates of a single protein kinase to understand its integration into intracellular signaling networks. In this work, we developed an assay that holds promise as being useful for the identification of phosphorylation substrates of a given protein kinase of interest. The method relies on irreversible inhibition of endogenous kinase activities with the ATP analogue 5′-fluorosulfonylbenzoyladenosine (5′FSBA). 5′FSBA-treated cell extracts are then combined with a purified activated kinase to allow phosphorylation of putative substrate proteins, followed by a two-step purification protocol and identification by fingerprint mass spectrometry. Specifically, we applied this method to identify new phosphorylation substrates of the *Drosophila* p21-activated kinase (PAK) protein Mbt. Among candidate proteins identified by mass spectrometry, the dynactin complex subunit dynamitin was verified as a bona fide Mbt phosphorylation substrate and interaction partner, suggesting an involvement of this PAK protein in the regulation of dynactin-dependent cellular processes.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Identification of kinase phosphorylation substrates; *Drosophila* S2 cells; 5′FSBA inhibition of kinases; p21-activated kinase; Dynamitin

p21-activated kinases (PAKs)[1] are a family of serine/threonine kinases that are regulated through binding of Cdc42 and/or Rac-like RhoGTPases. By this means, they can act as central nodes in a variety of RhoGTPase-controlled cellular processes such as proliferation, migration, polarization, and reorganization of the cytoskeleton [1,2]. Based on structural features, PAK proteins are classified into groups I and II. Representative members of group I are human PAK1, -2, and -3 as well as *Drosophila* D-PAK and D-PAK3, whereas group II members include human PAK4, -5, and -6 as well as *Drosophila* Mbt. Members of both subgroups share an N-terminal binding site for Cdc42 and/or Rac GTPases and a C-terminal kinase domain, but they differ in the regulation of their kinase activity and the presence of additional protein interaction sites, which allow integration into different signaling pathways at distinct subcellular sites [3,4]. This can be exemplified for *Drosophila* D-PAK and Mbt, which are found in growth cones and at adherens junctions of developing photoreceptor cells, respectively [5–7].

Group II PAKs have received considerable interest recently because of their involvement in morphogenetic

* Corresponding author. Fax: +49 931 20145835.
E-mail address: thomas.raabe@mail.uni-wuerzburg.de (T. Raabe).

[1] *Abbreviations used:* PAK, p21-activated kinase; Limk1, LIM kinase 1; Ssh, Slingshot; cDNA, complementary DNA; PBS, phosphate-buffered saline; EDTA, ethylenediaminetetraacetic acid; EGTA, ethylene glycol-bis(β-aminoethylether)-*N,N,N′,N′*-tetraacetic acid; PMSF, phenylmethylsulfonyl fluoride; HRP, horseradish peroxidase; ECL, enhanced chemiluminescence; 5′FSBA, 5′-fluorosulfonylbenzoyladenosine; DTT, dithiothreitol; TCA, trichloroacetic acid; TOF, time-of-flight; BSA, bovine serum albumin; KESTREL, kinase substrate tracking and elucidation; MacMARCKS, macrophage-enriched myristoylated alanine-rich C kinase substrate.

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.05.022

processes. Constitutive activation of PAK4 in cultured cells leads to decreased cell adhesion, cell rounding, and induction of anchorage-independent growth. The PAK4 transcription unit localizes to a genomic region on chromosome 19 amplified in several tumors, and increased PAK4 activity can be detected in cell lines derived from many different cancers [8,9]. Studies in *Xenopus laevis* have revealed a role of X-PAK5 as a regulator of blastomeric adhesive properties by modulating calcium-dependent cell–cell adhesion [10]. These findings have been further corroborated by studies of *Drosophila* Mbt, which is involved in morphogenesis of photoreceptor cells [7]. In contrast to group I PAKs, few phosphorylation targets of group II PAKs have been described so far [3,11]. Cytoskeletal changes regulated by PAK4 are mediated by LIM kinase 1 (Limk1) and the phosphatase slingshot (Ssh), which act in an antagonistic manner on the actin depolymerization factor cofilin [12,13]. Constitutive activation of *Drosophila* Mbt also induces changes of the actin cytoskeleton, yet the phenotypic observations made in transgenic flies imply that additional proteins are subject to regulation by Mbt [14].

In this work, we set forth to identify additional phosphorylation substrates for Mbt. This was achieved by the supplementation of cell lysates with an activated Mbt variant after irreversible inhibition of endogenous kinase activities. Candidate substrate proteins were identified by peptide fingerprint mass spectrometry and were verified by in vitro kinase assays and coimmunoprecipitation studies. This method identified the *Drosophila* orthologue of the dynactin complex subunit dynamitin as a novel substrate of Mbt. In summary, this work has established a generic method for the identification of bona fide substrates of known kinases and implies a function of group II PAKs in the regulation of dynactin in cells.

Materials and methods

Cloning procedures

For expression of Myc–Mbt variants in S2R cells, the coding sequences of Mbt, Mbt^{T525A}, and $Mbt^{S492N/S521E}$ [14] were cloned 3′ to a 6× myc-tag sequence and then inserted into the Asp718/Bsp120i sites of pMT/V5 (Invitrogen). The complementary DNAs (cDNAs) for *Drosophila* CG1532 (LD36566), CG4802 (GM01885), dynamitin (LD07994), fimbrin (LD05347), junctophilin (GH05993), and thioredoxin reductase 1 (LD21729) were obtained from the Drosophila Genomics Resource Center. The coding sequences of CG1532, CG4802, and dynamitin were amplified by PCR and cloned as *Nco*I/*Not*I fragments into the GST fusion vector pETM-30 (a gift from A. de Marco, European Molecular Biology Laboratory, Heidelberg, Germany). Fimbrin, junctophilin, and thioredoxin reductase 1 were ligated as *Eco*RI/*Xba*I, *Eco*RI/*Xho*I, and *Bam*HI/*Xba*I fragments into pGEX-KG. All constructs were verified by sequencing.

Expression and purification of recombinant proteins

Bacterial expression and purification of GST fusion proteins were performed as described previously [14].

Cell culture and immunoprecipitation

Drosophila S2R cells [15] were cultured at 25 °C in Schneider's Drosophila medium (Biowest) supplemented with 10% fetal calf serum and 1% penicillin–streptomycin (Gibco). Cells were detached with a cell scraper in phosphate-buffered saline (PBS). Stable S2R cell lines expressing Myc–Mbt, Myc–$Mbt^{S492N/S521E}$, and Myc–Mbt^{T525A} were generated by transfection of the constructs with calcium phosphate and selection with 100 μg/ml hygromycin B (Calbiochem). Myc–Mbt protein expression was induced by the addition of 0.8 mM $CuSO_4$ for 36 h. Cells were homogenized in lysis buffer (25 mM Tris [pH 7.5], 150 mM NaCl, 2 mM ethylenediaminetetraacetic acid [EDTA], 2 mM ethylene glycol-bis(β-aminoethylether)-*N,N,N′,N′*-tetraacetic acid [EGTA], 10% glycerol, 0.1% NP-40, 5 μg/ml antipain, 1 μg/ml aprotinin, 0.5 μg/ml leupeptin, 0.7 μg/ml pepstatin, 100 μg/ml phenylmethylsulfonyl fluoride [PMSF]). Total protein concentrations were determined by Bradford assay. For immunoprecipitations, lysates were mixed overnight at 4 °C with 30 μl G protein agarose (Roche) and 8 μl monoclonal anti-Myc antibody (9E10, Santa Cruz Biotechnology). Bound proteins were washed with lysis buffer prior to SDS–PAGE and then transferred to a nitrocellulose membrane. Membranes were probed with anti-Myc, anti-Mbt [7], and anti-dynamitin (BD Bioscience Pharmingen) antibodies. Detection was done with horseradish peroxidase (HRP)-conjugated secondary antibodies and enhanced chemiluminescence (ECL, GE Healthcare).

In vitro kinase assays

For in vitro kinase assays, immunoprecipitated Mbt proteins were washed in kinase buffer (20 mM Hepes [pH 7.6], 150 mM NaCl, 20 mM $MgCl_2$) and incubated with cell lysate fractions in kinase buffer containing 5 μCi of [γ-^{32}P]ATP (3000 Ci/mmol) for 30 min at 30 °C. To test for phosphorylation of GST fusion proteins, 30 μg of the corresponding GST protein was added to the reaction mixture. Reactions were terminated by boiling in SDS–PAGE loading buffer, followed by SDS–PAGE and autoradiography.

Preparation of ATPγS and 5′-fluorosulfonylbenzoyladenosine-treated S2R cell extracts

Confluent S2R cells from 14 10-cm-diameter dishes were washed three times with PBS and harvested. The cleared lysate was dialyzed against kinase buffer. To this extract, ATPγS or 5′-fluorosulfonylbenzoyladenosine (5′FSBA, Sigma) was added to 30 mM and incubated for 45 min at

30 °C. Excess ATPγS was removed by dialysis, and 5′FSBA was inactivated by the addition of 5 mM dithiothreitol (DTT).

Fractionation of cell extracts

First, 5 ml 5′FSBA-treated cell extract in kinase buffer was applied to a 1-ml HiTrap Q HP column (Amersham Biosciences). After washing the column with 10 ml kinase buffer, proteins were eluted with a three-step linear gradient against kinase buffer containing 1 M NaCl collecting 1-ml fractions (gradient: 0.15–0.35 M NaCl in 4 ml, 0.35–0.65 M NaCl in 19 ml, 0.65–1.00 M NaCl in 4 ml). PHOS-Select Iron Affinity Gel (Sigma) chromatography was performed according to the manufacturer's instructions with minor modifications. In brief, following anion exchange chromatography and the radioactive kinase assay, the fraction containing putative substrate proteins was adjusted to 250 mM acetic acid and 30% acetonitrile (final pH 2.5) and incubated with 80 µl PHOS-Select resin for 45 min at 4 °C on a rotator. Unbound proteins were removed by washing the beads twice with 500 µl of the same solution and once with 500 µl water. Phosphorylated proteins were eluted by the addition of 200 µl of 150, 250, and 400 mM ammonium hydroxide and were precipitated with trichloroacetic acid (TCA). Protein pellets were resuspended in SDS–PAGE loading buffer and subjected to SDS–PAGE, Coomassie staining, and autoradiography.

In-gel digestion and protein identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry

Bands were excised from the gel, and proteins were carbamidomethylated and digested in-gel using mass spectrometry-grade trypsin (Trypsin Gold, Promega) in 25 mM NH_4HCO_3. Mass spectra of resulting tryptic peptides were recorded using an Ultraflex time-of-flight/time-of-flight (TOF/TOF) (Bruker Daltonics) in the positive ion reflectron mode with dihydroxy benzoic acid (20 mg/ml) in 50% acetonitrile/0.85% *o*-phosphoric acid as matrix. Peptide mass profiles were analyzed with Mascot (www.matrixscience.com) using the National Center for Biotechnology Information database.

Results

Our goal was to devise a strategy with high selectivity that allows the identification of new substrates of the PAK protein Mbt from total lysates of *Drosophila* cells. The general strategy was to irreversibly block all endogenous kinase activities and then to perform in vitro kinase assays in the presence of exogenously provided Mbt. First, we searched for a cell line with no detectable Mbt expression. We reasoned that although Mbt is absent from these cells, at least a subset of substrates would be expressed and present in a hypophosphorylated form. On the other hand, such a cell line would allow ectopic expression of Mbt variants without possible interference by the endogenous protein. S2R cells [15] fulfill these criteria. As shown by Western blot with an anti-Mbt antibody [7], S2R cells do not express detectable levels of endogenous Mbt (Fig. 1A). Stable cell lines were established, allowing inducible expression of Myc epitope-tagged wild-type Mbt (Myc–Mbt), constitutively activated Mbt (Myc–$Mbt^{S492N,S521E}$), and kinase-dead Mbt (Myc–Mbt^{T525A}), with the latter used as a control in Fig. 1A.

Next we determined conditions to inhibit endogenous kinases present in the extract. Untreated cell extracts show high endogenous kinase activities as detected by in vitro kinase assays with [γ-^{32}P]ATP (Fig. 1B, lane 1). Pretreatment of the extract with the slowly hydrolyzable ATP analogue ATPγS resulted in a strong reduction of endogenous kinase activities; however, the residual activities still were too high for a reproducible and unambiguous identification of Mbt substrate proteins (Fig. 1B, lane 2). Therefore, we replaced ATPγS by the adenine nucleotide analogue

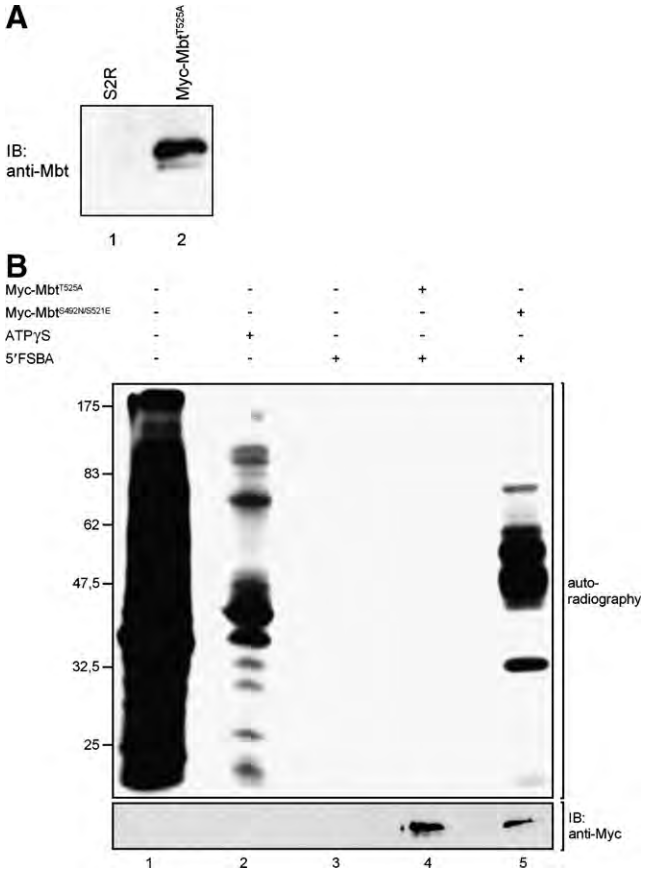


Fig. 1. (A) Total lysates of S2R cells (lane 1) and S2R cells expressing Myc–Mbt^{T525A} as control (lane 2) were analyzed for the presence of Mbt with an anti-Mbt antibody. (B) In vitro kinase assays of S2R cell extracts without pretreatment (lane 1) or after pretreatment with ATPγS (lane 2) or 5′FSBA (lane 3) are shown. The addition of purified activated Myc–$Mbt^{S492N,S521E}$ protein (lane 5), but not of kinase-dead Myc–Mbt^{T525A} protein (lane 4), to 5′FSBA-treated cell extracts results in the appearance of several distinct phosphorylation signals. The presence of the respective Myc–Mbt proteins was verified by Western blot (IB) with an anti-Myc antibody. Molecular weight markers are indicated at the left side.

5′FSBA, which targets a critical lysine residue in the ATP binding cleft and forms a sulfonamide, resulting in irreversible inhibition of the enzyme's ATPase activity [16]. In the presence of 30 mM 5′FSBA, endogenous kinase activities are completely blocked (Fig. 1B, lane 3). After inactivation of unbound 5′FSBA, in vitro kinase reactions were performed with purified Myc–Mbt^{S492N,S521E} or Myc–Mbt^{T525A}. In the presence of Myc–Mbt^{S492N,S521E}, several phosphorylation signals became visible (Fig. 1B, lane 5). The lack of any phosphorylation signal after supplementation with Myc–Mbt^{T525A} (Fig. 1B, lane 4) confirms the dependence on Mbt kinase activity and argues against the possibility that a copurified kinase performs the phosphorylation reaction.

Next we tried to reduce the complexity of proteins in the sample by fractionation of the 5′FSBA-treated extract on an anion exchange column. Proteins were eluted by applying a salt gradient, and the collected fractions (Fig. 2A, upper panel) were subjected to in vitro kinase reactions in the presence of Myc–Mbt^{S492N,S521E} (Fig. 2A, lower panel). Note the enrichment of Mbt phosphorylated proteins from fractions 25 to 27. In a further fractioning step, we performed affinity chromatography with the PHOS-Select resin. This method is based on the ability of Fe(III) to chelate phosphate groups on amino acid residues under denaturating conditions at pH 2.0. Elution of proteins from the column is achieved by the addition of increasing concentrations of $NH_4(OH)$. Fraction 25 from the anion exchange column (Fig. 2A) was first subjected to an in vitro kinase reaction with Myc–Mbt^{S492N,S521E} and then applied to the PHOS-Select resin (Fig. 2B). Note the enrichment of several Mbt phosphorylated polypeptides in the first elution fraction (Fig. 2B, autoradiography panel).

Gel slices from the indicated positions (Fig. 2B, bands a–e) were excised, and candidate proteins were identified by fingerprint mass spectrometry. Fig. 2C shows the list of identified proteins. To verify phosphorylation by Mbt, all candidate proteins (with the exception of bovine serum albumin [BSA], which was considered as a contaminant) were overexpressed as GST fusion proteins in *Escherichia coli* and were subjected to in vitro kinase assays with purified Myc–Mbt^{S492N,S521E} or Myc–Mbt^{T525A} (Fig. 3A). Only in the case of GST–dynamitin was efficient phosphorylation by Myc–Mbt^{S492N,S521E} observed. We further noticed a weak but nonreproducible phosphorylation of thioredoxin reductase 1 in some experiments, leaving open the possibility that this protein might be a substrate for Mbt. Kinase reactions in the presence of Myc–Mbt^{T525A} failed to yield any signal on autoradiography, confirming dependence of the phosphorylation reaction on intrinsic Mbt kinase activity.

Finally, we investigated whether Mbt and endogenous dynamitin are associated with each other in cells. This was assessed by coimmunoprecipitation experiments using the cell lines stably expressing Myc–Mbt, Myc–Mbt^{T525A}, or Myc–Mbt^{S492N,S521E}. Endogenous dynamitin was copurified with Myc–Mbt and Myc–Mbt^{S492N,S521E} (Fig. 3B, lanes 4 and 6) but not with the catalytically inactive mutant Myc–Mbt^{T525A} (Fig. 3B, lane 5). In summary, these results demonstrate that dynamitin is a bona fide substrate for the group II PAK protein Mbt.

Discussion

In this study, we have developed a highly stringent in vitro kinase assay to identify physiological phosphorylation substrates of a kinase of interest in cell extracts, exemplified by the identification of dynamitin as a novel substrate protein of the *Drosophila* kinase Mbt. In general, incubation of native cell extracts with [γ-^{32}P]ATP and a kinase of interest are of limited value to identify substrate proteins because the presence of multiple endogenous kinase activities results in high background phosphorylation, thereby masking kinase-specific phosphorylation signals. KESTREL (kinase substrate tracking and elucidation), a method that has been applied successfully to several protein kinases, tries to reduce the background problem by a combination of short incubation times, high concentrations of the added kinase, depletion of the endogenous ATP pool, the use of Mn^{2+} instead of Mg^{2+} as a cofactor, and a chromatography step to separate endogenous kinases from substrate proteins [17,18]. Another approach uses synthetic ATP derivatives, which can be used only by mutated forms of protein kinases [19]. We improved the signal/noise ratio by preincubation of the cell lysate with the nucleotide analogue 5′FSBA, which was formerly used in enzyme kinetic studies and for affinity labeling [16]. Because most ATP-hydrolyzing enzymes contain an evolutionary conserved motif for nucleotide binding and 5′FSBA reacts covalently, we expected a broad spectrum of inactivated ATP-, ADP-, NAD-, and NADH-dependent proteins [16]. Indeed, in vitro kinase assays with 5′FSBA-treated cell extracts showed complete blockage of endogenous kinase activities. The ATP analogue ATPγS turned out to be less effective, probably because it hydrolyzes slowly and then dissociates from the kinase. The assay requires the availability of a purified activated kinase, which in our case was isolated from a stably transfected cell line. One potential problem might be the copurification of other kinases. Control experiments with a kinase-dead variant of the kinase of interest or a comparative analysis of several closely related kinases should help to distinguish between background signals and physiologically relevant phosphorylation signals. For the identification of the proteins by peptide fingerprint mass spectrometry, it was necessary to reduce the complexity of the protein sample, first, by anion exchange chromatography and, second, by the PHOS-Select resin. Besides binding phosphorylated proteins, the PHOS-Select matrix is in principle also capable of retaining acidic proteins and proteins with two or more histidine residues in close proximity. Therefore, we consider the anion exchange column as an essential step and the PHOS-Select matrix as an

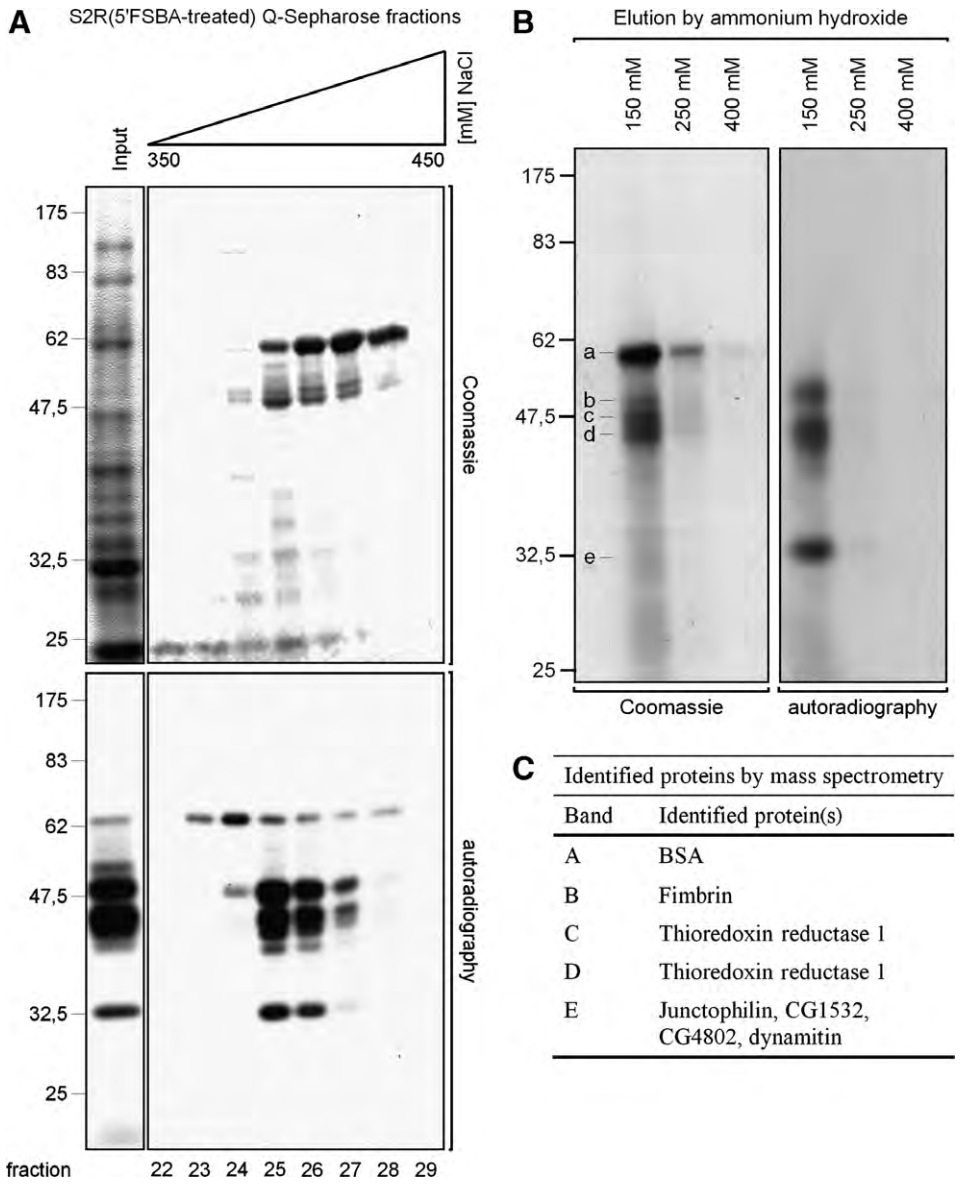


Fig. 2. (A) The 5′FSBA-treated S2R extract was loaded onto a HiTrap Q HP column and eluted in 60 1-ml fractions by applying a sodium chloride gradient (see Materials and methods). Then 30 μl of each fraction was analyzed for the presence of Mbt substrate proteins in radioactive kinase assays with purified Myc–Mbt^{S492N,S521E} (autoradiography) and in parallel for total protein content by Coomassie blue staining after SDS–PAGE. Only fractions 22 to 29 are shown. Molecular weight markers are indicated. (B) Enrichment of Mbt phosphorylated proteins by PHOS-Select Iron Affinity Gel chromatography is shown. Fraction 25 (panel A) was supplemented with Myc–Mbt$^{S492N/S521E}$ and [γ-^{32}P]ATP to label Mbt substrate proteins and then passed over the PHOS-Select matrix. After washing, phosphorylated proteins were eluted with increasing concentrations of ammonium hydroxide, separated by SDS–PAGE, and analyzed by autoradiography and Coomassie blue staining. Bands a to e were cut out for subsequent mass spectrometry analysis. (C) A list of candidate proteins identified by mass spectrometry is shown.

additional step to remove some, but certainly not all, unwanted proteins. Our experience with Mbt substrates, where one of six potential candidates was verified as true, underlines the necessity of appropriate purification protocols. Although we have been successful with anion exchange chromatography, other applications might warrant the replacement of this step by cation exchange chromatography.

In the case of Mbt, dynamitin could be verified as a true phosphorylation and interaction partner. Interestingly, only the wild-type and constitutively active version of Mbt were found to be associated with dynamitin, suggesting that this interaction is dependent on Mbt's kinase activity. Our results are also in line with a recent protein network analysis of the human Pak1–βPix–GIT–paxillin complex, which identified (among many other proteins) two components of the dynactin complex: dynamitin and p150^{Glued} [20].

The multisubunit complex dynactin is composed of two structural units: a rod-like unit, consisting of an Arp1 polymer and actin-capping proteins that provide the interface for cargo binding, and a projecting unit, formed by a dimer

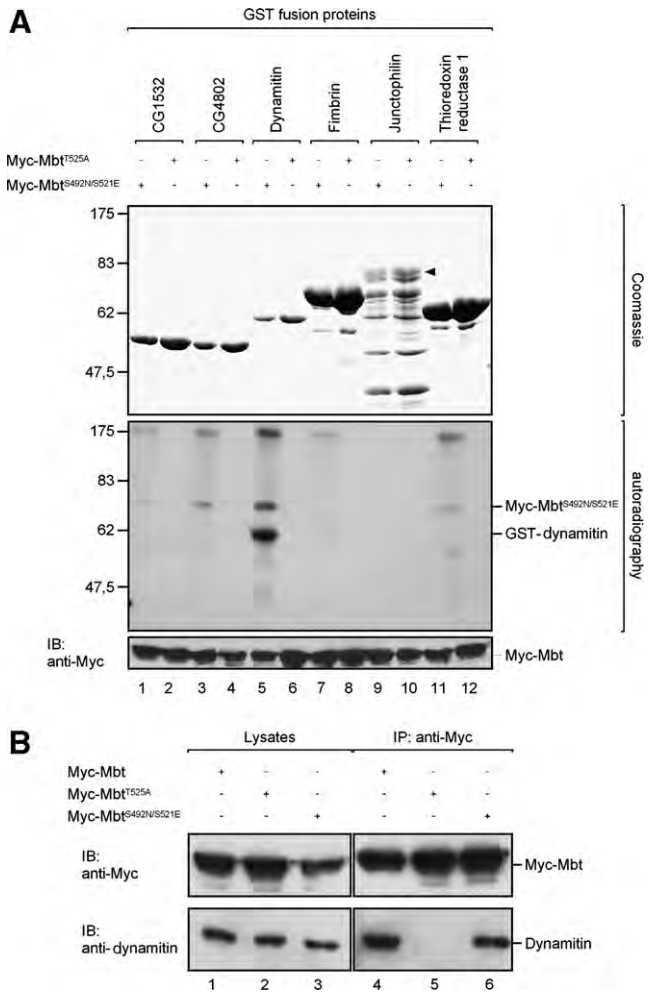


Fig. 3. Mbt phosphorylates and interacts with dynamitin. (A) The indicated candidate proteins were expressed as GST fusion proteins and used in kinase assays together with Myc–Mbt^{S492N,S521E} and Myc–Mbt^{T525A}. Reactions were subjected to SDS–PAGE and analyzed by autoradiography. The presence of the GST proteins was verified by Coomassie blue staining, and Myc–Mbt proteins were detected by Western blot analysis with the anti-Myc antibody (IB). GST–junctophilin is subjected to degradation, with the expected full-length protein being indicated by an arrowhead. Myc–Mbt^{S492N,S521E} also becomes autophosphorylated. (B) Mbt interacts with endogenous dynamitin. S2R cells stably expressing Myc–Mbt, Myc–Mbt^{T525A}, and Myc–Mbt^{S492N,S521E} were induced for expression of Mbt proteins. Immunoprecipitations from cell lysates were done with the anti-Myc antibody (IP). Immunoprecipitations were analyzed by SDS–PAGE and Western blot with the anti-Myc and anti-dynamitin antibodies (IB). The corresponding lysate controls are shown in lanes 1 to 3.

of p150^{Glued} that mediates binding to dynein motor proteins and microtubules. Both units are connected by dynamitin (p50). The main function of the dynactin complex is to allow the attachment of dynein motor proteins to their cargos, but it also contributes to the binding of dynein to centrosomes, the nucleus, and the cell cortex [21]. Dynactin acts also independently of dynein to anchor microtubules at centrosomes, and p150^{Glued} is required to organize radial microtubule arrays [22,23]. Components of the dynactin complex have also been implicated in regulation of adhesive structures of cells. Dynamitin interacts with MacMARCKS (macrophage-enriched myristoylated alanine-rich C kinase substrate), and both microtubules and MacMARCKS are part of the cytoskeletal constraint on β_2 integrin molecules that regulate integrin molecular mobility and activation [24]. Interference with the function of *Drosophila* p150^{Glued} induces defects in the organization of adherens junctions during morphogenesis of photoreceptor cells and mislocalization of armadillo (β-catenin) as a central component of the cadherin–catenin complex [25]. How p150^{Glued} exerts its effect on adherens junctions is not known; however, there are some striking similarities to the adherens junction defects observed in *mbt* mutant photoreceptor cells [7]. Therefore, it is tempting to speculate that Mbt and dynactin act at least in part synergistically to control adherens junction organization during photoreceptor cell morphogenesis. The phosphorylation of dynamitin by Mbt might control assembly, localization, or function of the dynactin complex at adherens junctions. Further studies will focus on the identification of the phosphorylation sites to investigate their physiological role in developmental processes.

Acknowledgments

A. de Marco is gratefully acknowledged for the kind gift of the pETM30 plasmid. This work was funded by grants from the Deutsche Forschungsgemeinschaft to T.R. and U.F. (SFB487 and SFB581).

References

[1] K. Burridge, K. Wennerberg, Rho and Rac take center stage, Cell 116 (2004) 167–179.
[2] A.B. Jaffe, A. Hall, Rho GTPases: Biochemistry and biology, Annu. Rev. Cell Dev. Biol. 21 (2005) 247–269.
[3] G.M. Bokoch, Biology of the p21-activated kinases, Annu. Rev. Biochem. 27 (2003) 743–781.
[4] Z.S. Zhao, E. Manser, PAK and other rho-associated kinases: Effectors with surprisingly diverse mechanisms of regulation, Biochem. J. 386 (2005) 201–214.
[5] H. Hing, J. Xiao, N. Harden, L. Lim, S.L. Zipursky, Pak functions downstream of Dock to regulate photoreceptor axon guidance in *Drosophila*, Cell 97 (1999) 853–863.
[6] T.P. Newsome, S. Schmidt, G. Dietzl, K. Keleman, B. Asling, A. Debant, B.J. Dickson, Trio combines with Dock to regulate Pak activity during photoreceptor axon pathfinding in *Drosophila*, Cell 101 (2000) 283–294.
[7] D. Schneeberger, T. Raabe, Mbt, a *Drosophila* PAK protein, combines with Cdc42 to regulate photoreceptor cell morphogenesis, Development 130 (2003) 427–437.
[8] M.G. Callow, F. Clairvoyant, S. Zhu, B. Schryver, D.B. Whyte, J.R. Bischoff, B. Jallal, T. Smeal, Requirement for PAK4 in the anchorage-independent growth of human cancer cell lines, J. Biol. Chem. 277 (2002) 550–558.
[9] J. Qu, M.S. Cammarano, Q. Shi, K.C. Ha, P. de Lanerolle, A. Minden, Activated PAK4 regulates cell adhesion and anchorage-independent growth, Mol. Cell. Biol. 21 (2001) 3523–3533.
[10] S. Faure, J. Cau, P. de Santa Barbara, S. Bigou, Q. Ge, C. Delsert, N. Morin, Xenopus p21-activated kinase 5 regulates blastomeres' adhesive properties during convergent extension movements, Dev. Biol. 277 (2005) 472–492.

[11] R. Kumar, A.E. Gururaj, C.J. Barnes, p21-Activated kinases in cancer, Nat. Rev. Cancer 6 (2006) 459–471.
[12] C. Dan, A. Kelly, O. Bernard, A. Minden, Cytoskeletal changes regulated by the PAK4 serine/threonine kinase are mediated by LIM kinase 1 and cofilin, J. Biol. Chem. 276 (2001) 32115–32121.
[13] J. Soosairajah, S. Maiti, O. Wiggan, P. Sarmiere, N. Moussi, B. Sarcevic, R. Sampath, J.R. Bamburg, O. Bernard, Interplay between components of a novel LIM kinase–slingshot phosphatase complex regulates cofilin, EMBO J. 24 (2005) 473–486.
[14] N. Menzel, D. Schneeberger, T. Raabe, The *Drosophila* p21 activated kinase Mbt regulates the actin cytoskeleton and adherens junctions to control photoreceptor cell morphogenesis, Mech. Dev. 124 (2007) 78–90.
[15] S-I. Yanagawa, J-S. Lee, A. Ishimoto, Identification and characterization of a novel line of *Drosophila* Schneider S2 cells that respond to wingless signaling, J. Biol. Chem. 273 (1998) 32353–32359.
[16] R.F. Colman, Affinity labeling of purine nucleotide sites in proteins, Annu. Rev. Biochem. 52 (1983) 67–91.
[17] P. Cohen, A. Knebel, KESTREL: A powerful method for identifying the physiological substrates of protein kinases, Biochem. J. 393 (2006) 1–6.
[18] A. Knebel, N. Morrice, P. Cohen, A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta, EMBO J. 20 (2001) 4360–4369.
[19] K. Shah, Y. Liu, C. Deirmengian, K.M. Shokat, Engineering unnatural nucleotide specificity for *Rous sarcoma* virus tyrosine kinase to uniquely label its direct substrates, Proc. Natl. Acad. Sci. USA 94 (1997) 3565–3570.
[20] M.W. Mayhew, D.J. Webb, M. Kovalenko, L. Whitmore, J.W. Fox, A.F. Horwitz, Identification of protein networks associated with the PAK1–βPIX–GIT1–paxillin signaling complex by mass spectrometry, J. Proteome Res. 5 (2006) 2417–2423.
[21] T.A. Schroer, Dynactin, Annu. Rev. Cell Dev. Biol. 20 (2004) 759–779.
[22] H. Kim, S-C. Ling, G.C. Rogers, C. Kural, P.R. Selvin, S.L. Rogers, V.I. Gelfand, Microtubule binding by dynactin is required for microtubule organization but not cargo transport, J. Cell Biol. 176 (2007) 641–651.
[23] N.J. Quintyne, T.A. Schroer, Distinct cell cycle-dependent roles for dynactin and dynein at centrosomes, J. Cell Biol. 159 (2002) 245–254.
[24] T. Jin, J. Li, Dynamitin controls β_2 integrin avidity by modulating cytoskeletal constraint on integrin molecules, J. Biol. Chem. 277 (2002) 32963–32969.
[25] S.S. Fan, Dynactin affects extension and assembly of adherens junctions in *Drosophila* photoreceptor development, J. Biomed. Sci. 11 (2004) 362–369.

Curriculum Vitae

Personal Details

First Name, Name:	Ashwin, Chari
Date of Birth:	26. December 1979
Place of Birth:	Bangalore (Indien)
Citizenship:	Switzerland
Marital status:	not married

Scientific Education

Jul. 1999:	Eidgenössische Matura Typus B, Kantonsschule Baden, Switzerland
Oct. 1999- Oct. 2004:	Studies in Biology, Specialization in Biochemistry and Molecular Biology/ Biophysics, ETH Zürich, Switzerland
Jan. 2003- Aug. 2003:	Diploma Thesis at the Paul Scherrer Institut/ ETH Zürich: „Expression, Characterisation and Structural Studies of the snRNP Assembly Factors pICln and the Protein Arginine Methyltransferase PRMT5, Supervisor Dr. Christian Kambach
Oct. 2004:	Diploma at the ETH Zürich (Dipl. natw. ETH)
Oct. 2004- present:	PhD student at the Chair of Biochemistry, AG Fischer, Thesis title: „The Reaction Mechanism of Cellular U snRNP Assembly“

Würzburg, 2009

Ashwin Chari

Scientific Conferences

1. **12th Annual Meeting of the RNA Society, Madison, USA, 2007**

   "A Unique Chaperone System Prevents Sm Protein Aggregation and Ensures Proper U snRNA Assembly", poster presentation
   Chari A, Klingenhäger M, Fischer U

   "Evolution of an RNP assembly system: A minimal SMN-complex facilitates formation of U snRNPs in Drosophila melanogaster", poster presentation
   Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U

2. **13th Annual Meeting of the RNA Society, Berlin, Germany, 2008**

   „A Coupled Chaperone-Enzyme System Facilitates U snRNP Assembly by a Mechanism Similar to DNA Clamp Loading", oral presentation
   Chari A, Golas MM, Neuenkirchen N, Klingenhäger M, Sander B, Englbrecht C, Stark H, Fischer U

   "LSm1-7 complex binds to essential translation and replication signals in Brome Mosaic Virus RNAs", oral presentation
   Galao RP, Alves-Rodrigues I, Chari A, Fischer U, Díez J

   "IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disease distal SMA type 1 (DSMA1)", poster presentation
   Guether UP, Handoko L, Laggerbauer B, Schuelke M, Chari A, von Au K, Fischer U

   "The 68 kDa subunit of cleavage factor I is symmetrically dimethylated by PRMT5", poster presentation
   Martin G, Ostareck-Lederer A, Chari A, Neuenkirchen N, Fischer U, Keller W

3. **Swiss RNA Workshop, Bern, Switzerland, 2009**

„An Assembly Chaperone Collaborates with the SMN-Complex to Generate Spliceosomal snRNPs“, oral presentation

4. **14th Annual Meeting of the RNA Society, Madison, USA, 2009**

„Reconstitution of the U snRNP Assembly Machinery: Insights into Mechanisms and Spinal Muscular Atrophy“, oral presentation

List of Publications

Original Manuscripts:

1. **Toward an assembly line for U7 snRNPs: interactions of U7-specific Lsm proteins with PRMT5 and SMN complexes.**

   Azzouz TN, Pillai RS, Dapp C, Chari A, Meister G, Kambach C, Fischer U, Schumperli D: **J Biol Chem 2005**, 280:34435-34440.

   Thesis author´s contribution:

Conception:	10 %
Experimental contributions:	10 %
Formulation of results:	10 %

2. **Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization.**

   Grimmler M, Otter S, Peter C, Müller F, Chari A, Fischer U: **Hum Mol Genet 2005**, 14:3099-3111.

   Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formulation of results:	30 %

3. **A comprehensive interaction map of the human survival of motor neuron (SMN) complex.**

   Otter S, Grimmler M, Neuenkirchen N, Chari A, Sickmann A, Fischer U**: J Biol Chem 2007,** 282:5825-5833.

   Thesis author´s contribution:

Conception:	30 %
Experimental contribution:	10 %
Formulation of results:	30 %

4. **A 5´-fluorosulfonylbenzoyladenosine-based method to identify physiological substrates of a Drosophila p21-activated kinase.**

   Menzel N, Chari A, Fischer U, Linder M, Raabe T: **Anal Biochem 2007**, **368**:178-184.

   Thesis author´s contribution:

Conception:	20 %
Experimental contribution:	20 %
Formulation of Results:	20 %

5. **Evolution of an RNP assembly system: A minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster.**

   Kroiss M, Schultz J, Wiesner J, Chari A, Sickmann A, Fischer U: **Proc Natl Acad Sci 2008**, 105:10045–10050.

   Thesis author´s contribution:

Conception:	30 %
Experimental contribution:	10 %
Formulation of results:	10 %

6. **An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs.**

   Chari A, Golas MM, Klingenhager M, Neuenkirchen N, Sander B, Englbrecht C, Sickmann A, Stark H, Fischer U: **Cell 2008**, 135:497-509.

   Thesis author´s contribution:

Conception:	60 %
Experimental contribution:	80 %
Formulation of results:	80 %

7. **Ighmbp2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1.**

Günther UP, Handoko L, Laggerbauer B, Chari A, Sickmann A, Gehring N, von Au K, Schülke D, Fischer U: **Hum Mol Genet 2009**, 18: 1288-1300.

Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formlation of results:	20 %

8. **Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates**.

Scheller N, Mina LC, Galao RP, Chari A, Gimenez-Barcons M, Noueiry A, Fischer U, Meyerhans A, Diez J: **Proc Natl Acad Sci 2009**, Epub

Thesis author´s contribution:

Conception:	10 %
Experimental contribution:	10 %
Formulation of results:	10 %

9. **LSm1-7 complexes bind to specific sites in viral RNA genomes and regulate their translation and replication.**

Galao RP, Chari A, Alves-Rodriguez I, Lobao D, Mas A, Kambach C, Fischer U, Diez J: **RNA**, in revision

Thesis author´s contribution:

Conception:	40 %
Experimental contribution:	10 %
Formulation of results:	20 %

10. **Arginine methylation in subunits of mammalian pre-mRNA cleavage factor I**

Martin G, Ostareck-Lederer A, Chari A, Neuenkirchen N, Dettwiler S, Blank D, Rüegsegger U, Fischer U, Keller W: **RNA**, in revision

Thesis author´s contribution:

Conception:	20 %
Experimental contribution:	10 %
Formulation of results:	10 %

Reviews:

1. **Spinal muscular atrophy: the RNP connection.**

Eggert C, Chari A, Laggerbauer B, Fischer U: **Trends Mol Med 2006**, 12:113-121

2. **Wenn Muskeln die Nerven verlieren**

Fischer U, Kroiss M, Chari A: **Biospektrum 2007**, 13. Jahrgang, Ausgabe 06.07

3. **Deciphering the assembly pathway of Sm-class U snRNPs**

Neuenkirchen N, Chari A, Fischer U: **FEBS Lett 2008**, 582:1997-2003

4. **The role of RNP biogenesis in Spinal Muscular Atrophy**

Chari A, Paknia E, Fischer U: **Curr Opin Cell Biol 2009**, 21:387-393

5. **Molekulares Bodybuilding: Wie Zellen RNA-Proteinkomplexe herstellen**

Chari A, Fischer U: **Biospektrum 2009**, 15. Jahrgang, in press

Acknowledgements

First and foremost I would like to thank Prof. Dr. U. Fischer for his encouragement and support during the course of this thesis. He encouraged me at all times and the present work is a result of the many constructive discussions we have led over the last four and a half years.

I would also like to thank Prof. Dr. U. Scheer for readily agreeing to act as a supervisor for my thesis.

I thank the many colleagues, with whom I have been engaged in various projects both outside and inside the Fischer lab. Although I will not mention each of you by name you know who is meant and my cordial thanks goes out to you for the many successful collaborations.

I have been very fortunate to supervise three talented diploma students in the course of my thesis. Lasse, Clemens, and Patrick thank you for your engagement. In the later stages of this work I have greatly benefited of the expert technical assistance of Jürgen Ohmer. Thanks Jürgen for your engagement. A special thanks goes out to my colleagues Bernhard and Basti for the many late night discussions both scientific and not so scientific. It was fun and sometimes the only thing that kept me going, especially in the early stages of this thesis.

My friends Nicolas and Reinhard, thank you for the many late nights out. It was always fun and I would do it anytime again. My friends back home in Switzerland Sonja, Mäsi and Andy thanks for sticking with me, even though I have not been around very much. I always enjoy coming back and catching up with you guys.

My parents I thank for all the assistance during my entire life. None of this would have been possible without you; I hope I have made you proud!

Lastly, thank you my dear Elham for all the support, patience and love. Living with you makes me appreaciate each and every moment, and detail so much more.