Supplementary information

Functional organization of the Sm core in the crystal structure of human U1 snRNP

Gert Weber[1,2], Simon Trowitzsch[1,#], Berthold Kastner[1],
Reinhard Lührmann[1,*], Markus C. Wahl[1,2,*]

Max-Planck-Institut für biophysikalische Chemie
[1] Zelluläre Biochemie/Makromolekulare Röntgenkristallographie
Am Faßberg 11, D-37077 Göttingen, Germany

[2] Freie Universität Berlin
AG Strukturbiochemie
Takustr. 6, D-14195 Berlin, Germany

Running title

Structure of human U1 snRNP

[#] Present address: EMBL Grenoble, BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[*] To whom correspondence should be addressed.
RL: Tel.: +49 (0)551 201-1407; Fax: +49 (0)551 201-1197; eMail: Reinhard.Luehrmann@mpi-bpc.mpg.de
MCW: Tel.: +49 (0)30 838-53456; Fax: +49 (0)30 838-56981; eMail: mwahl@chemie.fu-berlin.de

Supplementary materials and methods

Purification and crystallization of human U1 snRNP

Total snRNPs were isolated from HeLa cell nuclear extract by immunoaffinity purification as described (Bach et al, 1990) and applied on a 10-30 % low-salt glycerol gradient (Kastner & Lührmann, 1999). The 12S fractions of the gradient containing U1 and U2 snRNPs were diluted to a NaCl concentration of 100 mM, loaded on a Heparin sepharose column (GE Healthcare) in buffer A (20 mM Tris-HCl, pH 7.9, 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT) and eluted in a linear gradient from 100 mM to 800 mM NaCl. U1 and U2 snRNPs eluted from the column well separated at 450 mM and 250 mM NaCl, respectively. The U1 snRNP fractions were buffer-exchanged to buffer B (10 mM Tris-HCl, pH 7.9, 100 mM NaCl, 1 mM DTT) and concentrated by ultracentrifugation in a S100-AT rotor (Sorvall).

U1 snRNP (15 - 20 mg/ml) was mixed with the DNA nonamer 5'-AGGTAAGTA-3' (IBA BioTAGnology) in a 1:2 molar ratio and trace amounts of chymotrypsin were added (between 1:200 and 1:10000 w/w chymotrypsin:U1). Crystals were obtained by the sitting drop vapor diffusion method at 277 K within several days to weeks, depending on the amount of chymotrypsin used. The reservoir contained 100 mM Na-citrate, pH 5.6, 8 – 10 % PEG 4000, 100 mM NaCl, 2 mM EDTA and 1-2 % of the detergent Anapoe X-305 (Anatrace). For data collection at cryogenic temperatures, crystals were transferred to reservoir supplemented with 15 – 17.5 % propylene glycol. Heavy atom derivatives were obtained by soaking crystals in 100 mM Na-citrate, pH 5.6, 8 – 10 % PEG 4000, 100 mM NaCl and variable amounts of heavy atom substances. Derivative crystals were cryo-protected by adding 15 – 17.5 % propylene glycol.

Data collection and processing

Native datasets were collected at 100 K on the X06SA beamline of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland) using a Pilatus 6M pixel detector (Dectris; Table I). Data were indexed, integrated and scaled using the XDS software package (Kabsch,

1993). Derivative datasets were collected at 100 K on the X10SA beamline using a MAR225 CCD detector (Table I). All derivative data were indexed, integrated and scaled using the HKL software package (Minor et al, 2006). Native and derivative data were scaled together with the program SCALEIT of the CCP4 software suite (Winn, 2003).

Structure solution

The structure of native human U1 snRNP was solved by a combined molecular replacement and multiple isomorphous replacement strategy. A model of a heptameric ring of Sm proteins was constructed from the crystal structures of the human Sm protein dimers D1-D2 and D3-B (Kambach et al, 1999) and from homology models of the Sm proteins E, F and G. The crystal structure of the yeast Sm-F protein (PDB ID 1N9R) (Collins et al, 2003) and the C_α-coordinates of the respective proteins of recombinant U1 snRNP (Pomeranz Krummel et al, 2009) served as templates for modeling of the Sm proteins E, F and G by the HHpred server (Soding et al, 2005). *Via* molecular replacement searches with PHASER (McCoy, 2007) we positioned two Sm ring models in an asymmetric unit. Subsequently, two copies of the N-terminal RRM of U1-A in complex with the tip of SL2 of U1 snRNA (Oubridge et al, 1994) could also be localized with the initially placed Sm ring models held fixed. The non-crystallographic symmetry of the molecular replacement solution was validated by a self-rotation function of the diffraction data (Vagin & Teplyakov, 2010).

Phases from the above molecular replacement solution were used to calculate difference Fourier maps, in which heavy atom sites could be located. The sites were refined and used for phasing with autoSHARP (Vonrhein et al, 2007) followed by solvent flattening, histogram matching and NCS-averaging with PARROT employing the PDB-structure 1Z1G as a reference (Cowtan & Zhang, 1999). Molecular replacement and heavy atom phases were combined using PHENIX.REFINE (Adams et al, 2002). The resulting experimentally phased and phase-combined maps were of sufficient quality to place a homology model of the U1-70K RRM based on the RRM of the arginine/serine-rich splicing factor 10 (PDB ID 2CQC) and portions of the structural model of U1 snRNA obtained by solution studies (Krol et al,

1990) (Supplementary figure S2). Additional model building was guided by the recently published backbone model of a recombinant U1 snRNP (Pomeranz Krummel et al, 2009).

Model building and refinement

The model of U1 snRNP was refined by alternating rounds of automated refinement with PHENIX.REFINE and manual model building with COOT (Emsley & Cowtan, 2004). α-helices not contained in the initial model were localized using an automatic routine in PHENIX.FIND_HELICES_STRANDS (Adams et al, 2002). During initial refinement steps, each protein chain was treated as a rigid body and the RNA was divided into five rigid bodies (SL1, SL2, SL3, SL4 and the remainder of the U1 snRNA). During final model building stages, coordinates and torsion angles were refined with tight restraints on geometry and NCS between corresponding portions of the two particles in an asymmetric unit using PHENIX.REFINE. One B-factor was assigned to each amino acid residue or nucleotide. In addition, backbone hydrogen-bonding in α-helices, β-sheets, the Watson-Crick interface of canonical base pairs, the pairing of G•U wobble pairs, the G16•G118 pair across the four-way junction and in the Watson-Crick-like protein-RNA contacts of the Sm site nucleotides were restrained. The final rounds of refinement were carried out with CNS employing deformable elastic networks (DEN) (Schroder et al, 2010) and using the PHENIX-refined U1 snRNP model as a reference model. After a two-dimensional grid search as described in (Schroder et al, 2010), γ (the parameter balancing the influences of the diffraction data and the reference model) and w_{DEN} (the weighting factor for the DEN potentials) were set to 0.6 and 40, respectively. The same set of reflections was maintained for cross validation during the change of the refinement programs.

Author contributions

RL and BK conceived the project; GW and ST purified and crystallized the particle, produced heavy atom derivatives and collected diffraction data; GW and MCW carried out

the crystallographic analyses; all authors participated in data interpretation and writing of the paper; RL and MCW acted as project leaders.

Supplementary references

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D* **58:** 1948-1954

Avis JM, Allain FH, Howe PW, Varani G, Nagai K, Neuhaus D (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J Mol Biol* **257:** 398-411

Bach M, Bringmann P, Lührmann R (1990) Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. *Methods Enzymol* **181:** 232-257

Collins BM, Cubeddu L, Naidoo N, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, Mabbutt BC (2003) Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF. *J Biol Chem* **278:** 17291-17298

Cowtan KD, Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* **72:** 245-270

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60:** 2126-2132

Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26:** 795-800

Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, Lührmann R, Li J, 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, Lührmann R (1999) Purification of U small nuclear ribonucleoprotein particles. *Methods Mol Biol* **118:** 289-298

Keel AY, Rambo RP, Batey RT, Kieft JS (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure* **15:** 761-772

Krol A, Westhof E, Bach M, Lührmann R, Ebel JP, Carbon P (1990) Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res* **18:** 3803-3811

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948

McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D* **63:** 32-41

Minor W, Cymborowski M, Otwinowski Z, Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D* **62:** 859-866

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372:** 432-438

Pomeranz Krummel DA, Oubridge C, Leung AK, Li J, Nagai K (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458:** 475-480

Schroder GF, Levitt M, Brunger AT (2010) Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464:** 1218-1222

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33:** W244-248

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D* **66:** 22-25

Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364:** 215-230

Winn MD (2003) An overview of the CCP4 project in protein crystallography: an example of a collaborative project. *J Synchr Radiat* **10:** 23-25

Supplementary tables

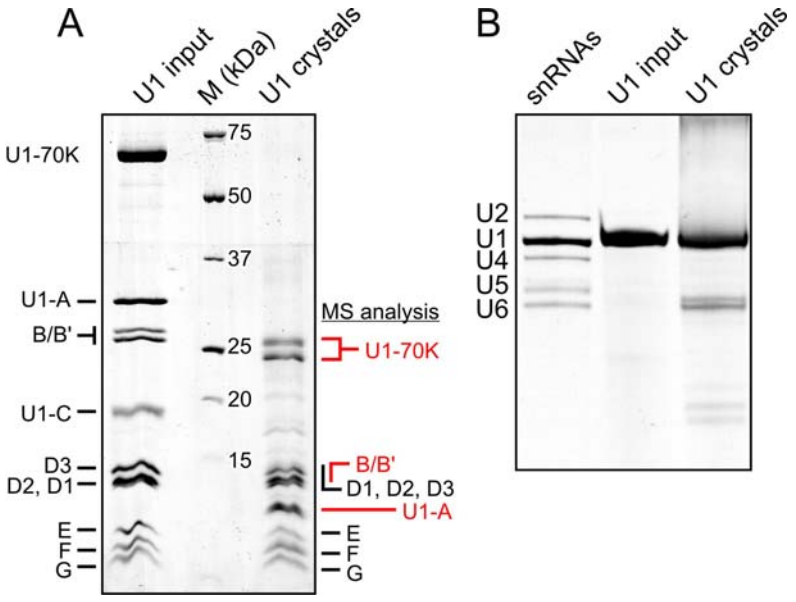
Supplementary table SI U1 snRNP components contained in the final model

U1 snRNP component	Residues	Residue ranges complex 1	Residue ranges complex 2
U1 snRNA	164	1-164	1-164
U1-A	282	5-112	2-114
U1-70K (Isoform 1)	437	34-183	34-181
U1-C	159	-	-
SmB	231	6-86	5-53, 64-90
SmD3	126	5-84	3-94
SmD2	118	2-112	2-76, 90-117
SmD1	119	1-85	1-85
SmE	92	7-90	1-89
SmF	86	2-86	2-75
SmG	76	1-76	1-76

Supplementary figures

Supplementary figure S1

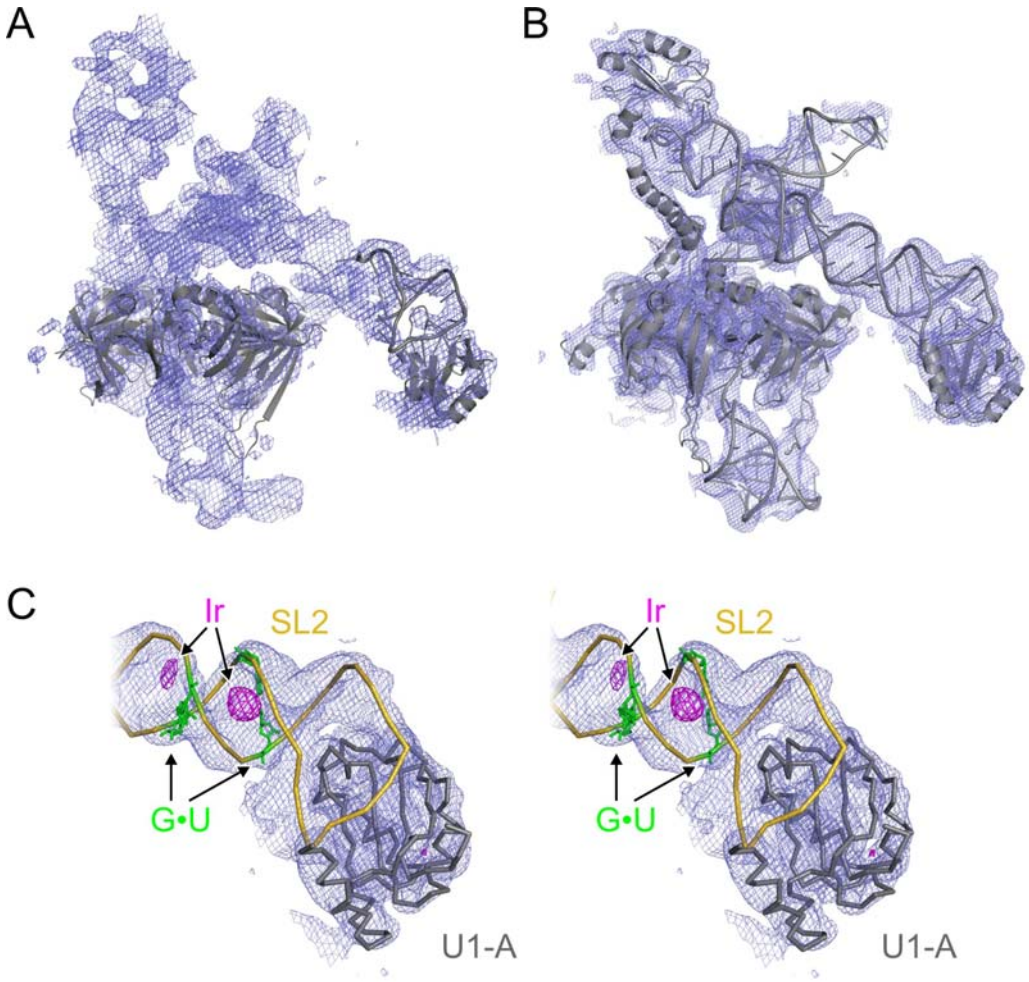
Contents of the U1 snRNP crystals. (**A**) SDS PAGE analysis of U1 snRNP crystals. M - molecular weight marker; marker sizes in kDa are indicated. U1 input - protein composition of natively purified human U1 snRNP. Protein bands are labeled at the left. U1 crystals - protein composition of U1 snRNP crystals obtained upon limited proteolysis with chymotrypsin. Protein bands as identified by mass spectrometric fingerprinting are labeled at the right ("MS analysis"). U1-70K, U1-A and B/B' (red) have been truncated by chymotrypsin (indicated in red). U1-C was not found in dissolved crystals. In contrast, other extensions, such as the RG-rich regions in the C-termini of Sm proteins D1 and D3, appear to remain intact upon limited proteolysis. (**B**) Urea PAGE analysis of U1 snRNP crystals. snRNAs – U1, U2, U4, U5 and U6 snRNA references. U1 input - RNA content of natively purified human U1 snRNP. U1 crystals - RNA content of U1 snRNP crystals obtained in the presence of chymotrypsin. Only minor amounts of degraded bands are seen.



Supplementary figure S1

Supplementary figure S2

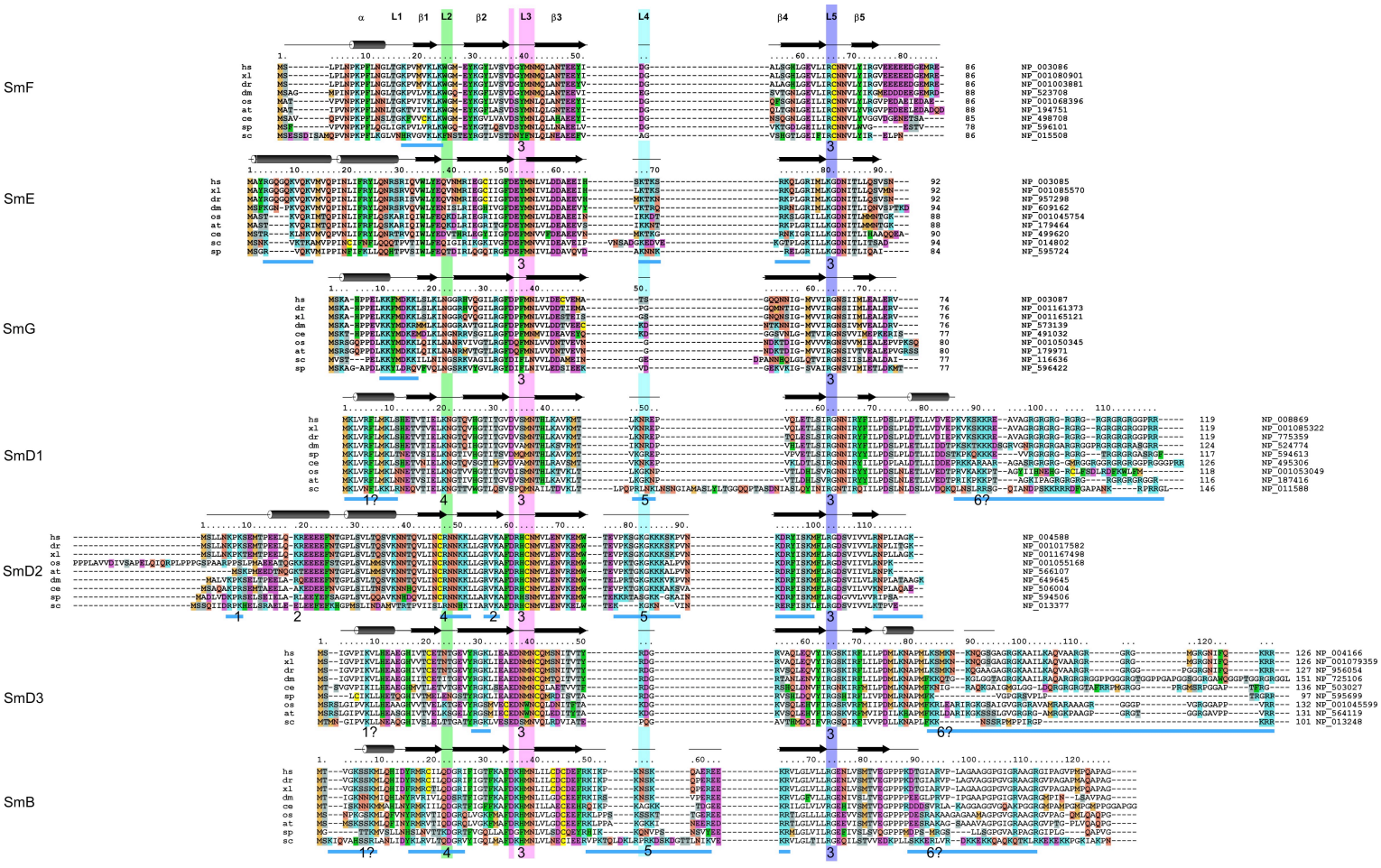
Electron density. (**A**) Experimental MIRAS electron density map contoured at the 1σ level (blue mesh) with a model of the seven-membered Sm ring and the U1-A N-terminal RRM in complex with the tip of SL2 (gray ribbons) positioned in a molecular replacement search. (**B**) Final $2F_o$-F_c phase-combined electron density map contoured at the 1σ level (blue mesh) with the final model of native U1 snRNP (gray ribbons). (**C**) Stereo plot showing the final $2F_o$-F_c phase-combined electron density map at the 1σ level (blue mesh) and an isomorphous difference Fourier map at the 5σ level (magenta mesh) indicating iridium hexammine ions bound at the SL2 region. U1-A - gray; RNA - gold; G79•U60 (left) and G56•U83 (right) wobble base pairs known to form iridium hexammine binding sites (Keel et al, 2007) are shown in green. In all panels the view is the same as in Figure 1A.

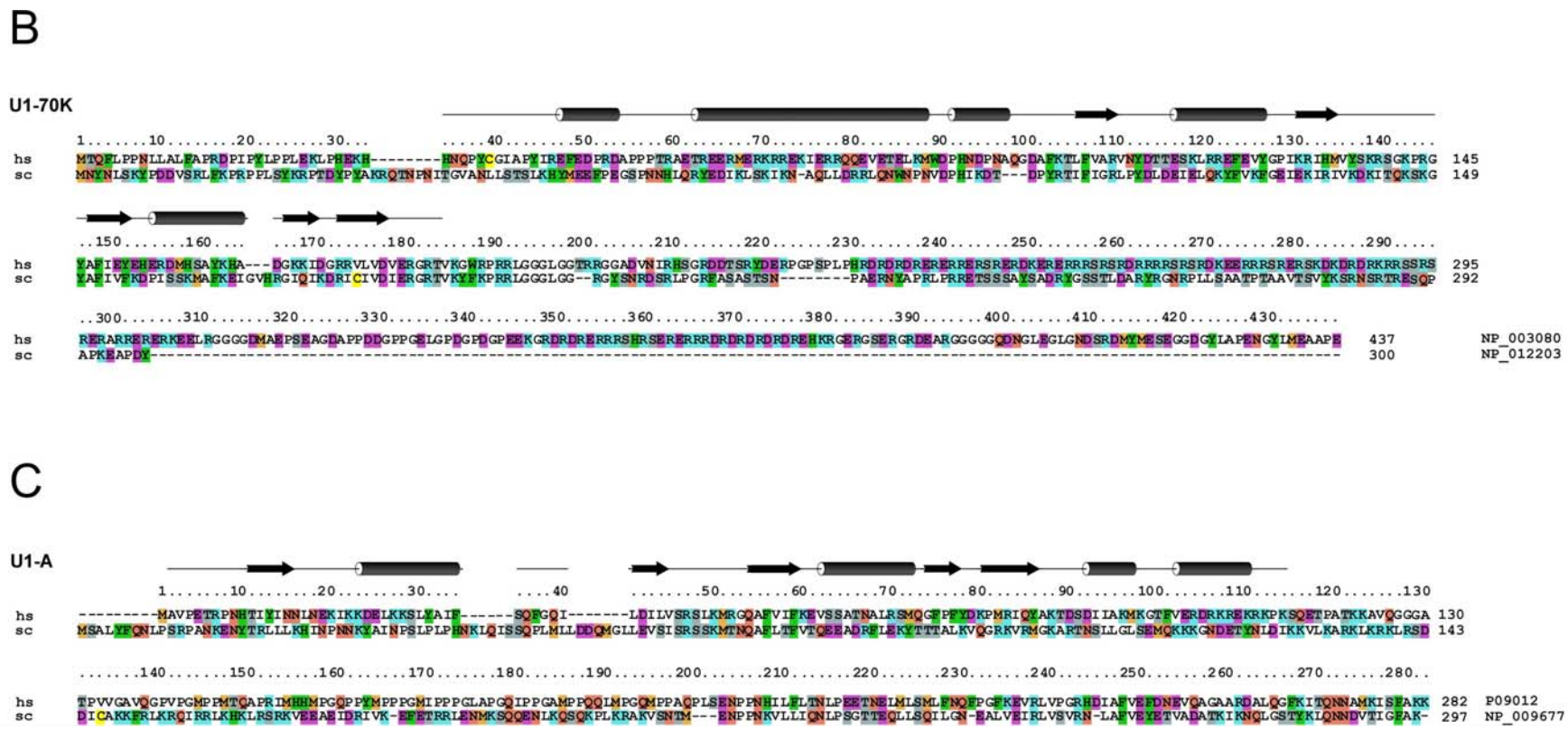


Supplementary figure S2

Supplementary figure S3

Multiple sequence alignments. (**A**) For each of the seven Sm proteins present in the U1 snRNP, the sequences from eight organisms are aligned to the human sequence (at the top of each block). All sequences were retrieved *via* NCBI Entrez Protein (http://www.ncbi.nlm.nih.gov/protein). Alignments were done by using the clustalX2 program (Larkin et al, 2007) and were locally adjusted by visual inspection. The output order of the clustalX2 program, reflecting the degree of conservation, was maintained in the figure. To highlight groups of chemically related amino acids, the following color code was used: basic amino acids (R and K) - blue; acidic amino acids (E, D) - purple; amino acids with amide groups (N, Q) - orange; aromatic amino acids (H, F, Y, W) - green; polar amino acids (S, T) - gray; methionine (M) - dark yellow; cysteine (C) - yellow. Above each sequence block, the numberings of the human proteins are given. Above the numberings, the secondary structure elements as found in the U1 snRNP crystal structure are indicated by cylinders (α-helices) and arrows (ß-strands). The line connecting these elements marks the residue ranges contained in the present structural model. Colored columns indicate the positions of the loops important for U1 snRNA binding. For loops L2 and L4, the amino acids at the tips of the loops in the crystal structure are taken as loop positions. Below each sequence block, blue bars mark conserved patches of basic amino acids. Numbers below the blocks denote elements seen in or deduced from (with question mark) the crystal structure to be involved in snRNA binding (see also Figure 3A of the main text). The abbreviations used for the organisms are defined at the bottom left. (**B**), (**C**) Alignment of human and yeast U1-70K and U1-A sequences. Color coding of the amino acids is the same as in (A). Secondary structure elements contained in the present model are shown above the alignments as in (A). Portions of the proteins contained in the present structure are indicated by a line connecting these elements above the alignment.

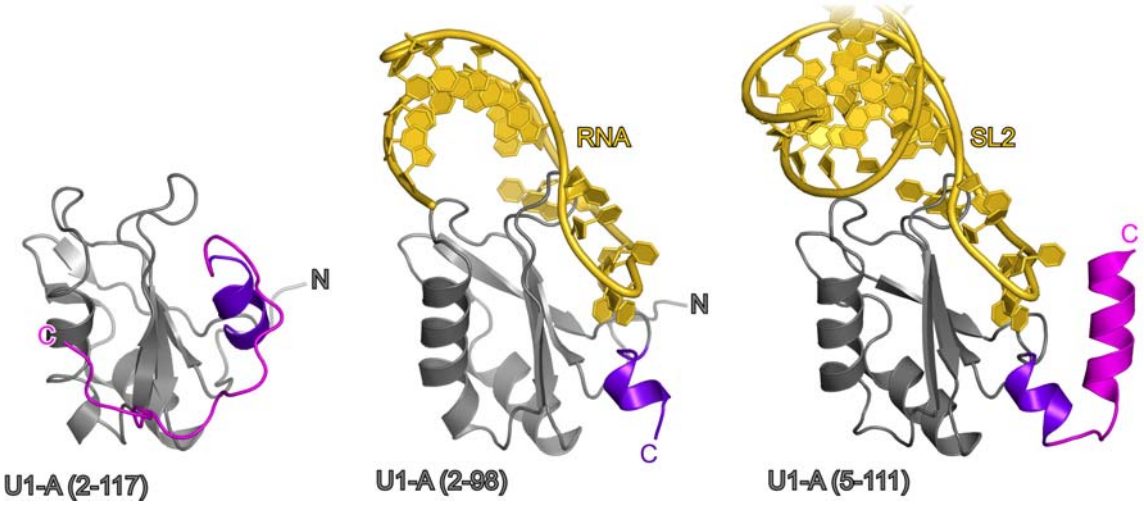

SmF
SmE
SmG
SmD1
SmD2
SmD3
SmB



Supplementary figure S3

Supplementary figure S4

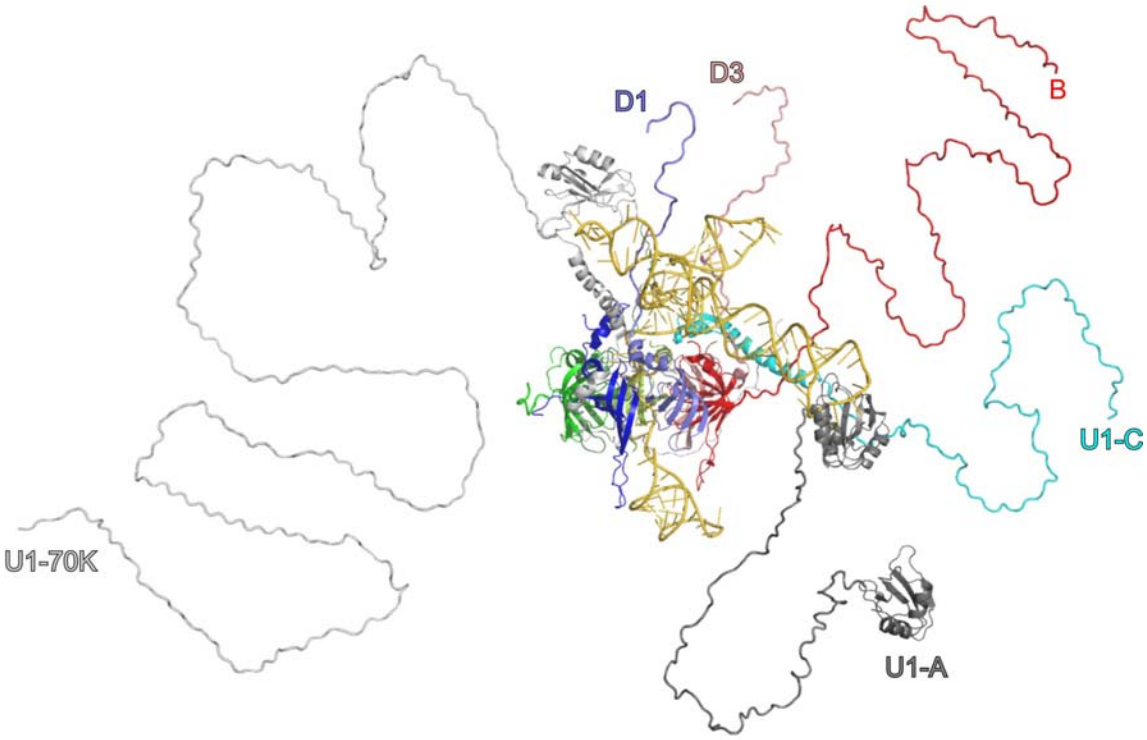
C-terminal extension of the U1-A N-terminal RRM. Ribbon plot of the N-terminal RRM of U1A in isolation (left; PDB ID 1FHT (Avis et al, 1996)); the U1A N-terminal RRM bound to a mimic of SL2 (center; PDB ID 1URN (Oubridge et al, 1994)); and in the present U1 snRNP crystal structure (right). U1-A - gray; RNA - gold. A short C-terminal helix covers the RNA binding platform in the isolated protein and is relocated upon RNA binding (purple). More distal C-terminal elements are unstructured in isolation (left) and form an additional helix that clamps the RNA in the present U1 snRNP structure (magenta). This distal region is not contained in the U1A construct used for crystal structure analysis in complex with a SL2 mimic (center). The residue ranges of U1A contained in the models are indicated in parentheses. Rotated 90° about the Y-axis compared to Figure 1A. The requirement for alleviation of an auto-inhibited conformation and the folding upon binding transition in the protein upon complex formation will increase the specificity of the interaction.



Supplementary figure S4

Supplementary figure S5

Model of the complete U1 snRNP. Ribbon plot of U1 snRNP after modeling portions that are missing in the present crystal structure. The 5'SS and U1-C have been modeled based on superimposition of the recombinant U1 snRNP backbone model (PDB ID 3CW1) (Pomeranz Krummel et al, 2009). Presumed intrinsically unstructured elements and the flexibly linked C-terminal RRM of U1-A are shown in arbitrary conformations. Proteins with flexible peripheral elements are labeled. Coloring and view are the same as in Figure 1A.



Supplementary figure S5