Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary crystallographic data of the m_3G cap-binding domain of human snRNP import factor snurportin 1

Anja Strasser,[a] Achim Dickmanns,[a] Ute Schmidt,[b]‡ Elke Penka,[b] Henning Urlaub,[b] Mitsuo Sekine,[c] Reinhard Lührmann[b] and Ralf Ficner[a]*

[a]Abteilung Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität, Justus-von-Liebig Weg 11, 37077 Göttingen, Germany, [b]Abteilung für Zelluläre Biochemie, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, 37077 Göttingen, Germany, and [c]Department of Life Science, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8501, Japan

‡ Present address: Abteilung Molekulare Genetik, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

Correspondence e-mail: rficner@gwdg.de

Received 8 April 2004
Accepted 23 June 2004

The nuclear import of spliceosomal UsnRNPs is mediated by the transport adaptor snurportin 1 (SPN1), which specifically recognizes the 2,2,7-trimethylguanosine (m_3G) cap at the 5′ end of UsnRNAs. Human SPN1 was overexpressed as a GST-fusion protein in *Escherichia coli* and purified to homogeneity. Since full-length SPN1 did not crystallize, limited proteolysis experiments were performed and stable digestion products were analyzed for functionality with respect to m_3G cap-binding activity and subsequently used for crystallization trials. Well diffracting single crystals of a truncated SPN1 m_3G cap-binding domain (residues 79–300) were obtained after two rounds of seeding. The crystals belong to space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 57.47$, $c = 130.09$ Å, $\alpha = \beta = \gamma = 90^\circ$. Crystals contain one molecule in the asymmetric unit and diffract to a resolution limit of 2.9 Å.

1. Introduction

Eukaryotic cells exhibit a high degree of organization and stringency with respect to control and regulatory mechanisms as a result of compartmentalization of their cells into mitochondrial, cytoplasmic, endoplasmatic reticulum and nuclear space. These compartments are separated from each other by a monolayer or bilayer of membranes. The crosstalk between the compartments is mediated by transport machineries traversing the membranes and, in the case of the nucleus, permitting the shuttling of transport factors back and forth between the cytoplasm and nucleoplasm. The exchange between the compartments is mediated by nuclear pore complexes (NPCs), large supramolecular structures of 125 MDa (Reichelt *et al.*, 1990) that span the inner and outer membrane, comprising approximately 30 different proteins (Allen *et al.*, 2001; Cronshaw *et al.*, 2002; Rout *et al.*, 2000; Stoffler *et al.*, 2003). NPCs form a central aqueous channel that permits the exchange of ions, small molecules, polypeptides and RNA. Generally, the cargo, *i.e.* larger molecules of a molecular weight above 20–60 kDa, is not transported by itself but requires soluble transport factors to mediate the translocation or at least to increase the transfer rates. The cargo itself is recognized by nuclear transport signals, which typically consist of short stretches of amino acids that differ for import and export cargo [reviewed by Dingwall & Laskey (1991) and Gerace (1995)]. The transport factors identified so far belong to two classes. The first group, mainly nuclear export factors, are involved in the export of many RNAs (Gruter *et al.*, 1998; Herold *et al.*, 2000; Katahira *et al.*, 1999). The majority of import and export receptors belong to the second class, the importin β superfamily, and share common properties such as NPC, Ran and cargo binding (Chaillan-Huntington *et al.*, 2000; Fornerod *et al.*, 1997; Gorlich *et al.*, 1997). In some cases, the interaction with the cargo is not achieved directly but requires adaptor molecules that specifically recognize nuclear transport signals. These adaptor molecules interact *via* their N-terminal domain with the receptor importin β and *via* their C-terminal domain with the cargo, such as importin α in the case of the basic-type NLS (Gorlich *et al.*, 1994; Moroianu *et al.*, 1995; Weis *et al.*, 1995), Rip α in the case of RPA (replication protein A; Jullien *et al.*, 1999) and snurportin1 (SPN1) in the case of spliceosomal small nuclear ribonucleoprotein particles (snRNPs; Huber *et al.*, 1998). SPN1 specifically recognizes the hypermethylated 5′ cap of the uridyl-rich small nuclear RNAs (UsnRNAs) present in snRNPs (Huber *et al.*, 1998). During biogenesis of UsnRNPs, the UsnRNA is transcribed in the nucleus and then exported to the cytoplasm *via* recognition of the 7-methylguanosine (m^7G) cap by the cap-binding complex (CBC) and involvement of the export factors PHAX and CRM1 (Ohno *et al.*, 2000; Segref *et al.*, 2001). In the cytoplasm, a set of seven homologous proteins, the Sm proteins, assemble specifically with the UsnRNA, a process mediated by the SMN complex (Massenet *et al.*, 2002; Meister *et al.*, 2001). This assembly is a prerequisite for the hypermethylation of the m^7G cap to the

2,2,7-trimethylguanosine (m_3G) cap (Mattaj, 1986), which in turn is recognized by SPN1 and importin β for snRNP import into the nucleus (Huber *et al.*, 1998; Palacios *et al.*, 1997; reviewed by Will & Lührmann, 2001). SPN1 strictly discriminates between monomethylated m^7G and the hypermethylated m_3G cap RNAs (Huber *et al.*, 1998). The aim of the present work was the expression and purification of SPN1 and the crystallization of SPN1 in the presence of an m_3G cap oligonucleotide.

2. Methods, results and discussion

2.1. Cloning procedures, overexpression and purification of full-length SPN1

The coding sequence of human SPN1 (Genbank accession code AF039029) was amplified and cloned into the overexpression vector pGEX6-P3 (Amersham Biosciences, Germany) *via Nde*I and *Xho*I restriction sites, introducing an additional five amino acids at the N-terminus of the polypeptide. The clone used for overexpression was verified by sequencing. SPN1 was overexpressed as a glutathione-*S*-transferase (GST) fusion protein. In expression trials, *Escherichia coli* cells differing with respect to control of induction (LysS, LysE), codon usage and protease deficiencies were tested at different temperatures. Expression turned out to work best in *E. coli* strain BL21(DE3) RIL (Stratagene) cultivated in 2YT medium at a temperature of 291 K. Expression was induced by 0.5 m*M* IPTG at an OD_{600} of ∼0.8. After 12 h, cells were pelleted and resuspended in 1 × PBS (8 m*M* Na_2HPO_4, 2 m*M* KH_2PO_4, 128 m*M* NaCl, 2 m*M* KCl pH 7.5) containing 2.5 m*M* $MgCl_2$, 2 m*M* EDTA, 2 m*M* DTT and a protease-inhibitor cocktail (one tablet of Complete protease-inhibitor set per 50 ml buffer; Roche, Germany). Resuspended cells were lysed using a fluidizer (Microfluidics, USA) for three cycles at a pressure of 621 kPa. After centrifugation at 100 000*g* at 277 K for 1 h, the supernatant was bound to a pre-equilibrated 30 ml bed volume glutathione (GSH) Sepharose column (100 m*M* NaCl, 50 m*M* Tris–HCl pH 7.5, 2.5 m*M* $MgCl_2$, 2 m*M* EDTA, 2 m*M* DTT). Unbound protein was removed by washing with two column volumes of loading buffer. Bound GST–SPN1 was eluted using 10–15 m*M* reduced glutathione at pH 8.0. For cleavage of the GST-fusion protein, pooled fractions were incubated with 0.5 mg of PreScission protease (Amersham Biosciences, Germany) per 50 mg of fusion protein overnight at 277 K. SPN1 was separated from GST using a heparin column. The sample was loaded onto a heparin column equilibrated in HEPES–NaOH buffer pH 7.5 and eluted with a gradient of 100 m*M* to 1 *M* NaCl. Fractions containing SPN1 were concentrated using Vivaspin concentrators with a cutoff of 10 kDa (Sartorius, Germany) to a total volume of 5 ml and further purified by applying the sample onto a Superdex 75 (26/60) gel-filtration column (Amersham Biosciences, Germany) equilibrated in 100 m*M* NaCl, 50 m*M* Tris–HCl pH 7.5, 2 m*M* EDTA and 2 m*M* DTT to remove the remaining impurities. The chromatographic purification steps were carried out at room temperature using either ÄKTA Prime or ÄKTA Purifier platforms (Amersham Biosciences, Germany). Pure SPN1 was concentrated using Vivaspin concentrators with a cutoff of 10 kDa (Sartorius, Germany) to a final concentration of ∼10 mg ml^{-1} and used for crystallization trials.

2.2. Limited proteolysis of SPN1 and *in vitro* m_3G cap-binding activity of proteolytic fragments

Since no crystals of full-length SPN1 were obtained in the crystallization trials either in the absence or in the presence of m_3G cap dinucleotides (data not shown), limited proteolysis experiments were performed in order to define stable fragments of SPN1 that still retain m_3G cap-binding activity.

Initially, various proteases, *i.e.* plasmin, thrombin, trypsin, LysC, elastase and V8 protease (GluC, *Staphylococcus aureus*; Roche, Germany), were tested for proteolytic activity in the presence or absence of an m_3GpppAmpUmpA-cap ribo-oligonucleotide, henceforth termed m_3G cap oligonucleotide (Sekine *et al.*, 1996; data not shown). From these proteases, only treatment with GluC resulted in the formation of larger stable fragments of SPN1 (Fig. 1*a*). The cleavage pattern obtained by GluC did not differ significantly in the presence or absence of the m_3G cap oligonucelotide (data not shown). The digestion kinetics of full-length SPN1 (denoted 1 in Fig. 1*a*) revealed that for GluC protease three major fragments of apparent molecular weights of approximately 40, 28 and 26 kDa occurred (denoted 2, 3 and 4 in Fig. 1*a*). At early time points, the major bands were those of 40 and 28 kDa, whereas the shortest fragment (26 kDa) showed up later. In a subsequent step, the fragments obtained by GluC digest were tested for the remainder of the m_3G cap-binding activity. These experiments revealed that all three fragments obtained by GluC protease digest bind the m_3G cap oligonucleotide (Fig. 1*b*). Polypeptide fragments obtained by GluC digest were analyzed by automated N-terminal sequencing. The N-termini of the 40, 28 and 26 kDa fragments were determined to be Lys4, Asn79 and Met87, respectively. The C-termini were predicted by combination of the estimated molecular weight obtained by SDS–PAGE and analysis of the SPN1 amino-acid sequence for putative GluC-specific cleavage sites (Fig. 1*c*), resulting in calculated molecular weights of 37, 26 and 24 kDa, respectively.

Thus, as it resulted in higher yields of active products, the GluC protease-digestion approach was used to prepare large amounts of the 37 kDa SPN1 fragment for crystallization. After purification of full-length SPN1, it was digested with GluC protease in a 50:1 molar ratio in the presence of 1 m*M* $CaCl_2$ for 40 min at a temperature of 296 K. Protease activity was inhibited by the addition of PMSF to a final concentration of 1 m*M* and the digestion mixture was separated using a Superdex S75 26/60 column (Amersham Biosciences, Germany). The peak fractions were analyzed by SDS–PAGE and subsequent Coomassie staining; the fractions containing the 37 kDa fragment of SPN1 were pooled and concentrated to 7 mg ml^{-1} using Vivaspin concentrators and an m_3GpppG cap dinucleotide (Kedar SC, Warsaw, Poland) was subsequently added in fivefold molar excess.

2.3. Crystallization of a 26 kDa fragment of SPN1

All crystallization experiments were set up at 293 K in Chryschem sitting-drop plates sealed with clear tape (Hampton Research, USA). 2 µl of a solution of human SPN1 37 kDa fragment (∼7 mg ml^{-1}) with a fivefold molar excess of m_3GpppG cap dinucleotide were mixed with 2 µl reservoir solution. Needle-like crystals grew in 8%(*w*/*v*) PEG 20K and 100 m*M* MES pH 6.0 at 293 K within 4 d (Fig. 2*a*). These crystals were crushed and were used for microseeding at different dilutions or by transfer with a cat whisker to 8%(*w*/*v*) PEG 10K and 200 m*M* sodium citrate pH 5.5. Hexagonal-shaped single crystals of dimensions of 50 × 50 × 30 µm grew out of seeds within 3 d (Fig. 2*b*), but these crystals only diffracted X-rays up to a resolution of ∼4 Å. These crystals belong to space group *P*3, with unit-cell parameters $a = b = 56.3$, $c = 269.1$ Å, $\alpha = \beta = 120$, $\gamma = 90^\circ$.

The hexagonal-shaped crystals were broken into pieces and used for a second

round of seeding in the same buffer [8%(*w*/*v*) PEG 10K, 200 m*M* sodium citrate pH 5.5]. The crystals grew further and, surprisingly, a different crystal form appeared after 5 d (100 × 100 × 200 µm; Fig. 2*c*).

The complex seeding procedure required to produce crystals of adequate size led to the assumption that either the protein has to undergo molecular rearrangements or the protein assembling in the crystals is of a minor species in the crystallization drop. In order to test this hypothesis, crystals were harvested out of crystallization droplets and their content was analyzed by SDS–PAGE (Fig. 3). As supposed, the SPN1 fragment in the crystal is of smaller size (∼26 kDa) than the truncated SPN1 (∼37 kDa) used for setting up the crystallization trials. The molecular weight of the crystallized fragment was determined by mass spectrometry to be ∼26 000 Da (data not shown). The 37 kDa truncated SPN1 was excluded from the crystal. This observation clearly shows that further degradation occurred in the droplet before crystal formation, most likely because of incomplete inhibition of the GluC protease.

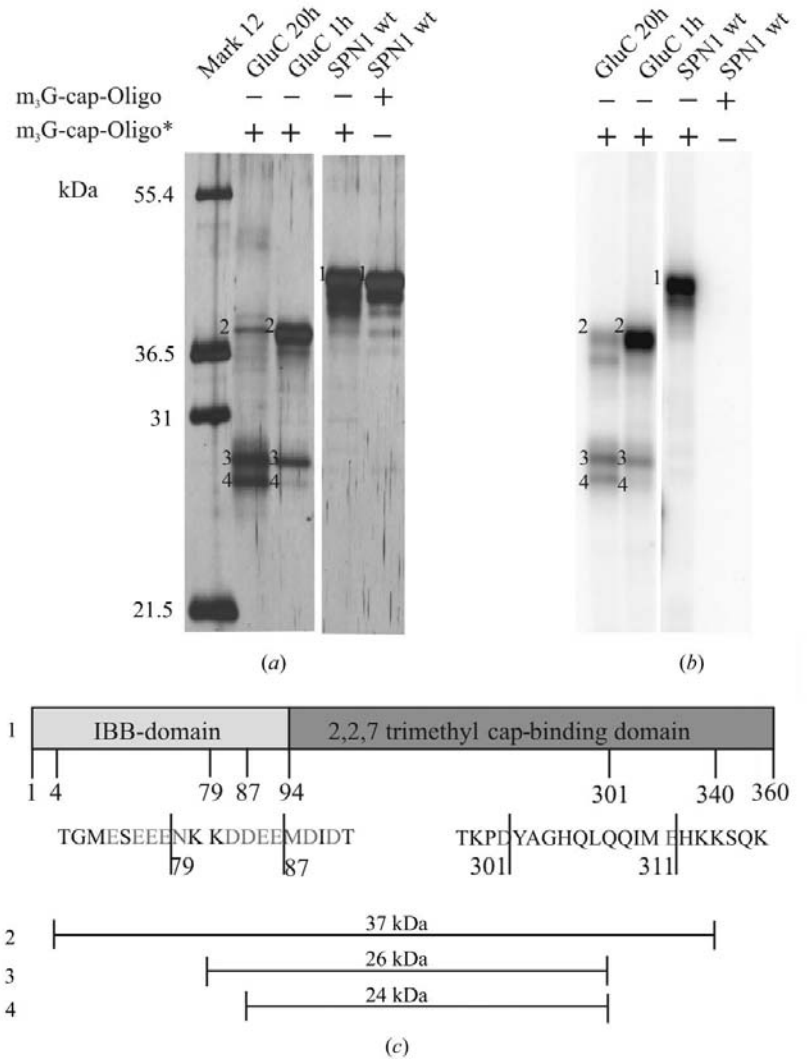

Figure 1
Proteolytic studies on SPN1 and subsequent UV crosslinking in the presence of a radiolabeled m_3G cap oligonucleotide. SPN1 was digested with GluC protease at a protein:protease ratio of 50:1 at 296 K and aliquots were harvested at the time points (hours) indicated. Subsequently, the digestion products were mixed with a ^{32}P-labeled m_3G cap oligonucleotide and UV-crosslinked at 254 nm on ice for 5 min. Separation was achieved by 12.5% SDS–PAGE. The gel was silver-stained and dried (*a*) before autoradiographic analysis (*b*). The numbers to the left of the bands indicate (1) the position of full-length SPN1 and (2), (3) and (4) the N-terminally sequenced fragments of 40, 26 and 24 kDa, respectively, on the gel. The molecular weight of the fragments was calculated on the basis of the known N-terminus and possible cleavage sites for GluC at the C-terminus (*c*). Note that the apparent molecular weights of the various fragments differ from the calculated molecular weight, which seems to be a property of the domain, at least in part, common to all fragments.

2.4. X-ray diffraction experiments and crystal characterization.

Single crystals obtained by the double-seeding procedure were initially of poor diffraction quality using standard cryo-conditions and procedures. However, the diffraction quality significantly increased when the crystals were transferred into cryobuffer and equilibrated in this buffer

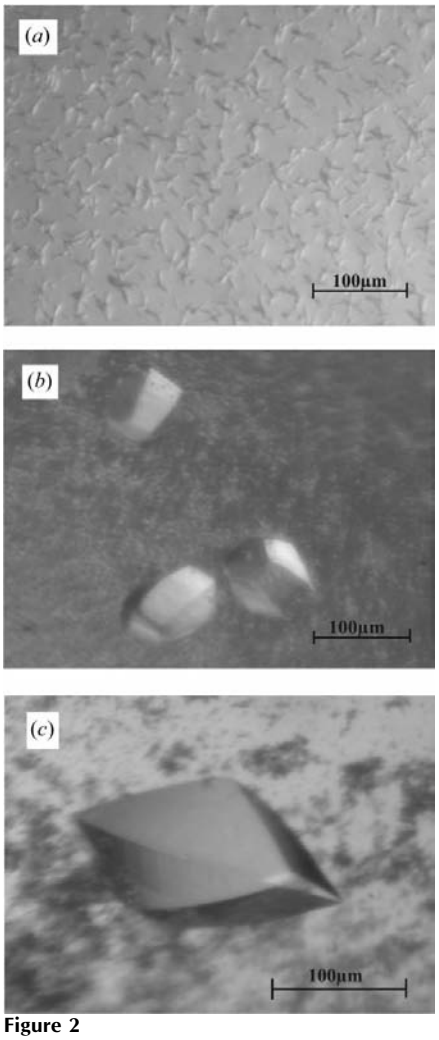

Figure 2
Optimization of crystal-growth conditions obtained from primary screens. (*a*) Crystals grown in 8%(*w*/*v*) PEG 20K and 0.1 *M* MES pH 6.0 were used for different seeding techniques, such as dilution of crushed crystals in mother liquor (1:200, 1:400, 1:800) or using a cat whisker. (*b*) Seeds were diluted (1:200) in 8%(*w*/*v*) PEG 10K and 0.2 *M* sodium citrate pH 5.5 and 0.2 µl of this mixture was added to a new 2 + 2 µl drop with fresh buffer. Hexagonal-shaped crystals grew within 3 d, with an edge length of up to 30–50 µm. (*c*) Seeds of the hexagonal crystals were crushed, diluted (1:200) and seeded with a cat whisker in 8%(*w*/*v*) PEG 10K, 0.2 *M* sodium citrate pH 5.5. Crystals grew within 4–5 d, with an edge length of up to 200 µm. Before the crystals were mounted on a rotating anode, they were equilibrated overnight in a 6 µl drop of 10%(*w*/*v*) PEG 10K and 5%(*v*/*v*) glycerol in a sealed sitting-drop well plate containing 10%(*w*/*v*) PEG 10K and 40%(*v*/*v*) glycerol in the reservoir solution.

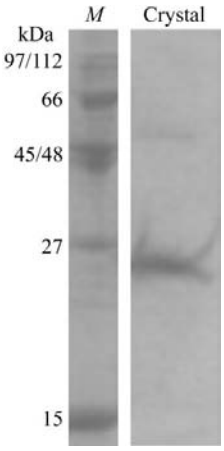

Figure 3
SDS–PAGE of dissolved SPN1 crystals. SPN1 crystals used in diffraction experiments (see Fig. 2*c*) were washed in crystallization buffer and dissolved in Laemmli buffer. The solubilized crystals were applied to SDS–PAGE gel together with molecular-weight markers and bands were visualized by Coomassie staining. Bands from the solubilized protein are shown as well as the molecular-weight markers (lane *M*).

overnight. The best diffraction was obtained by transferring the crystals to a sitting drop containing cryobuffer composed of 5%(*v*/*v*) glycerol and 10%(*w*/*v*) PEG 10K, while the reservoir solution contained 40%(*v*/*v*) glycerol and 10%(*w*/*v*) PEG 10K. Crystals were mounted in a fibre loop and flash-frozen in a 100 K nitrogen-gas stream. X-ray diffraction data were collected using a Micromax 007 rotating-anode generator (Rigaku/MSC, USA) operating at 40 kV and 20 mA equipped with Osmic focusing mirrors and a MAR345dtb detector system (X-ray Research, Germany). The crystal-to-detector distance was 250 mm. A data set of 100 frames was collected applying 1° rotation and a 20 min exposure time per image. Crystals diffracted to a maximum resolution of 2.9 Å, with an estimated mosaicity of 0.7°. Processing the data with *DENZO*/*SCALEPACK* (HKL Research, USA) revealed a primitive lattice with tetragonal symmetry and unit-cell parameters $a = b = 57.47$, $c = 130.09$ Å, $\alpha = \beta = \gamma = 90^\circ$. According to systematic absences, the space group was determined to be $P4_12_12$ (or its enantiomorph $P4_32_12$), with one molecule in the asymmetric unit, corresponding to a solvent content of ∼44%. The data-collection statistics are summarized in Table 1.

Table 1
X-ray data-collection statistics.

Values in parentheses are for data in the highest resolution shell.

Space group	$P4_12_12/P4_32_12$
Wavelength (Å)	1.54
Resolution range (Å)	50–2.9 (3.0–2.9)
Measured reflections	33 262 (3319)
Unique reflections	5151 (490)
Completeness (%)	98.6 (98.0)
Mean $I/\sigma(I)$	18.2 (13.6)
$R_{\rm merge}$† (%)	7.1 (33.9)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_{hkl} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i \langle I_{hkl}\rangle$.

We thank Klaus Reuter for support in constructing the initial expression plasmids and Annette Berndt for technical assistance in purifying large amounts of SPN1. We are indebted to Bernhard Schmidt for N-terminal sequencing of the SPN1 fragments. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 523: TP A11 to RF and TP A8 to RL).

References

Allen, N. P., Huang, L., Burlingame, A. & Rexach, M. (2001). *J. Biol. Chem.* **276**, 29268–29274.
Chaillan-Huntington, C., Braslavsky, C. V., Kuhlmann, J. & Stewart, M. (2000). *J. Biol. Chem.* **275**, 5874–5879.
Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T. & Matunis, M. J. (2002). *J. Cell Biol.* **158**, 915–927.
Dingwall, C. & Laskey, R. A. (1991). *Trends Biochem. Sci.* **16**, 478–481.
Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. (1997). *Cell*, **90**, 1051–1060.
Gerace, L. (1995). *Cell*, **82**, 341–344.
Gorlich, D., Dabrowski, M., Bischoff, F. R., Kutay, U., Bork, P., Hartmann, E., Prehn, S. & Izaurralde, E. (1997). *J. Cell Biol.* **138**, 65–80.
Gorlich, D., Prehn, S., Laskey, R. A. & Hartmann, E. (1994). *Cell*, **79**, 767–778
Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B. K. & Izaurralde, E. (1998). *Mol. Cell*, **1**, 649–659.
Herold, A., Suyama, M., Rodrigues, J. P., Braun, I. C., Kutay, U., Carmo-Fonseca, M., Bork, P. & Izaurralde, E. (2000). *Mol. Cell Biol.* **20**, 8996–9008.
Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. & Lührmann, R. (1998). *EMBO J.* **17**, 4114–4126.
Jullien, D., Gorlich, D., Laemmli, U. K. & Adachi, Y. (1999). *EMBO J.* **18**, 4348–4358.
Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J. U. & Hurt, E. (1999). *EMBO J.* **18**, 2593–2609.
Massenet, S., Pellizzoni, L., Paushkin, S., Mattaj, I. W. & Dreyfuss, G. (2002). *Mol. Cell Biol.* **22**, 6533–65341.
Mattaj, I. W. (1986). *Cell*, **46**, 905–911.
Meister, G., Hannus, S., Plottner, O., Baars, T., Hartmann, E., Fakan, S., Laggerbauer, B. & Fischer, U. (2001). *EMBO J.* **20**, 2304–2314.
Moroianu, J., Blobel, G. & Radu, A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 2008–2011.
Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. (2000). *Cell*, **101**, 187–198.
Palacios, I., Hetzer, M., Adam, S. A. & Mattaj, I. W. (1997). *EMBO J.* **16**, 6783–6792.
Reichelt, R., Holzenburg, A., Buhle, E. L. Jr, Jarnik, M., Engel, A. & Aebi, U. (1990). *J. Cell Biol.* **110**, 883–894.
Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y. & Chait, B. T. (2000). *J. Cell Biol.* **148**, 635–652.
Segref, A., Mattaj, I. W. & Ohno, M. (2001). *RNA*, **7**, 351–360.
Sekine, M., Kadokura, M., Satoh, T., Seio, K., Wada, T., Fischer, U., Sumpter, V. & Lührmann, R. (1996). *J. Org. Chem.* **61**, 4412–4422.
Stoffler, D., Feja, B., Fahrenkrog, B., Walz, J., Typke, D. & Aebi, U. (2003). *J. Mol. Biol.* **328**, 119–130.
Weis, K., Mattaj, I. W. & Lamond, A. I. (1995). *Science*, **268**, 1049–1053.
Will, C. L. & Lührmann, R. (2001). *Curr. Opin. Cell Biol.* **13**, 290–301.