

Structural and functional characterization of proteins involved in the biogenesis of spliceosomal U snRNPs

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Thomas Monecke
aus Leinefelde

Göttingen 2009

D7

Referent: Herr Prof. Dr. Ralf Ficner
Abteilung für Molekulare Strukturbiologie
Institut für Mikrobiologie und Genetik
Georg-August-Universität Göttingen

Korreferent: Herr Prof. Dr. Holger Stark
Max-Planck-Institut für biophysikalische Chemie und
Göttinger Zentrum für Molekulare Biowissenschaften

Tag der mündlichen Prüfung:

Table of contents

1 • Summary	1
2 • Zusammenfassung.....	4
3 • General Introduction	7
3.1 NUCLEOCYTOPLASMIC TRANSPORT	7
3.1.1 The Ran cycle.....	9
3.1.2 Importin β superfamily (β -karyopherins)	10
3.2 SPLICING OF MRNA BY THE SPLICEOSOME	14
3.2.1 Biogenesis of spliceosomal UsnRNPs.....	16
3.3 THE CAP DIMETHYLTRANSFERASE TGS1	18
3.4 THE NUCLEAR EXPORT COMPLEX CRM1·SPN1·RANGTP	20
3.4.1 The nuclear export receptor CRM1	21
3.4.2 The snRNP import adapter SPN1	22
3.5 DEADENYLATION DEPENDENT MRNA DECAY.....	23
3.5.1 The poly(A)-specific ribonuclease (PARN).....	24
4 • Crystal structure of the conserved methyltransferase domain of TGS1	27
5 • Structural basis for m⁷G-cap dimethylation by human TGS1	38
5.1 SUPPLEMENTARY DATA.....	58
6 • Crystal structure of the nuclear export complex CRM1·SPN1·RanGTP.....	59
6.1 SUPPORTING ONLINE MATERIAL.....	68
7 • Crystal structure of the RRM domain of PARN.....	79
8 • Additional Results and Discussion.....	92
8.1 THE TRIMETHYLGUANOSINE SYNTHASE 1	92
8.2 THE NUCLEAR EXPORT COMPLEX CRM1·SPN1·RANGTP	98
8.3 THE POLY(A)-SPECIFIC RIBONUCLEASE	103
9 • References	108
Appendix I • Abbreviations	119
Appendix II • Danksagungen.....	122
Appendix III • Curriculum Vitae.....	124

»Jedem tiefen Naturforscher muß eine Art religiösen Gefühls naheliegen, weil er sich nicht vorzustellen vermag, daß die ungemein feinen Zusammenhänge, die er erschaut, von ihm zum erstenmal gedacht werden.«

Albert Einstein

Chapter 1 • *Summary*

One major hallmark of eucaryotic organisms is the subdivision of their cells into different, membrane enclosed compartments. In their central compartment, the nucleus, eucaryotes store their condensed genetic information. In order to translate this genetic information into the protein sequence the DNA has to be transcribed into RNA. However, transcription does not immediately lead to mature messenger RNA (mRNA) but results in a pre-mature messenger RNA (pre-mRNA) containing coding regions (exons) as well as non-coding regions (introns). Prior to their transport to the cytoplasm the introns of pre-mRNAs are excised by a large ribonucleoprotein complex, the spliceosome. The major components of the spliceosome are the uridyl-rich small nuclear ribonucleoprotein particles (so-called UsnRNPs), whose biogenesis requires a nucleocytoplasmic transport cycle.

The major goal of the present work was the structural and functional characterization of two different proteins and a protein complex, which either are required for the biogenesis of UsnRNPs or which bind the m⁷G-cap. The first protein described is the dimethyltransferase TGS1 (Trimethylguanosine Synthase 1), which catalyzes the dimethylation of the 5'-guanine base of specific RNA caps. The second project concerned the characterization of a nuclear export complex, consisting of the cargo protein snurportin 1, its exportin (exportin 1) as well as the molecular switch RanGTP. The third project comprised the analysis and characterization of the binding mode of the poly(A)-specific ribonuclease (PARN) to the mRNA 5'-cap. The results of the three independent projects are briefly summarized below.

The methyltransferase domain of the Trimethylguanosine Synthase 1 (TGS1), which hypermethylates the m⁷G-cap of spliceosomal UsnRNAs during biogenesis of the corresponding RNP was crystallized and its crystal structure was determined. The active form of the methyltransferase domain comprises the structurally conserved methyltransferase fold as well as a small N-terminal and α -helical domain. Biochemical as well as further crystallographic analyses revealed that this additional N-terminal domain is strictly required for both, substrate binding and catalysis. This functional characterization was enabled by a newly established HPLC-based methyltransferase activity assay. Moreover, a previously postulated structure based reaction mechanism could be verified biochemically by a combination of this assay and site-directed mutagenesis studies.

The nuclear export complex comprising exportin 1 (Xpo1, CRM1), its cargo snurportin1 as well as the molecular switch Ran in its GTP bound form was recombinantly expressed, assembled *in vitro* and crystallized. The crystal structure analysis revealed that CRM1 adopts an overall superhelical, toroid-shaped conformation and that the GTPase Ran is enwrapped by the exportin's inner surface. Unexpectedly, the cargo snurportin 1 binds on the outer surface of CRM1 including three different areas and does not make a single direct contact to the molecular switch RanGTP. However, between the cargo and RanGTP indirect contacts are mediated by the so-called acidic loop of CRM1. This strategy explains on the one hand the extremely broad substrate spectrum of CRM1 and on the other hand the apparent cooperativity of binding between RanGTP and snurportin 1.

The RNA recognition motif (RRM) of the poly(A)-specific ribonuclease (PARN) was purified and crystallized in complex with the cap analog m⁷GTP. The crystal structure as well as further biochemical studies revealed that the positively charged m⁷G-cap is bound in an unexpected mode, which has not been observed so far. While other structurally defined cap-binding proteins bind the methylated purine base in between two aromatic or hydrophobic residues, PARN stacks the base only on one side by a tryptophan side chain and lacks a protein residue on the opposing side. The binding mode observed in the crystal structure could be verified by means of fluorescence spectroscopy. The change of the emitted tryptophan fluorescence upon cap binding allowed the determination of PARN-cap dissociation constants for the wild type protein as well as for some single amino acid mutants.

The present work resulted in the following four publications. Authors marked with † contributed equally to the corresponding publication.

1. Thomas Monecke, Achim Dickmanns, Anja Strasser and Ralf Ficner (**2009**). Structure analysis of the conserved methyltransferase domain of human trimethylguanosine synthase TGS1, *Acta Crystallographica Section D*, **65**(4), 332-38.

2. Thomas Monecke, Achim Dickmanns and Ralf Ficner (**2009**). Structural basis for m⁷G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1, *Nucleic Acids Research*, Epub ahead of print, Advance Access published April 22, 2009.

3. Thomas Monecke[†], Thomas Güttler[†], Piotr Neumann, Achim Dickmanns, Dirk Görlich and Ralf Ficner (**2009**). Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP, *Science*, **324**, 1087-91.

4. Thomas Monecke[†], Stephanie Schell[†], Achim Dickmanns and Ralf Ficner (**2008**). Crystal Structure of the RRM Domain of Poly(A)-Specific Ribonuclease Reveals a Novel m⁷G-Cap-Binding Mode, *Journal of Molecular Biology*, **382**(4), 827-34.

Chapter 2 • Zusammenfassung

Ein zentrales Merkmal eukaryotischer Zellen ist ihre Unterteilung in verschiedene von Membranen umschlossene Kompartimente. Die genetische Information eukaryotischer Organismen liegt in komprimierter Form in ihrem Zellkern, dem zentralen Kompartiment, vor. In eukaryotischen Zellen wird die Boten-RNA (engl.: *messenger RNA*) als Vorläufer-Boten-RNA transkribiert, deren kodierende Bereiche (Exons) durch nicht-kodierende Intronbereiche getrennt sind. Die Introns werden vor dem Transport der reifen Boten-RNA in das Zytoplasma durch einen großen Ribonukleoprotein-Komplex, das Spleißosom, entfernt. Das Spleißosom besteht hauptsächlich aus den so genannten uridin-reichen kleinen nukleären Ribonukleoproteinpartikeln (engl.: *uridyl-rich small nuclear ribonucleoprotein particles*, UsnRNPs). Die Biogenese der UsnRNPs schließt in höheren Eukaryoten einen nukleozytoplasmatischen Transportzyklus ein.

Das Ziel der vorliegenden Arbeit bestand in der strukturellen sowie funktionellen Charakterisierung zweier verschiedener Proteine und eines Proteinkomplexes, die im Zusammenhang mit der UsnRNP Biogenese stehen bzw. das RNA 5'-cap binden. Dabei handelte es sich um die Dimethyltransferase TGS1 (Trimethylguanosin Synthase 1), die die zweifache Methylierung des 5'-Guanosins spezieller RNA-caps katalysiert. Der zweite Schwerpunkt lag auf der Charakterisierung eines Exportkomplexes, welcher aus dem zu transportierenden Protein Snurportin 1, dessen Exportin (Exportin 1) und dem kleinen molekularen Schalter RanGTP besteht. Darüberhinaus sollte der Bindungsmodus der Poly(A)-spezifischen Ribonuklease an das 5'-mRNA-cap strukturell untersucht und beschrieben werden. Im Folgenden werden die Ergebnisse der einzelnen Projekte kurz zusammengefasst.

Die Methyltransferase-Domäne der Trimethylguanosin Synthase 1 (TGS1), die das m⁷G-cap der spleißosomalen UsnRNAs während ihrer Reifung im Zytoplasma zweifach methyliert, wurde kristallisiert und ihre dreidimensionale Struktur wurde bestimmt. Die aktive Form dieser Domäne besteht aus dem strukturell konservierten Methyltransferase-Faltungsmotiv sowie einer kleinen N-terminalen, α -helikalen Domäne. Durch biochemische Analysen, sowie weitere kristallographische Untersuchungen konnte gezeigt werden, dass die N-terminale Domäne für die korrekte Bindung beider Substrate, sowie für die katalytische Aktivität des Enzyms essentiell ist. Die funktionelle Analyse wurde durch einen neu entwickelten HPLC-

gestützten Methyltransferase-Aktivitätstest ermöglicht. Desweiteren konnte durch die Kombination dieses Aktivitätstests mit Mutagenese-Studien ein Reaktionsmechanismus verifiziert werden, der auf Grundlage der Kristallstruktur postuliert worden war.

Der Exportkomplex bestehend aus dem Exportin 1 (Xpo1, CRM1), seinem Kargo Snurportin 1 und dem molekularen Schalter Ran in seiner GTP gebundenen Form wurde *in vitro* assembliert und kristallisiert. Die Analyse der Kristallstruktur zeigt, dass das Exportin in seiner superhelikalen, torusartigen Gestalt die kleine GTPase Ran in ihrem Inneren einschließt. Dagegen bindet das zu transportierende Snurportin 1 über drei verschiedene Regionen auf der Außenseite des Exportins und steht nicht in direkter Verbindung mit RanGTP. Zwischen beiden Proteinen vermittelt eine saure Schleife des Exportins indirekte Kontakte. Durch diesen Bindungsmodus lässt sich das beobachtete außerordentlich breite Substratspektrum des Exportins 1 erklären. Außerdem ist auch die Kooperativität bei der Bildung des Exportkomplexes aus seinen drei Bestandteilen auf diese Anordnung der Komponenten zurückzuführen.

Die RNA-Bindedomäne der Poly(A)-spezifischen Ribonuklease (PARN) wurde gereinigt und im Komplex mit dem *cap*-Analog m^7GTP kristallisiert. Die Analyse der Kristallstruktur sowie weitere biochemische Charakterisierungen zeigten, dass das positiv geladene m^7G-cap in einer bisher nicht beobachteten Form an das Protein gebunden ist. Dabei wird die Purin-Base des $m^7G-caps$ auf einer Seite von der Seitenkette eines Tryptophanrestes flankiert, während sich überraschenderweise auf der anderen Seite der Base keine Proteinseitenkette befindet. Die in der Kristallstruktur beobachtete Bindung konnte über Fluoreszenzspektroskopie für das Wildtypprotein sowie verschiedene Mutanten verifiziert werden. Dabei ermöglichte die Änderung der emittierten Fluoreszenz von Tryptophanen nach der Bindung des zugegebenen $m^7G-caps$ die Bestimmung von Dissoziationskonstanten.

Die vorliegende Arbeit führte zu den vier nachfolgend aufgeführten Publikationen. Autoren, deren Name mit † markiert ist, haben zu gleichen Teilen zu der entsprechenden Publikation beigetragen.

1. Thomas Monecke, Achim Dickmanns, Anja Strasser and Ralf Ficner (**2009**). Structure analysis of the conserved methyltransferase domain of human trimethylguanosine synthase TGS1, *Acta Crystallographica Section D*, **65**(4), 332-38.

2. Thomas Monecke, Achim Dickmanns and Ralf Ficner (**2009**). Structural basis for m⁷G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1, *Nucleic Acids Research*, Epub ahead of print, Advance Access published April 22, 2009.

3. Thomas Monecke[†], Thomas Güttler[†], Piotr Neumann, Achim Dickmanns, Dirk Görlich and Ralf Ficner (**2009**). Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP, *Science*, **324**, 1087-91.

4. Thomas Monecke[†], Stephanie Schell[†], Achim Dickmanns and Ralf Ficner (**2008**). Crystal Structure of the RRM Domain of Poly(A)-Specific Ribonuclease Reveals a Novel m⁷G-Cap-Binding Mode, *Journal of Molecular Biology*, **382**(4), 827-34.

Chapter 3 • General Introduction

The major hallmark of eucaryotes is the division of their cells into different, membrane enclosed compartments. Besides mitochondria, chloroplasts, the Golgi apparatus, vesicles and lysosomes eucaryotic cells contain a central compartment, the nucleus. The cell nucleus is surrounded by a double membrane, the nuclear envelope (NE), and contains the genetic information of the cell in form of the deoxyribonucleic acid (DNA). As a consequence of this segregation, transcription and translation are strictly separated from each other allowing for specific and diverse regulation at different levels. Typically and in contrast to procaryotes, eucaryotic genes contain non-coding sequences (introns), which are inserted in between the coding regions (exons). The DNA is transcribed into a messenger RNA precursor (pre-mRNA), which contains introns as well as exons. Large assemblies called spliceosomes remove the introns prior to the transport of the mRNA to the cytoplasm, where it is translated at the ribosomes into the encoded protein sequence. In addition to mRNA, the nucleus supplies the cytoplasm with other important RNA species such as transfer RNAs (tRNAs) or ribosomal subunits, whereas the organelle itself is provided with cytoplasmic products such as proteins, small molecules and other solutes. Consequently, thousands of macromolecules are transferred every second between the nuclear and the cytoplasmic compartment passing the nuclear envelope (Chook *et al.*, 1999; Conti and Izaurralde, 2001; Damelin *et al.*, 2002; Fried and Kutay, 2003; Gorlich and Kutay, 1999; Kuersten *et al.*, 2001; Macara, 2001).

3.1 Nucleocytoplasmic transport

Nuclear transport of macromolecules proceeds through nuclear pore complexes (NPCs), which are large gates embedded in the nuclear envelope (Maco *et al.*, 2006; Schwartz, 2005; Tran and Wente, 2006). While ions, small molecules and proteins with a molecular weight below approximately 20-30 kDa are able to pass the nuclear pores passively, large macromolecules have to be transported by specialized proteins (Cook *et al.*, 2007; Gorlich and Kutay, 1999; Paine *et al.*, 1975; Peters, 2006). Thereby, active transport does not only allow the directional translocation of the cargo in or out of the nucleus but also enables transport processes against an existing concentration gradient (Gorlich and Kutay, 1999). Potential cargoes usually harbor a nuclear localization signal (NLS) for nuclear import,

whereas export cargoes are characterized by a nuclear export signal (NES). The classical NLS (cNLS) is present in a multitude of eukaryotic proteins and consists of a short amino acid sequence that includes at least four lysine residues (Kalderon *et al.*, 1984; Lanford and Butel, 1984; Lanford *et al.*, 1986). An example for a nuclear export signal is the leucine rich NES, which contains at least four characteristically spaced hydrophobic amino acids that are recognized prior to nuclear export (Fornerod *et al.*, 1997; Fukuda *et al.*, 1997; Ossareh-Nazari *et al.*, 1997; Stade *et al.*, 1997). Further signals include the bipartite NLS (bpNLS) (Robbins *et al.*, 1991) or the M9 transport signal (Pollard *et al.*, 1996).

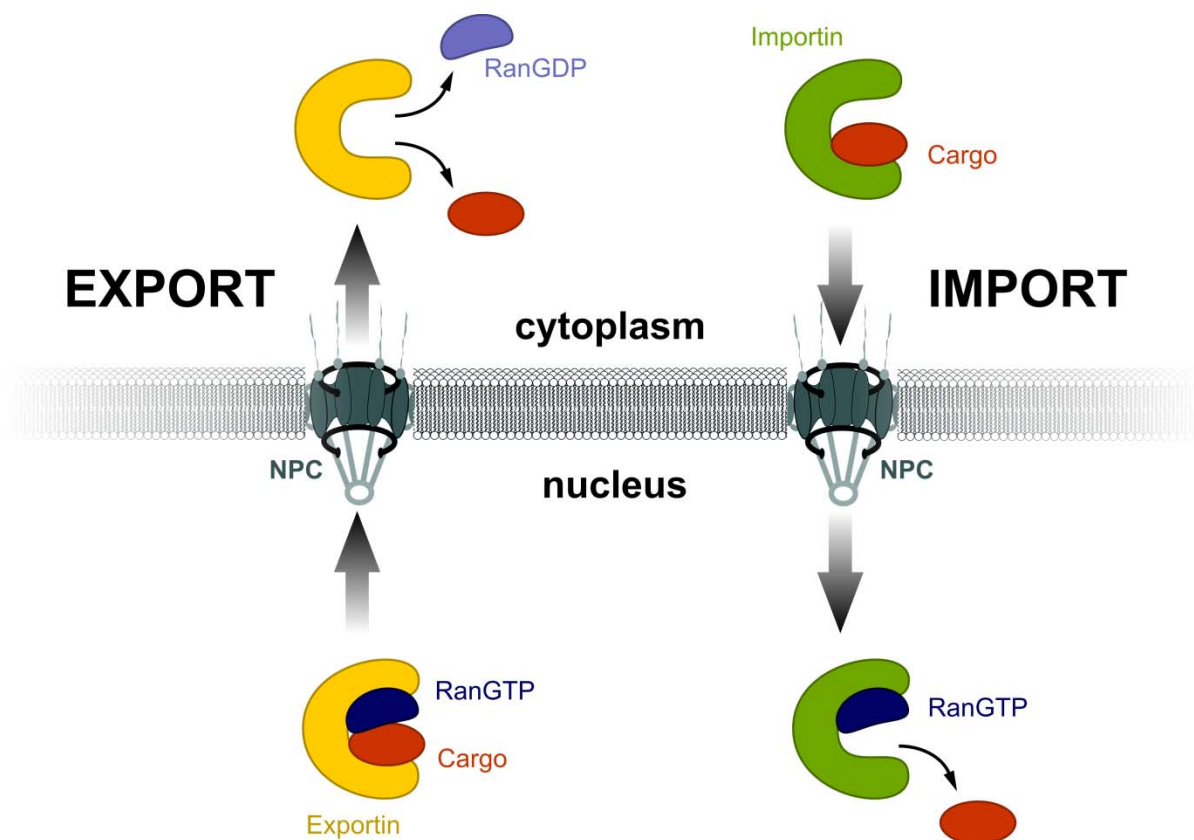


Figure 3-1: Schematic representation of nuclear export and import by β -karyopherins. In nuclear export (left hand side), the exportin binding to RanGTP in the nucleoplasm is a prerequisite for cargo binding. The ternary complex translocates through the nuclear pore complex (NPC) and is disassembled in the cytoplasm upon RanGTP hydrolysis. Conversely, importins (right hand side) bind cargoes in the cytoplasm, pass the NPC and release import cargoes in the nucleus upon RanGTP binding.

The signal sequences of the cargoes are either recognized directly by the transport receptor or the interaction is mediated *via* adapter molecules (Fried and Kutay, 2003; Weis, 2003). Given that there are numerous cargoes with different import and export signals, it is surprising that, with only a few exceptions, all transport receptors belong to a single protein family called importin β superfamily (Cook *et al.*, 2007; Gorlich and Kutay, 1999; Lei and Silver, 2002; Macara, 2001; Mattaj and Englmeier, 1998; Strom and Weis, 2001; Weis, 2002). These

proteins, also known as β -karyopherins, share weak sequence homology and are subdivided into receptors mediating the import (importins) and exportins, which mediate export processes (Figure 3-1) (Cook *et al.*, 2007). In nuclear import the importin binds the cargo in the cytoplasm and traverses the nuclear pore. On the nuclear side the small GTPase Ran (rat sarcoma related nuclear antigen) in its GTP bound form (RanGTP) binds to the importin and functions as a molecular switch mediating the release of the bound cargo. In nuclear export, by contrast, the binding of RanGTP to the exportin is a prerequisite for cargo binding. Accordingly, only the ternary complex of exportin-RanGTP-cargo is able to cross the nuclear envelope through the NPC. In the cytoplasm, hydrolysis of the GTP molecule bound by Ran to GDP mediates export complex disassembly. Thus, RanGTP controls both transport processes conversely; it is an integral component of the nuclear export complex, while importins release their cargoes upon RanGTP binding (Gorlich and Kutay, 1999). By using this strategy, the cell ensures that on the one hand an import complex cannot leave the nucleus prior to cargo release. On the other hand a complex consisting of an exportin and a cargo molecule alone cannot persist in the cytoplasm ensuring the return of the exportin into the nucleus.

3.1.1 The Ran cycle

The regulatory role of Ran is achieved by a steep concentration gradient of RanGTP with a high concentration in the nucleus and a low concentration in the cytoplasm facilitating the directionality of both transport processes (Figure 3-1, Figure 3-2) (Gorlich *et al.*, 1996). As this gradient is the major control element of nucleocytoplasmic transport, the question arises how such a difference in concentration is established, maintained and controlled. Like most GTPases (Bourne *et al.*, 1991), Ran hydrolyzes GTP very slowly and as a result its nucleotide binding state is regulated by interaction with different regulatory proteins. In fact, the low hydrolysis rate of Ran can be enhanced by a factor of 10,000 by the Ran GTPase activating protein 1 (RanGAP1) (Bischoff *et al.*, 1995). In the presence of the Ran binding protein 1 (RanBP) the GTPase activity is even increased by a factor of 100,000. Thus, in the cytoplasm, the concentration of RanGTP is kept low by RanGAPs and RanBPs, while a Ran nucleotide exchange factor (RanGEF), the regulator of chromosome condensation 1 (RCC1), ensures high concentrations of RanGTP in the nucleus (Figure 3-2). RCC1 is associated with the histones H2A and H2B in the nucleus (Nemergut *et al.*, 2001) and is able to accelerate the nucleotide exchange of Ran 10,000-fold (Bischoff and Ponstingl, 1991).

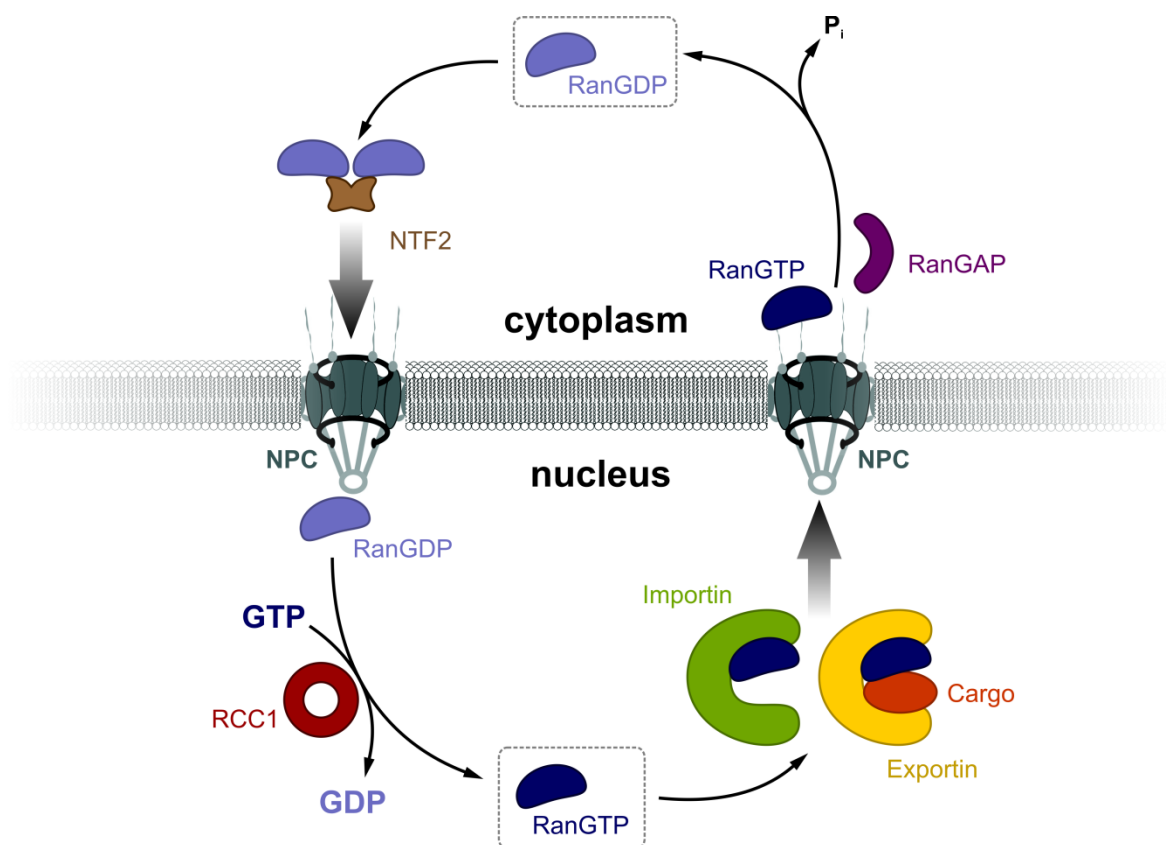


Figure 3-2: Schematic representation of the Ran cycle. RanGTP which enters the cytoplasm *via* binding to exportins and importins is hydrolyzed in the cytosol by the GTPase-activating protein (RanGAP). Ran in its GDP bound form is subsequently imported into the nucleus by the nuclear transport factor 2 (NTF2). RCC1 (regulator of chromosome condensation 1) maintains a high nuclear RanGTP concentration by acting as Ran guanine exchange factor (RanGEF). Consequently, the high concentration of RanGTP allows the unidirectional transport of cargoes across the nuclear envelope.

The asymmetric distribution of these factors with RanGAPs being exclusively present in the cytoplasm (Hopper *et al.*, 1990) and conversely RanGEFs in the nucleus (Nemergut *et al.*, 2001) allows the generation and maintenance of the RanGTP gradient, which is the driving force for unidirectional nuclear transport mediated by β -karyopherins.

3.1.2 Importin β superfamily (β -karyopherins)

Commonly, β -karyopherins exhibit a molecular weight of about 90-150 kDa and consist of repetitive elements, the so-called HEAT repeats (Chook and Blobel, 1999; Cingolani *et al.*, 1999; Cook *et al.*, 2007; Lee *et al.*, 2005; Matsuura and Stewart, 2004; Vetter *et al.*, 1999). These repeats were first identified in Huntingtin, elongation factor 3, PR65/A subunit of protein phosphatase 2A and TOR lipid kinase (HEAT). This tandem-arranged structural element comprises ~ 40 amino acids (Andrade *et al.*, 2001) the consensus sequence of which

is highly degenerated and therefore difficult to predict. Typically, the motif consists of two antiparallel α -helices (A and B), which are separated by a short loop and they are slightly tilted with respect to each other resulting in a molecule with a superhelical twist (Figure 3-3).

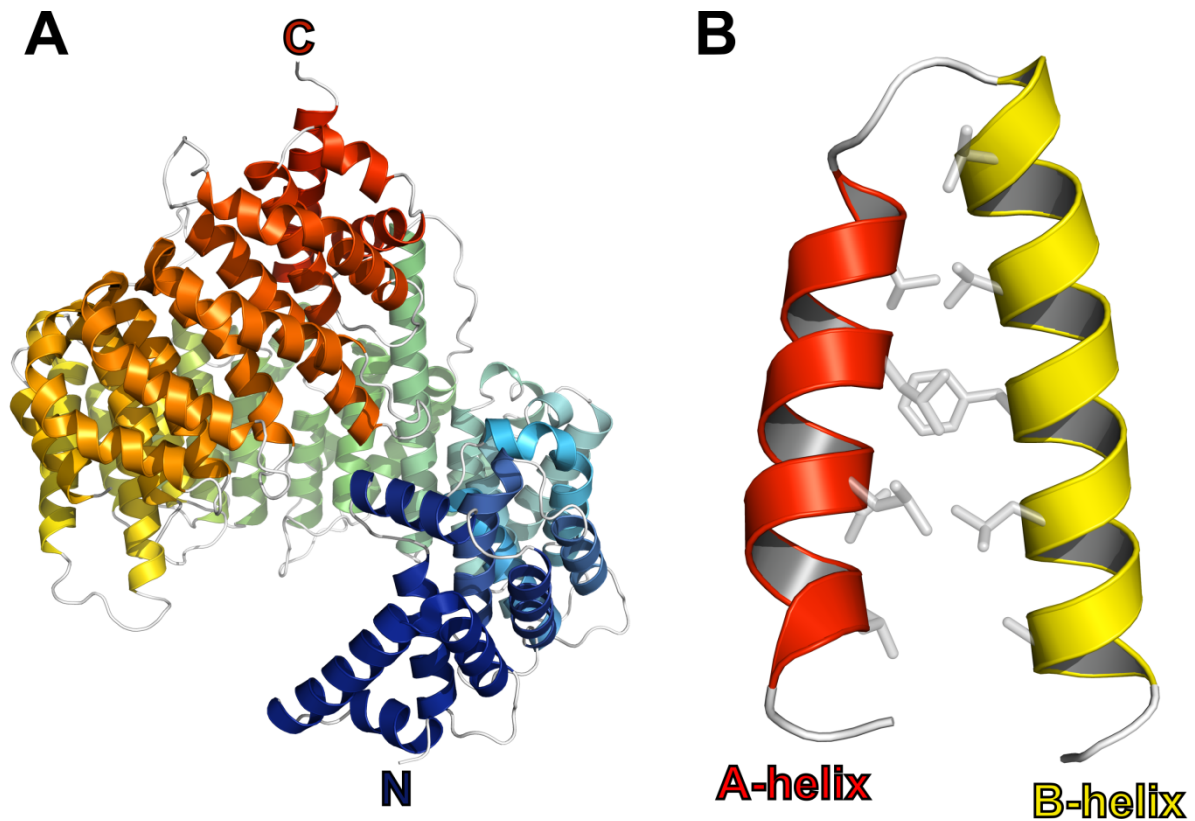


Figure 3-3: Structural organization of β -karyopherins. (A) Crystal structure of importin β . The cartoon representation shows the organization of importin β as crystallized in complex with the importin β binding domain of importin α (importin α_{IBB}) (Cingolani *et al.*, 1999). The IBB domain of importin α is not shown for clarity reasons. Importin β consists of 19 consecutive HEAT repeats (rainbow-colored from the N-terminus in blue to the C-terminus in red) separated by short loops (white). (B) Typical arrangement of the two helices of a HEAT repeat. In the ribbon representation the HEAT repeat 11 of importin β is shown. Typically, helix A (red) is located at the outer, convex surface of the molecule and helix B (yellow) at the inner, concave surface of the superhelical molecule. A and B helix mainly interact *via* hydrophobic residues located in the enclosed cleft. The small tilt of helix A with respect to helix B finally results in the remarkable superhelical twist of importin β .

The first identified β -karyopherin and eponym is the import receptor importin β , which imports a variety of proteins bearing a classical NLS directly or *via* diverse adapter molecules (Harel and Forbes, 2004). Importin α is such an adapter for various cargoes mediating their import indirectly by binding to importin β (Cook *et al.*, 2007; Stewart, 2007). In this case the interaction is mediated by the importin β binding domain (IBB) of importin α (Cingolani *et al.*, 1999). Figure 3-3 shows the crystal structure of importin β as crystallized in complex with the importin β binding domain of importin α (importin α_{IBB}) (Cingolani *et al.*, 1999). The overall structure of importin β is shown in cartoon representation and it consists of 19 HEAT repeats which are rainbow-colored from dark blue (HEAT 1) to dark red (HEAT 19). The

small clockwise tilt of helix A with respect to helix B (Figure 3-3 B) leads to a remarkable superhelical twist and intrinsic flexibility of the whole structure. In fact, the possibility to arrange different numbers of consecutive HEAT repeats and to vary the amino acid sequence selectively offers a high diversity with respect to the establishment of numerous cargo specificities. Table 3-1 shows a selection of importins and exportins identified and gives an impression of the diversity of the transported molecules.

Table 3-1: Selection of different transport receptors mediating import and export processes.

Transport receptor		Cargoes
Importins	Importin β	classical NLS bearing substrates via importin α (Chi <i>et al.</i> , 1995), m ³ G capped snRNPs via SPN1 (Huber <i>et al.</i> , 1998), XRIP α (Jullien <i>et al.</i> , 1999), HIV-Tat und -Rev (Truant and Cullen, 1999), Cyclin B1 (Moore <i>et al.</i> , 1999; Takizawa <i>et al.</i> , 1999), ribosomal proteins (Jakel and Gorlich, 1998), Smad proteins (Xiao <i>et al.</i> , 2000), tyrosine phosphatase (Tiganis <i>et al.</i> , 1997)
	Transportin 1	hnRNP proteins (Pollard <i>et al.</i> , 1996; Siomi <i>et al.</i> , 1997), ribosomal proteins (Jakel and Gorlich, 1998), c-Fos from HIV-1 (Arnold <i>et al.</i> , 2006b)
	Transportin 2	proteins with SR-domain (Kataoka <i>et al.</i> , 1999; Lai <i>et al.</i> , 2000)
	Importin 5	ribosomal proteins (Jakel and Gorlich, 1998)
	Importin 7	H1 linker histones with importin β (Bauerle <i>et al.</i> , 2002; Jakel <i>et al.</i> , 1999), core histones (Muhlhauser <i>et al.</i> , 2001), ribosomal proteins (Jakel and Gorlich, 1998), HIV1-integrase (Fassati <i>et al.</i> , 2003), HIV1-Rev (Arnold <i>et al.</i> , 2006a), p35 in neurons (Fu <i>et al.</i> , 2006), glucocorticoid receptor (Freedman and Yamamoto, 2004)
	Importin 9	core histones (Muhlhauser <i>et al.</i> , 2001)
	Importin 11	UbcM2 (Plafker and Macara, 2000)
	Importin 13	Ubc9, RBM8, eIF1A (Mingot <i>et al.</i> , 2001), NF-YB, NF-YC (Kahle <i>et al.</i> , 2005)
	Exportin 1 (CRM1)	leucine rich NES substrates (Fornerod <i>et al.</i> , 1997; Fukuda <i>et al.</i> , 1997; Ossareh-Nazari <i>et al.</i> , 1997; Stade <i>et al.</i> , 1997), SPN1 (Paraskeva <i>et al.</i> , 1999), UsnRNAs (Ohno <i>et al.</i> , 2000)
Exportins	CAS	importin α (Kutay <i>et al.</i> , 1997)
	Exportin-t	tRNAs (Arts <i>et al.</i> , 1998; Kutay <i>et al.</i> , 1998)
	Exportin 4	eIF5A (Lipowsky <i>et al.</i> , 2000), Sox family transcription factors (Gontan <i>et al.</i> , 2009)

An important feature of transport receptors, in contrast to all other proteins, is their ability to traverse the nuclear pores without hindrance irrespective of their size. The reason for that remarkable fact lies in the architecture of the NPC. This strictly controlled nuclear gate can be divided into three major parts (Figure 3-4) and the large majority of NPC components belongs to the family of so-called nucleoporins (Nups) (Cook *et al.*, 2007; Stewart, 2007).

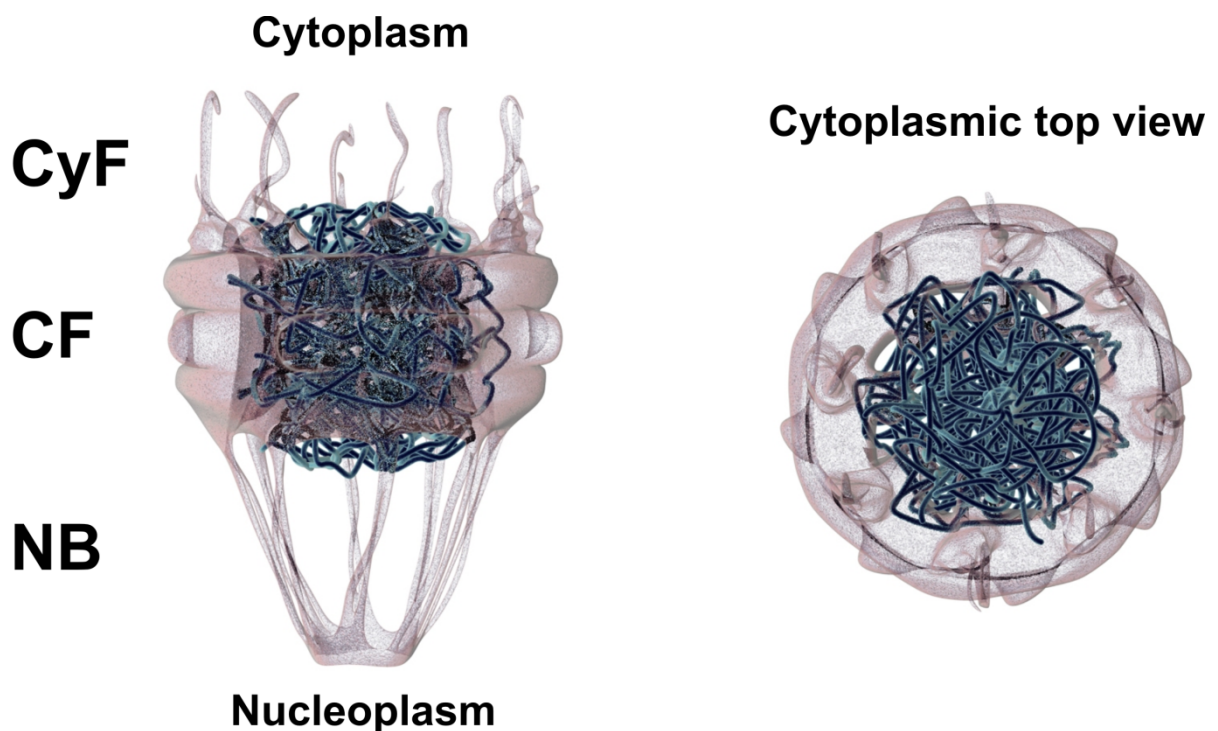


Figure 3-4: Different views of a schematic representation of the nuclear pore complex (NPC). The NPC can be subdivided into cytoplasmic fibrils (CyF), a central framework (CF) spanning the nuclear envelope and the nuclear basket (NB) reaching into the nucleoplasm (left hand side). The eight-fold symmetry of the NPC is clearly visible from the top view from the cytoplasmic side (right hand side). Proteins which are thought to build up and fill the central channel frequently contain phenylalanine-glycine (FG) repeats generating a hydrophobic meshwork shown as dark cyan filaments. In contrast to soluble cellular proteins, transport receptors are able to interact with these FG repeats *via* patches on their surface thereby traversing the NPC. NPC schemes were taken from <http://sspatel.googlepages.com>; © 2000-2006 Samir S. Patel.

Cytoplasmic filaments (CyF) are located on the cytoplasmic side of the NPC (Beck *et al.*, 2004; Beck *et al.*, 2007), they mainly consist of the nucleoporin Nup358 and were shown to interact with diverse proteins like exportins and RanGAP. The central framework (CF) is an assembly of different proteins built around the central pore. Especially the nucleoporins, which constitute this central channel of the NPC, frequently contain phenylalanine-glycine (FG) repeats. The proteins containing FG repeats are thought to generate a hydrophobic meshwork filling the central pore of the NPC and thus preventing the passage of hydrophilic cellular proteins. Conversely, importins and exportins are able to interact *via* specific sites on their surface with the FG repeats of Nups enabling their translocation through the NPC

(Bayliss *et al.*, 2000; Chook and Blobel, 2001; Suntharalingam and Wente, 2003). On the nuclear side of the NPC proteins of the nuclear basket (NB) provide final docking sites for import complexes. Although the overall shape of the NPC is known, the detailed molecular mechanisms, which underlie the NPC passage by transport receptors, are discussed controversially.

3.2 Splicing of mRNA by the spliceosome

In eucaryotic cells mRNA is transcribed from the DNA sequence as a premature precursor (pre-mRNA) by RNA-polymerase II. The pre-mRNA contains both, introns as well as exons. The introns are removed post-transcriptionally by a process called splicing (Brow, 2002; Burge, 1999; Hastings and Krainer, 2001; Wahl *et al.*, 2009; Will, 2006). This multi-step process is catalyzed by the spliceosome, a huge supramolecular complex consisting of various RNA species as well as proteins and may reach a size of up to 2 MDa. This highly dynamic device is assembled during the splicing cycle from different subunits, the so-called UsnRNPs (uridyl-rich small nuclear ribonucleoprotein particles) onto the premature mRNA (Figure 3-5).

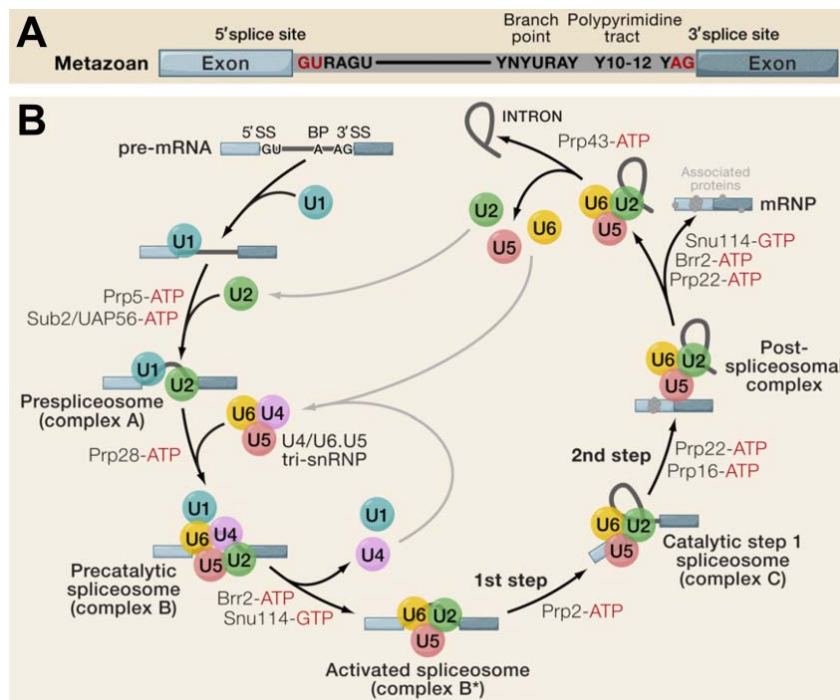


Figure 3-5: Splicing of premature mRNA. (A) Sequence elements of metazoan pre-mRNA. The exons are interrupted by the intronic sequence (gray). 5' and 3' splice sites as well as the branch point sequence are indicated. (B) The splicing cycle by the major spliceosome. The snRNPs U1, U2, U4, U5 and U6 catalyze two transesterification reactions resulting in the ligated exons and the intron lariat. Besides the UsnRNPs, additional non-snRNP proteins like helicases are essential for splicing. Schemes were taken from (Wahl *et al.*, 2009).

UsnRNPs are composed of an snRNA (100-200 bases) as well as different proteins, latter of which can be divided into proteins that are common to all snRNPs and those specific for one particular snRNP. The major spliceosome is responsible for removal of U2-type introns and is assembled from the five UsnRNPs U1, U2, U4, U5 and U6 (Burge, 1999; Hartmuth *et al.*, 2002). Additionally, numerous non-snRNP proteins like helicases mediate or facilitate the necessary RNA-RNA and protein-RNA rearrangements (Figure 3-5) (Ares and Weiser, 1995; Brow, 2002; Madhani and Guthrie, 1994; Staley and Guthrie, 1998).

The definition of introns in pre-mRNAs and therefore the accuracy of splicing merely relies on the recognition of highly conserved intron sequences by UsnRNPs (Wahl *et al.*, 2009). These characteristic sequences have to be arranged in a defined order and distance to each other for the particular introns types. In human U2-type introns the 5' and 3' splice site marks the beginning and the end of an intron, respectively. However, the important sequences are even extended beyond the intron-exon limits (Figure 3-5 A). The branch point adenine, which is mechanistically highly important and a central point in the process of splicing, is located 18-100 nucleotides upstream of the 3' splice site.

Chemically, the process of splicing comprises two consecutive S_N2 like transesterification reactions. The first transesterification is initiated by a nucleophilic attack of the 2' hydroxyl of the branch point adenine onto the phosphate group of the 5' splice site. As a result of this first reaction a free 5' splice site with an accessible 3' OH is formed and second a lariat like structure connected to the 3' splice site is build. In a second transesterification the accessible 3' OH of the 5' site attacks the 3' splice site phosphate. Finally, both exons are connected by a newly formed phosphodiester bond and the lariat like structure consisting of the intron is released and degraded. Interestingly, the two transesterification steps and therefore the splicing reaction itself does not require additional energy, since the two phosphodiester bonds are only rearranged (Wahl *et al.*, 2009).

Spliceosome assembly is initiated by the recognition of the 5' splice site by the U1snRNP (Figure 3-5 B) (Seraphin and Rosbash, 1989). In this initial process the 5' splice site forms base pairs with a specific region of the U1snRNA. Simultaneously, the U2snRNP binds to the branch point and the pre-spliceosomal complex A emerges (Makarov *et al.*, 2002). A previously formed complex, consisting of the three snRNPs U4, U6 and U5, in which U4 and U6 interact by extensive base pairing, joins the pre-spliceosomal complex. The mature spliceosome now referred to as complex B (Malca *et al.*, 2003; Stevens *et al.*, 2002) undergoes several rearrangements and finally represents the activated spliceosome. This

transition includes that U1 and U4 leave the complex as a consequence of which U6 is able to form base pairs with U2 as well as with the 5' splice site (Kim and Lin, 1996; Lesser and Guthrie, 1993; Tarn and Steitz, 1994; Yean and Lin, 1991). The activated spliceosome (complex B*) subsequently catalyzes the first transesterification reaction and results in the formation of complex C (Brow, 2002; Madhani and Guthrie, 1992). After the second transesterification the intron-free mRNA as well as the intron lariat is released and the remaining spliceosome dissociates into the separate components (Jurica and Moore, 2003).

3.2.1 Biogenesis of spliceosomal UsnRNPs

The major constituents of the spliceosome are the UsnRNPs U1, U2, U4, U5 and U6, each of which consists of a small nuclear RNA (snRNA) as well as common and RNP-specific proteins. These proteins are assembled onto the snRNA in a specific and regulated manner (Dickmanns, 2005). Nuclear transport processes play an important role in the biogenesis of spliceosomal UsnRNPs, since their maturation, at least in higher eucaryotes, includes a nucleocytoplasmic transport cycle. [Figure 3-6](#) shows a simplified scheme of the maturation cycle of human U1snRNP. After transcription of the UsnRNA by RNA polymerase II (Hernandez, 2001) in the nucleus the snRNA acquires an m⁷G-cap on its 5'-end (Coppola *et al.*, 1983; Cougot *et al.*, 2004; Shatkin, 1976). This process includes three enzymes organized in a protein complex, namely a triphosphatase, a guanylyltransferase and a N⁷G-methyltransferase (Salditt-Georgieff *et al.*, 1980). First, the triphosphatase catalyzes the hydrolysis of the 5'-mRNA triphosphate followed by the transfer of guanosine monophosphate (GMP) to the existing RNA chain mediated by the guanylyltransferase. This leads to the unusual 5'-5' triphosphate linkage which protects the RNA from degradation by exoribonucleases. Subsequently, the 5' guanine base is methylated on N7 by the N⁷methyltransferase (Shuman, 2002). The attached m⁷G-cap is recognized by the cap binding complex (CBC), which consists of the two cap binding proteins 20 and 80 (CBP20 and CBP80), respectively (Ohno *et al.*, 1990). The CBC stacks the positively charged m⁷guanine between the two aromatic amino acid residues Tyr20 and Tyr43 of CBP20 (Calero *et al.*, 2002; Mazza *et al.*, 2002). The phosphorylated adapter for RNA export (PHAX) bridges the interaction between the CBC-snRNA complex and the actual export receptor CRM1 (Ohno *et al.*, 2000; Segref *et al.*, 2001). The directionality of this transport process is ensured by the small GTPase Ran in its GTP bound form, which is present in high concentration in the nucleus. In the cytoplasm, CRM1 dissociates from the complex due to the hydrolysis of the

GTP molecule bound to Ran and the dephosphorylation of PHAX. The next step is the assembly of the seven Sm-proteins (B/B', D1, D2, D3, E, F, G) by the SMN (survival of motor neurons)-complex onto the snRNA (Chari *et al.*, 2009). The SMN-complex with bound sub-complexes of the Sm-proteins B/D3 and D1/D2/E/F/G mediates the assembly and ring closure generating the typically doughnut shaped core-snRNP (Kambach *et al.*, 1999).

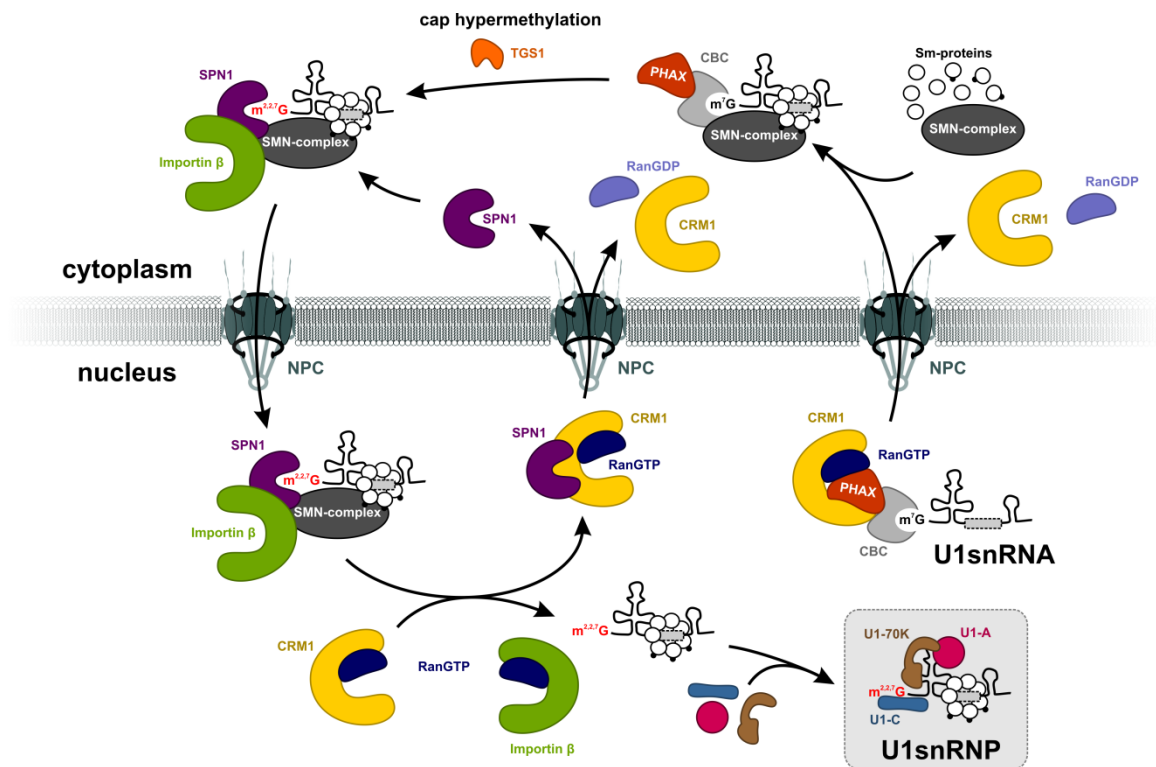


Figure 3-6: Schematic representation of the biogenesis cycle of human spliceosomal U1snRNP. The transcribed U1snRNA is exported to the cytoplasm by CRM1 (chromosome region maintenance 1). The cap binding complex (CBC) as well as the phosphorylated adapter for RNA export (PHAX) participate in this process. In the cytoplasm the assembly of seven Sm-proteins is mediated by the SMN complex. Subsequently, the dimethyltransferase TGS1 (Trimethylguanosine Synthase 1) catalyzes the hypermethylation of the m⁷G-cap. The generated m^{2,2,7}G (m₃G) cap is specifically recognized by the import adapter SPN1 (snurportin 1), which binds to the import receptor importin β. In the nucleus, the U1snRNP specific proteins U1-A, U1-C and U1-70K bind to the snRNP-core and complete its maturation. SPN1 is recycled to the cytoplasm by the export receptor CRM1 bound to RanGTP.

Prior to assembly, the molecular chaperone pICln transfers the joined sub-complexes to the SMN-complex (Chari *et al.*, 2008). The assembled core-snRNP was shown to be a prerequisite *in vivo* for the subsequent cap hypermethylation (Massenet *et al.*, 2002; Mattaj, 1986; Plessel *et al.*, 1994; Raker *et al.*, 1996). After dissociation of PHAX and CBC the dimethyltransferase TGS1 (Trimethylguanosine Synthase 1) catalyzes the addition of two further methyl groups to the guanine N2 of the m⁷G-cap (Mouaikel *et al.*, 2002; Plessel *et al.*, 1994; Raker *et al.*, 1996). TGS1 utilizes *S*-adenosyl-L-methionine, the second most commonly used enzyme substrate, as cofactor and methyl group donor in the reaction (Hausmann and

Shuman, 2005). It was shown that TGS1 interacts with the Sm-proteins B and D1 as well as with the major component of the SMN-complex, the SMN-protein (SMNp) (Mouaikel *et al.*, 2003b; Mouaikel *et al.*, 2002). The hypermethylated cap $m^{2,2,7}G$ in turn is bound by snurportin 1 (SPN1), the import adapter for spliceosomal UsnRNPs (Huber *et al.*, 1998; Strasser *et al.*, 2005). *Via* its IBB domain SPN1 binds to the import receptor importin β , which enables nuclear import (Mitrousis *et al.*, 2008; Palacios *et al.*, 1997; Wohlgend *et al.*, 2007). Subsequently, the import complex dissociates in the nucleoplasm upon binding of RanGTP to importin β and SPN1 stays bound to the core-snRNP (Huber *et al.*, 2002). In the nucleus, the U1 specific proteins U1-A, U1-70K and U1-C bind to the core-snRNP and finally the assembled and mature snRNPs are stored in the nuclear cajal bodies (Sleeman *et al.*, 2001; Sleeman and Lamond, 1999; Will and Luhrmann, 2001). The further ways of the readily assembled snRNPs to the sites of spliceosome assembly are still largely unknown.

3.3 The cap dimethyltransferase TGS1

The Trimethylguanosine Synthase 1 (TGS1) catalyzes the *S*-adenosyl-L-methionine dependent dimethylation of the m^7G -cap of spliceosomal core-UsnRNPs in their cytoplasmic stage. The reaction is thought to proceed by an S_N2 -like mechanism (Figure 3-7) (Schubert *et al.*, 2003).

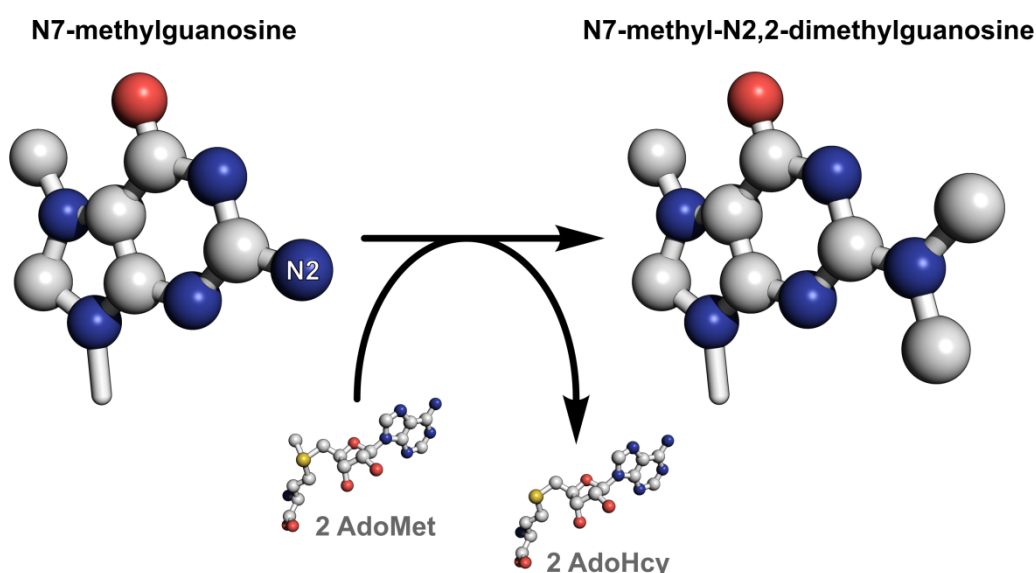


Figure 3-7: Reaction catalyzed by the Trimethylguanosine Synthase 1 (TGS1). The methylation most likely proceeds by an S_N2 -like mechanism. The N7-methylguanosine-cap (m^7G) is methylated twice on the N2 atom generating the N7-methyl-N2,2-dimethylguanosine-cap ($m^{2,2,7}G$; m_3G). The methyl group donor is the most commonly used methylation cofactor *S*-adenosyl-L-methionine (AdoMet), which is converted to *S*-adenosyl-L-homocysteine during the reaction. In the reaction scheme only the guanine base of the more complex 5'-cap structure is shown.

Additionally, the hypermethylation of the m⁷G-caps of several small nucleolar RNAs (snoRNAs) (Maxwell and Fournier, 1995) and the telomerase RNA TLC1 (Franke *et al.*, 2008) was found to depend on the catalytic activity of the dimethyltransferase as well. In the catalyzed reaction the enzyme converts the N7-methylguanosine (m⁷G) cap to the N7-methyl-N2,2-dimethylguanosine (m^{2,2,7}G; m₃G) cap by using *S*-adenosyl-L-methionine as methyl group donor.

TGS1 orthologs have been found and biochemically described to different extent in numerous organisms including *Homo sapiens*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Xenopus laevis*, *Trypanosoma brucei* and *Giardia lamblia* (Colau *et al.*, 2004; Enunlu *et al.*, 2003; Franke *et al.*, 2008; Girard *et al.*, 2008; Hausmann *et al.*, 2007; Hausmann and Shuman, 2005; Hausmann *et al.*, 2008; Mouaikel *et al.*, 2003a; Mouaikel *et al.*, 2002; Ruan *et al.*, 2007; Zhu *et al.*, 2001). The enzyme was first discovered in *H. sapiens* as PRIP-interacting protein with methyltransferase domain (PIMT) (Zhu *et al.*, 2001). It is located on chromosome 8q11, spans more than 40,000 base pairs and consists of 13 exons. In a yeast-two-hybrid screen it was found to interact with the protein PRIP (peroxisome proliferator-activated receptor-(PPAR)-interacting protein) and by this interaction, to activate the transcription factor PPAR (peroxisome proliferator-activated receptor) indirectly. PPAR belongs to the superfamily of nuclear receptors, which are known to be involved in the regulation of gene expression *via* binding to DNA response elements. It was shown that for this interaction only the N-terminal part of PIMT is needed, while the C-terminal methyltransferase domain seems to be dispensable for this function of the protein (Figure 3-8).

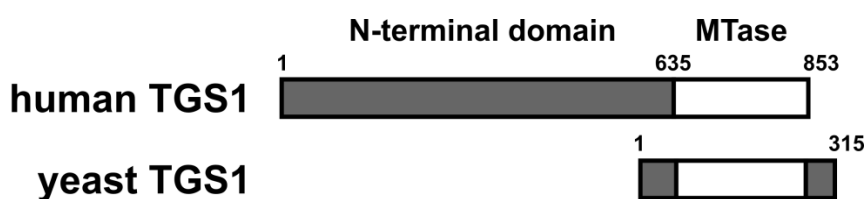


Figure 3-8: Domain organization of human and yeast TGS1. *H. sapiens* TGS1 (hTGS1) consists of a large N-terminal domain, which is involved in transcriptional coactivation and the conserved C-terminal methyltransferase domain (MTase). In contrast, the yeast ortholog only comprises the MTase domain and the short extensions on the N- and C-terminus were shown to be dispensable for methyltransferase activity.

Additionally, human PIMT was found to interact with the transcription coactivator proteins CBP (CREB binding protein), p300 (adenovirus E1A-binding protein p300) and PBP (PPAR-binding protein) (Misra *et al.*, 2002). However, the exact role of PIMT in this cascade still has not been fully described *in vivo* and remains elusive. The C-terminal methyltransferase

domain of human PIMT was shown to bind the most commonly used methyl group donor *S*-adenosyl-L-methionine (Zhu *et al.*, 2001). Hereafter, human PIMT is referred to as human TGS1 (hTGS1) to clarify its function in snRNP biogenesis and its relation to the characterized yeast enzyme. *S. cerevisiae* TGS1 (yTGS1) was identified in a yeast-two-hybrid screen and its ability to hypermethylate a 5'-cap analog was shown in a radioactive methylation assay (Mouaikel *et al.*, 2002). The yeast enzyme shares the conserved methyltransferase domain with hTGS1, while the extended N-terminal domain of hTGS1 is lacking (Figure 3-8). Therefore it appears likely that the N-terminal domain was fused with the C-terminal methyltransferase domain during evolution. Knockout of yeast TGS1 results in a cold sensitive phenotype as mutants show significant slower growth at lower temperatures and a mild splicing defect (Mouaikel *et al.*, 2002). In contrast, knockout of the *D. melanogaster* ortholog DTL (*d*rosophila-*t*at-*l*ike) led to severe defects in development, as the larvae died in early pupal stages (Komonyi *et al.*, 2005). The sequence identity of the MTase domains of human and yeast TGS1 proteins (aa635-853 in hTGS1 and aa1-315 in yTGS1), which were shown to be true functional orthologs (Hausmann *et al.*, 2008), amounts to 38% indicating a closely related fold. As shown by sequence and secondary structure analyses, the dimethyltransferase TGS1 belongs to the class I of methyltransferases (MTases) (Mouaikel *et al.*, 2003a). This subgroup is characterized by a central seven-stranded β -sheet with six strands in a parallel and the last one in an antiparallel orientation (Schubert *et al.*, 2003). The slightly twisted β -sheet of class I MTases is flanked by a variable number of α -helices on each side, generating the typical $\alpha\beta\alpha$ sandwich. Although secondary structure elements and the protein family can be predicted, the structural description of the protein itself as well as substrate binding and the catalytic mechanism are still unknown. Furthermore, it remains elusive how TGS1 is able to recognize and hypermethylate a variety of different RNA species such as snRNAs, snoRNAs and telomerase RNA, while the m⁷G-caps of other RNA types (e.g. mRNA) are protected.

3.4 The nuclear export complex CRM1·SPN1·RanGTP

CRM1 (chromosome region maintenance 1) is one of the most versatile export receptors as it exports hundreds of proteins and RNAs reaching from small molecules to huge macromolecular assemblies such as ribosomal subunits (Bohnsack *et al.*, 2002; Fornerod *et al.*, 1997; Gadgil *et al.*, 2001; Johnson *et al.*, 2002; Kutay and Guttinger, 2005; Moy and Silver, 2002; Stade *et al.*, 1997). During biogenesis of spliceosomal UsnRNPs the nuclear

export receptor CRM1 fulfills two distinct functions. First, the newly transcribed UsnRNA is exported by CRM1 (Fornerod *et al.*, 1997; Ohno *et al.*, 2000) and second, the import adapter SPN1 mediating the import of the core-snRNP is recycled back to the cytoplasm by CRM1 as well (Paraskeva *et al.*, 1999).

3.4.1 The nuclear export receptor CRM1

Initially, CRM1 was identified in *S. pombe* where mutations in the gene led to deformed nuclear chromosome domains (Adachi and Yanagida, 1989). The observed effects result from the function of CRM1 as major nuclear export receptor as well as from its additional function during mitosis. As a protein, which is composed out of HEAT repeats it belongs to the large importin β superfamily of transport receptors. CRM1 was found to interact with cargoes *via* the leucine rich nuclear export signal (LR-NES) (Fornerod *et al.*, 1997). This NES has been identified in various potential export cargoes like the viral HIV-Rev protein (Fischer *et al.*, 1995) or the protein kinase A inhibitor (PKI) (Wen *et al.*, 1995) and typically contains four characteristically spaced hydrophobic residues (Φ) (Fornerod and Ohno, 2002; la Cour *et al.*, 2004). The Φ -residues are basically leucines but can be in principle also isoleucine, valine, phenylalanine or methionine and are interrupted by a defined number of variable amino acids (Figure 3-9 A).

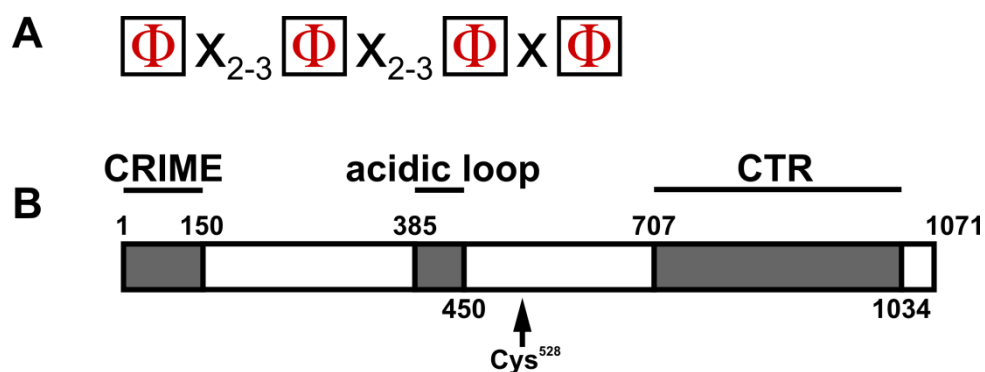


Figure 3-9: Characteristics of CRM1 (chromosome region maintenance 1). (A) Typical arrangement and spacing of hydrophobic residues within leucine rich nuclear export signals (LR-NES). Hydrophobic residues Φ (Leu, Ile, Val, Phe or Met) are characteristically interrupted by variable residues x. (B) Domain organization of CRM1. The N-terminal CRIME domain (CRM1, importin β , etc.) which shows significant similarity to importin β includes the amino acids 1-150. The acidic loop is predicted to be inserted within HEAT repeat 8 and comprises residues 385-450. The C-terminal fragment (CTR; amino acids 707-1034) has been crystallized and comprises six HEAT repeats. The position of cysteine 528, to which the CRM1 specific inhibitor leptomycin B (LMB) binds covalently, is indicated.

CRM1 comprises 1071 amino acids in *M.musculus* and *H. sapiens* and 1084 residues *S. cerevisiae*. It is predicted to be composed out of 19 HEAT repeats, which are separated by

short loops (Petosa *et al.*, 2004). An N-terminal CRIME domain (CRM1, importin β , etc.) spanning from amino acids 1-150 shares sequence homology with importin β (Figure 3-9 B). Importantly, the loop inserted in between the A- and B-helix of HEAT repeat 8 contains a remarkable high number of acidic amino acids and therefore is referred to as acidic loop (Petosa *et al.*, 2004). This 65 amino acids long loop is believed to be involved in binding of RanGTP, as a corresponding region in transportin 1 interacts with a basic patch of RanGTP (Chook and Blobel, 1999).

The C-terminal CRM1 region containing the residues 707-1027 has been characterized by means of X-ray crystallography (Petosa *et al.*, 2004). The crystal structure revealed that it comprises the six predicted C-terminal HEAT repeats 14-19. Between the HEAT repeats 18 and 19 an additional linker helix changes the directionality of the subsequent A and B helices of HEAT 19. Besides the crystallographic characterization of the last 6 repeats, cryo-EM studies showed that CRM1 without cargo and RanGTP adopts a ring like structure (Petosa *et al.*, 2004). Hence, it is thought that this ring may open upon RanGTP and cargo binding changing the overall structure of CRM1 dramatically. Despite intensive efforts to unravel the molecular details for export complex assembly and disassembly and the structural reasons for the very broad cargo spectrum of CRM1, no structural information defining these characteristics are available so far.

3.4.2 The snRNP import adapter SPN1

The import adapter snurportin 1 (SPN1) recognizes one part of the bipartite nuclear import signal of the UsnRNP-core prior to nuclear import by importin β (Huber *et al.*, 1998). SPN1 binds the m³G-cap subsequent to the hypermethylation by TGS1. *H. sapiens* SPN1 has a molecular weight of 42 kDa and contains two characterized domains (Figure 3-10). There is an N-terminal importin β binding domain (IBB) encompassing the 65 N-terminal residues, which is essential for the recognition by importin β and a central cap binding domain (CBD) (Figure 3-10, lower panel in green). The CBD comprises the residues 91-301 and is composed of two β -sheets and several α -helices (Strasser *et al.*, 2005). The cocrystallized m³G-cap is bound between the two β -sheets mainly involving the residues Trp276, Trp107, Leu104 and Ser105. Both nucleotides of the cap analog are bound in a stacked conformation and the m^{2,2,7}guanine is flanked by Trp276. It has been extensively discussed that CRM1 does not bind to SPN1 *via* a classical short leucine rich NES but rather by a large domain including N-terminal and central as well as C-terminal residues (Paraskeva *et al.*, 1999). Although

structural as well as biochemical information are available for the CBD and the IBB domain the binding mode of SPN1 to CRM1 remained completely unknown.

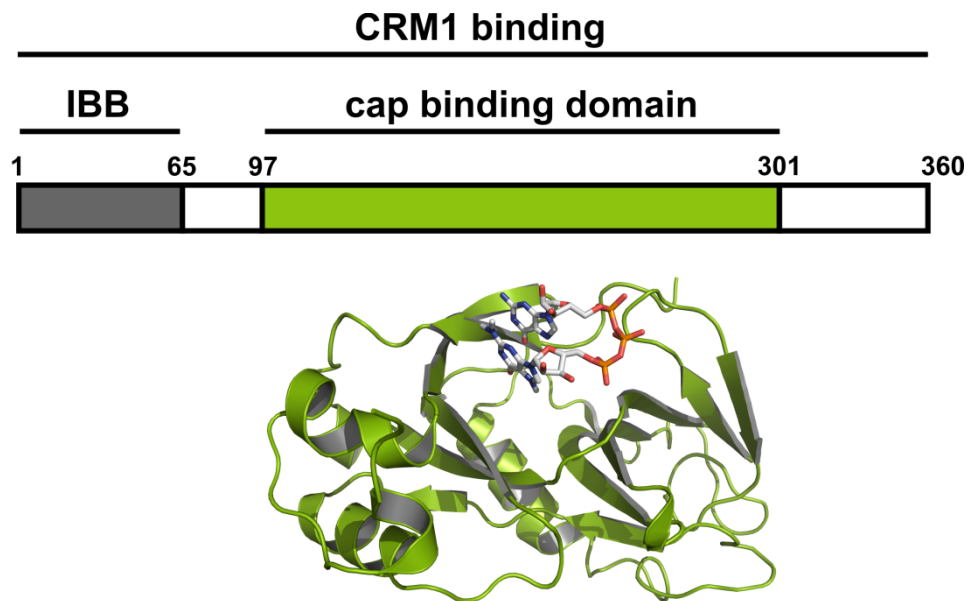


Figure 3-10: Domain organization of human SPN1. The protein comprises 360 amino acids. While the N-terminal 65 residues belong to the importin β binding domain (IBB) of SPN1 residues 97-301 represent the cap binding domain (CBD). It was shown that N- and C-terminal parts of SPN1 are required for high affinity CRM1 binding. A cartoon representation of the crystal structure of the CBD (green) bound to an m_3G -cap analog (carbons in light gray) is shown in the lower panel (Strasser *et al.*, 2005).

3.5 Deadenylation dependent mRNA decay

The compartmentalization of eucaryotic cells allows a differentiated and spatial separated regulation of gene expression in response to changes of the environmental conditions. The mechanisms include regulation of transcription, splicing, mRNA export and translation as well as numerous protein modifications in order to modulate their activity. Another possibility is the post-transcriptional regulation of mRNA stability, which depends on the ratio of mRNA synthesis and decay (Garneau *et al.*, 2007; Isken and Maquat, 2007; Maquat and Carmichael, 2001; Mitchell and Tollervy, 2001). Consequently, the shift of this equilibrium of a given mRNA allows a highly flexible and fast response mechanism to changing environmental conditions and signals. It is therefore not surprising that many regulators and mechanisms exist to control and modulate the rate of mRNA degradation. Most frequently, the mRNA decay in eucaryotes is initiated by 3'-poly(A) tail shortening, which is catalyzed by poly(A)-specific exoribonucleases (Coller and Parker, 2004; Parker and Song, 2004). [Figure 3-11](#)

shows the deadenylation dependent mechanisms of mRNA degradation in eucaryotes in both, the 5'→3' and the 3'→5' direction.

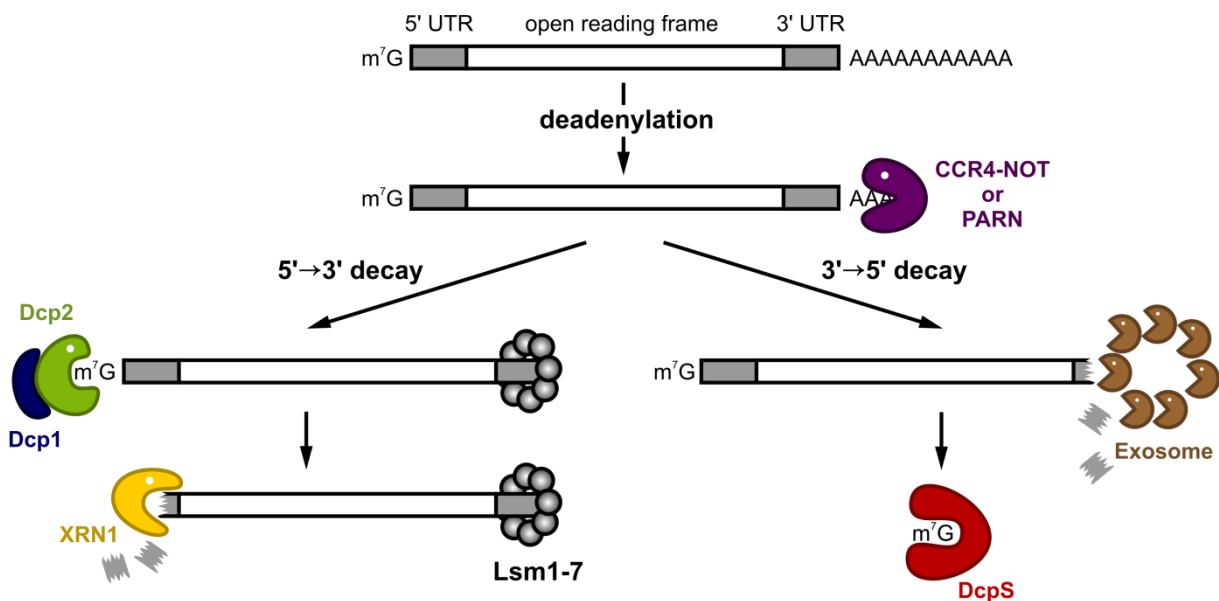


Figure 3-11: Mechanisms of mRNA degradation by the deadenylation dependent pathway in eucaryotes. Initially, the poly(A) tail is removed by deadenylases (CCR4-NOT complex or PARN). Subsequent deadenylation, two possible pathways lead to either the 5'→3' or 3'→5' degradation of the RNA. 5'→3' degradation is initiated by removal of the 5'-cap by the decapping enzyme consisting of the heterodimer Dcp1/Dcp2 and the Sm-like-complex (LSm) is associated with the 3' end of the mRNA. The mRNA then is susceptible to decay by the exoribonuclease XRN1. In the 3'→5' direction the mRNA is degraded by the exosome and the remaining 5'-cap is hydrolyzed by the scavenger-decapping enzyme DcpS. Modified scheme according to Garneau *et al.*, 2007.

3.5.1 The poly(A)-specific ribonuclease (PARN)

Independent of the direction of mRNA decay the first and initial step in both pathways is the deadenylation of the mRNA by the CCR4-NOT-complex or the poly(A)-specific ribonuclease (PARN) (Garneau *et al.*, 2007). PARN, which was initially named deadenylating nuclease (DAN) (Korner and Wahle, 1997), is unique compared to other deadenylases present in the cell as it not only binds the poly(A) tail but also the mRNA 5'-cap (Dehlin *et al.*, 2000; Gao *et al.*, 2000; Martinez *et al.*, 2001). Although a PARN ortholog is present in most eucaryotes, it is lacking in *S.cerevisiae* and *D. melanogaster*. Nevertheless, it is the major and most important deadenylase in human cells (Korner and Wahle, 1997) and additionally seems to be involved in the nonsense mediated mRNA decay (NMD) (Lejeune *et al.*, 2003). Human PARN is a 74 kDa protein and can be subdivided into three distinct domains (Figure 3-12) (Korner and Wahle, 1997; Nilsson *et al.*, 2007; Wu *et al.*, 2005). The nuclease domain encompassing the residues 1-391 mediates nucleolytic activity of PARN and a single-stranded nucleic acid binding R3H domain is inserted into the nuclease domain with respect to the

primary amino acid sequence. The structure of both associated domains was solved by means of X-ray crystallography with and without bound substrates illuminating the catalytic mechanism of the enzyme (Figure 3-12, lower panel) (Wu *et al.*, 2005). Besides the nuclease domain the R3H domain plays an important role in the binding of the mRNA poly(A) tail. The crystal structure revealed that PARN forms a homodimer, whereas mainly the nuclease domain contributes to the hydrophobic dimer interface. Interestingly, it was shown that the enzyme activity relies on dimer formation and the poly(A) tail binding sites are located antipodal with respect to each other (Figure 3-12). The R3H domain of one monomer thereby covers the poly(A) binding site of the opposing monomer and contributes to poly adenine binding.

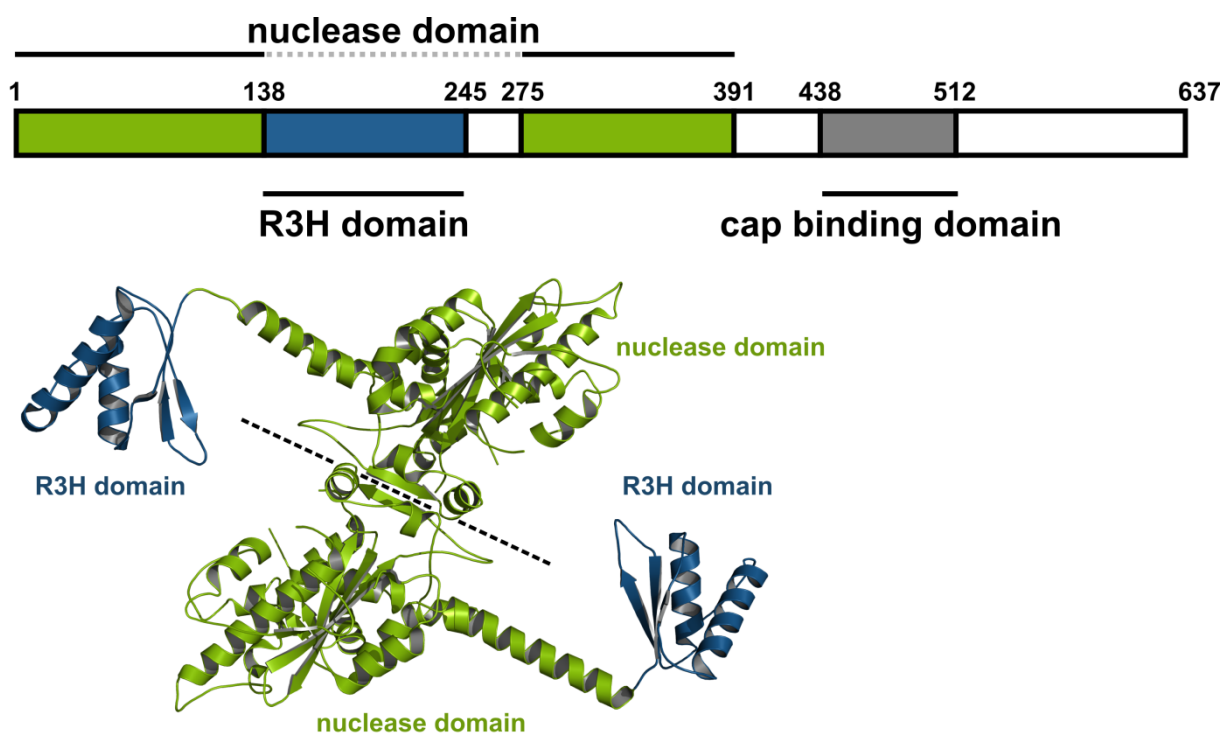


Figure 3-12: Domain organization of human poly(A)-specific ribonuclease (PARN). The protein comprises 637 amino acids and is divided into the nuclease (green), the R3H (blue) and the cap binding domain (gray). The lower panel shows a homodimer of the substrate free form of the human PARN nuclease-R3H domains (Wu *et al.*, 2005). The color code is as in the domain scheme (upper panel) and the dimer interface is indicated by a dashed line.

The C-terminal cap binding domain (CBD) encompassing the residues 438-512 is known to adopt an RNA recognition fold (RRM) and to bind the 5'-m⁷G-cap of the mRNA (Copeland and Wormington, 2001). Cap binding does not only stimulate the nucleolytic activity of the enzyme but it also enhances the processivity of the deadenylation reaction (Dehlin *et al.*, 2000; Gao *et al.*, 2001; Martinez *et al.*, 2001). It has been speculated that cap binding may induce a conformational change, which in turn could be responsible for stimulation. Although

the structure of the *M. musculus* PARN-RRM has been solved by means of nuclear magnetic resonance (NMR) the interaction mode with the cap remained unknown, since the structure lacks a bound m⁷G-cap (unpublished data; PDB ID 1WHV).

Cap binding by proteins is generally known to be mediated by two aromatic or hydrophobic residues, with the m⁷guanine stacked in between (Hsu *et al.*, 2000; Quijcho *et al.*, 2000) and it was thought that PARN may also follow this common strategy. The cap-free structure shows that there are two tryptophan residues (Trp449 and Trp468) in such proximity that they could interact and bind the cap in between them. However, the tryptophans are oriented antipodal with respect to each other and in order to bind the cap base they would have to change their conformations dramatically. This raised the question whether cap binding may induce such a tryptophan flip or even applies to a completely different mode.

Chapter 4 • Crystal structure of the conserved methyltransferase domain of TGS1

This manuscript has originally been published in *Acta Crystallographica Section D: Biological Crystallography*.

'Structure analysis of the conserved methyltransferase domain of human trimethylguanosine synthase TGS1'

Thomas Monecke¹, Achim Dickmanns¹, Anja Strasser² and Ralf Ficner^{1,‡}

¹ Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik, Abteilung für Molekulare Strukturbiologie (Göttinger Zentrum für Molekulare Biowissenschaften (GZMB), Ernst-Caspari-Haus), Justus-von-Liebig-Weg 11, D-37077, Göttingen

² Present address: Anja Strasser, Roche Diagnostics GmbH, Bereich Pharma Produktion, Im Nonnenwald 2, D-82377, Penzberg

‡ To whom correspondence should be addressed: E-mail: rficner@uni-goettingen.de
Phone: 0551 – 39 14071
Fax: 0551 – 39 14082

Keywords: dimethyltransferase; m⁷G; m^{2,2,7}G; cap, snRNA

Acta Crystallographica Section D; VOL. 65, NO. 4, pp. 332-38, April, 2009
© 2009 by International Union of Crystallography.

Received December 11, 2008; Accepted January 24, 2009

Preface – About the Manuscript

The Trimethylguanosine Synthase 1 (TGS1) is the enzyme responsible for the hypermethylation of the 5'-m⁷G-cap of certain small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and the telomerase RNA TLC1. While this enzyme, present in multiple organisms, is biochemically well characterized, its three-dimensional crystal structure has been unknown thus evading the characterization of its binding mode to the substrates as well as its catalytic mechanism.

The aim of this study was the structure determination of a fragment of the conserved methyltransferase domain of human TGS1 comprising the residues 653-853. As the crystal structure presented in this manuscript reveals an incomplete cap binding pocket, a non-radioactive activity assay was developed in order to investigate the catalytic activity and efficiency of the purified TGS1 fragment. Interestingly, the defective cap binding pocket results in a completely inactive enzyme and a maximum of 17 additional N-terminal amino acid residues is required to gain catalytic activity.

The coordinates and structure factors of the crystal structure described in the following publication have been deposited in the protein data bank (PDB) under the PDB ID 3EGI.

Author contributions:

Dr. Anja Strasser, Dr. Achim Dickmanns, Prof. Dr. Ralf Ficner and myself initiated the project and designed the research. My further contribution comprised molecular cloning, protein purification and crystallization as well as structure determination and analysis under supervision and guidance of all co-authors. Furthermore, Stephanie Schell contributed to the work in terms of protein purification and crystallization within a laboratory practical course.

Abstract

Methyltransferases play an important role in the post-transcriptional maturation of most ribonucleic acids. The modification of spliceosomal UsnRNAs includes N2-dimethylation of the m⁷G cap catalyzed by trimethylguanosine synthase 1 (TGS1). This 5'-cap hypermethylation occurs during the biogenesis of UsnRNPs as it initiates the m₃G cap-dependent nuclear import of UsnRNPs. The conserved methyltransferase domain of human TGS1 has been purified, crystallized and the crystal structure of this domain with bound substrate m⁷GpppA was solved by means of multiple wavelength anomalous dispersion. Crystal structure analysis revealed that m⁷GpppA binds *via* its adenosine moiety to the structurally conserved adenosylmethionine-binding pocket, while the m⁷guanosine remains unbound. This unexpected binding only occurs in the absence of AdoMet and suggests an incomplete binding pocket for the m⁷G cap which is caused by the N-terminal truncation of the protein. These structural data are consistent with the finding that the crystallized fragment of human TGS1 is catalytically inactive, while a fragment that is 17 amino acids longer exhibits activity.

1. Introduction

S-Adenosyl-L-methionine (AdoMet) dependent methyltransferases (MTases) are involved in many different cellular processes including the post-transcriptional modification of RNAs. Some 74 different methylated RNA nucleosides have been identified in the three kingdoms of life. The methylation of guanosine concerns atoms N1, N2, N7 and 2'O and various combinations of these methylations have been found (m¹G, m²G, m⁷G, Gm, m^{2,2}G, m^{2,7}G, m^{2,2}Gm, m^{2,2,7}G, m¹Gm, m^{2,7}Gm; Limbach *et al.*, 1994). All RNA methylations are introduced post-transcriptionally by AdoMet-dependent MTases, most of which belong to the class I MTases, which are characterized by a Rossmann-fold-like αβ structure (Schubert *et al.*, 2003).

One member of this family is the trimethylguanosine synthase 1 (TGS1), which catalyzes the N2-dimethylation of the m⁷G cap of spliceosomal uridyl-rich small nuclear RNAs (UsnRNAs) and of some small nucleolar RNAs (snoRNAs) (Hausmann & Shuman, 2005a; Maxwell & Fournier, 1995; Mouaikel, Bujnicki *et al.*, 2003; Mouaikel *et al.*, 2002). TGS1 enzymes from *Saccharomyces cerevisiae* (yTGS1), *Giardia lamblia*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Trypanosoma brucei* and human

cells (hTGS1) have been characterized with respect to their biochemical properties as well as their interaction with UsnRNPs or snoRNPs (small nuclear/nucleolar ribonucleoprotein particles) (Colau *et al.*, 2004; Enunlu *et al.*, 2003; Girard *et al.*, 2008; Gunzl *et al.*, 2000; Hausmann *et al.*, 2007, 2008; Hausmann & Shuman, 2005a,b; Komonyi *et al.*, 2005; Misra *et al.*, 2002; Mouaikel, Bujnicki *et al.*, 2003; Mouaikel, Narayanan *et al.*, 2003; Mouaikel *et al.*, 2002; Plessel *et al.*, 1994; Ruan *et al.*, 2007; Watkins *et al.*, 2004; Zhu *et al.*, 2001). There is a significant difference between organisms regarding the size of TGS1, which varies from 239 residues in *S. pombe* to 853 residues in *Homo sapiens*, as well as its cellular localization. Yeast TGS1 was shown to act in the nucleus exclusively (Mouaikel *et al.*, 2002), whereas hTGS1 methylates spliceosomal UsnRNAs in the cytoplasm and snoRNAs in the nucleus (Colau *et al.*, 2004; Mouaikel, Narayanan *et al.*, 2003; Verheggen *et al.*, 2002).

UsnRNA 5'-cap hypermethylation plays an important role during the biogenesis of UsnRNPs (Dickmanns & Ficner, 2005). In higher eukaryotes, the maturation of UsnRNPs comprises a nucleocytoplasmic transport cycle. Newly transcribed snRNAs U1, U2, U4 and U5 are exported to the cytoplasm in an m⁷G cap-dependent manner, where assembly with seven

Sm proteins occurs (Hamm *et al.*, 1990; Mattaj, 1986). This assembly process is mediated by the survival of motor neuron complex (SMN complex), a large multiprotein complex (Neuenkirchen *et al.*, 2008). The m⁷G cap is subsequently hypermethylated by TGS1 and the resulting m₃G cap is recognized by the nuclear import adaptor snurportin1 (Huber *et al.*, 1998; Strasser *et al.*, 2005), which binds to the general nuclear import receptor importin β . Hence, the m₃G cap serves as nuclear import signal that indicates the completed assembly of the core UsnRNP particle. Therefore, the interaction of TGS1 with UsnRNP proteins SmB/B' and D1 as well as with the SMN complex appears to correlate with the ordered process of RNP assembly and subsequent cap hypermethylation (Mouaikel, Narayanan *et al.*, 2003; Mouaikel *et al.*, 2002).

Biochemical studies have revealed that TGS1 is specific for m⁷G-capped RNA and m⁷GTP, which represents the minimal substrate, while nonmethylated 5'-cap RNA or GTP are not N₂-dimethylated (Hausmann & Shuman, 2005a,b; Hausmann *et al.*, 2008). TGS1 catalyzes two successive methyl-transfer reactions using AdoMet as a methyl-group donor, which includes the formation of the intermediate product N₂, N₇-dimethylguanosine (Hausmann & Shuman, 2005a; Hausmann *et al.*, 2008). Furthermore, a three-dimensional structure model of yTGS1 was generated by means of homology modeling, based on which the m⁷G cap-binding pocket was predicted (Mouaikel, Bujnicki *et al.*, 2003).

In order to verify the proposed structural model and the mode of m⁷G cap recognition, we have crystallized the predicted MTase domain of hTGS1 in the presence of the substrate dinucleotide m⁷GpppA. Our crystal structure analysis and additional biochemical studies demonstrate that the predicted MTase domain is catalytically inactive owing to a lack of m⁷G cap binding. We show that additional N-terminal residues enlarging the canonical MTase domain are required for enzymatic activity.

2. Materials and methods

2.1. Protein expression and purification

Human TGS1 fragments (amino acids 636–853 and 653–853) were subcloned from pGEX-6P-1 full-length TGS1 (accession No. Q96RS0) into the *Bam*HI/*Xho*I sites of pGEX-6P-3 (GE Healthcare, Germany) and verified by sequencing. The following primers were used for subcloning: MT636_forward, 5'-CGCGGA **TCCCCTGAAATAGCTGCTGTTCTGAGC**-3' (*Bam*HI site in bold), MT653_forward, 5'-CGCGGATCCAGGCTCTTCTCCCGTTTT GATG-3' (*Bam*HI site in bold), and MT853_reverse, 5'-CCGCTCGAGTTAGG TTTCAGAGGCTGGTCTTCG-3' (*Xho*I site in bold).

The native GST-fusion constructs for the activity tests were expressed in *Escherichia coli* BL21 (DE3) (Invitrogen, USA) at 289 K in ampicillin-containing 2YT medium, which was supplemented with 2% (w/v) glucose. Expression of constructs was induced at OD₆₀₀ = 0.8, adding IPTG to a final concentration of 500 μ M. Immediately after induction, 2%(v/v) ethanol and 50 mM K₂HPO₄ were added to the growth culture. The cells were harvested after 18 h by centrifugation (5000g, 20 min, 277 K) and resuspended in lysis buffer containing 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM EDTA and 2 mM DTT. All subsequent steps were carried out at 277 K unless stated otherwise. Cells were disrupted using a 110S microfluidizer (Microfluidics, USA). The clarified lysate (30 000g, 30 min, 277 K) was subsequently loaded onto a GSH-Sepharose column (GE Healthcare, Germany), which was equilibrated with lysis buffer. Unbound proteins were removed by washing with two column volumes (CV) of lysis buffer. In order to eliminate RNA contaminants, the loaded column was washed with one CV of a high-salt buffer containing 50 mM Tris-HCl pH 7.5, 1 M NaCl and 2 mM DTT. After re-equilibration in lysis buffer, the bound fusion protein was eluted with lysis buffer additionally containing 25 mM reduced glutathione. GST-hTGS1_{653–853} was incubated with PreScission

protease (GE Healthcare, Germany) at 277 K overnight in order to cleave the fusion protein into GST and hTGS1_{653–853} containing a multiple cloning site remainder of Gly-Pro-Leu-Gly-Ser at the N-terminus. hTGS1_{653–853} was further purified using a Superdex S75 (26/60) gel-filtration column (GE Healthcare, Germany) equilibrated in a buffer containing 20 mM Tris–HCl pH 7.5, 200 mM NaCl and 2 mM DTT. The elution volume of hTGS1_{636–853} and hTGS1_{653–853} on the gel-filtration column corresponded to a monomeric state of the protein. The resulting pure protein was concentrated to 6 mg ml^{−1} using a Vivaspin concentrator with MWCO 10 000 Da (Sartorius, Germany) and stored in aliquots at 193 K.

The selenomethionine-containing human TGS1 fragment encompassing amino acids 653–853 was expressed according to the protocol described by Reuter & Ficner (1999). The purification of SeMet-TGS1_{653–853} was performed as described for the native proteins with the exception that the DTT concentration was elevated to 5 mM in all buffers in order to prevent oxidation of the selenium.

2.2. HPLC-based activity assay

In order to analyze the activity of the purified human TGS1 fragments, an HPLC-based activity assay was developed. 25 µM purified protein was mixed with 0.5 mM cap analogue m⁷GpppA (KEDAR, Poland) and 2 mM AdoMet (Sigma–Aldrich, Germany) in 1 × PBS. The reaction mixture, with a total volume of 10 µl, was incubated at 310 K and the reaction was stopped by addition of 1 µl 1 M HClO₄ and incubation on ice for 1 min. The solution was neutralized by adding 20 µl 2 M sodium acetate. Precipitated protein was pelleted by centrifugation (16 000g, 10 min, 293 K) and the supernatant was loaded onto a reversed-phase HPLC column (Prontosil C18-AQ, Bischoff Chromatography, Germany), which was equilibrated in phosphate buffer A containing 100 mM K₂HPO₄/KH₂PO₄ pH 6.5. The substrates and products of the reaction were eluted from the column by applying a linear

gradient from 0 to 60% buffer B, which consisted of buffer A with an additional 50% (v/v) acetonitrile. Commercially available m⁷GpppA, AdoMet, AdoHcy (Sigma–Aldrich, Germany) and m^{2,2,7}GpppA (KEDAR, Poland) served as references for column calibration.

2.3. Crystallization and structure determination

The human TGS1 fragment was crystallized by the vapour-diffusion method in sitting-drop 24-well Cryschem plates (Hampton Research, USA). SeMet-containing TGS1 comprising amino acids 653–853 was crystallized in the presence of a sevenfold molar excess of the cap dinucleotide m⁷GpppA. 1 µl reservoir solution containing 1.5 M sodium formate and 0.1 M MES pH 6.1 was mixed with 1 µl of the prepared protein-substrate solution (6 mg ml^{−1}). Single crystals with dimensions of 70 × 70 × 300 µm grew within 3 d at 293 K and belonged to space group *P*₄₃₂₁₂, with unit-cell parameters *a*=*b*=213.9, *c*=62.4 Å. The crystals were soaked in cryosolution containing an additional 20% (v/v) glycerol for 5 s and subsequently mounted on a goniometer head in a 100 K cryostream. Peak, inflection-point and high-energy remote data sets were collected from an SeMet crystal on beamline BW7A at EMBL/DESY in Hamburg; the appropriate wavelengths were determined using a fluorescence scan. The remote data set was not used for phasing and refinement as it showed an increased *R*_{merge} compared with the peak and inflection-point data owing to radiation damage. The crystal was rotated in steps of 0.3° for the peak data set over a total range of 120° and in steps of 0.2° over the same range for the inflection-point data, while the distance was changed in between. Since SeMet-containing crystals diffracted to higher resolution than native crystals, phasing as well as refinement was performed using the SeMet-derivative crystals only. Data were integrated, scaled and reduced with the HKL-2000 suite (HKL Research, USA) and phases were obtained using HKL2MAP (Pape & Schneider, 2004).

Table 1. Statistics of data sets of selenomethionine (SeMet) TGS1₆₅₃₋₈₅₃

Values in parentheses are for the highest resolution shell.

Data set	Peak	Inflection
Data collection		
Space group	<i>P</i> 4 ₃ 2 ₁ 2	
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 213.9, <i>c</i> = 62.4, $\alpha = \beta = \gamma = 90$	
Wavelength (Å)	0.9799	0.9801
Resolution range (Å)	50.00-2.90 (3.00-2.90)	50.00-2.20 (2.28-2.20)
No. of reflections	32845	71603
Completeness (%)	100.0 (99.9)	97.0 (76.2)
R_{merge}^a (%)	5.6 (16.2)	4.8 (31.1)
Average $I/\sigma(I)$	31.8 (11.5)	34.7 (3.3)
Redundancy	9.4 (7.4)	8.7 (4.4)
Mosaicity (°)	0.35	0.34
No. of Se sites per ASU	12	
Refinement		
Resolution (Å)	50.0-2.2	
Molecules per ASU	4	
No. of atoms		
Protein	6030	
Ligand	108	
Waters	631	
R_{work}^b (%)	21.0	
R_{free}^c (%)	25.2	
Figure of merit	0.82	
Average B factors (Å ²)		
Protein	38.0	
Ligand	37.7	
Waters	46.4	
R.m.s. deviations		
Bond lengths (Å)	0.010	
Bond angles (°)	1.291	
Ramachandran statistics (%)		
Most favoured	92.0	
Allowed	8.0	
Generous	0.0	
Disallowed	0.0	

^(a) $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i \langle I_i(\text{hkl}) \rangle$, where the sum *i* is over all separate measurements of the unique reflection hkl.

^(b) $R_{\text{work}} = \sum_{\text{hkl}} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{\text{hkl}} |F_{\text{obs}}|$.

^(c) R_{free} as R_{work} , but summed over a 5% test set of reflections.

The resulting experimental electron-density map was used in ARP/wARP (Morris *et al.*, 2003) to build an initial model. The model was refined against the high-resolution inflection-point data set by iterative cycles of REFMAC5 (Murshudov *et al.*, 1997) and manual model building in Coot (Emsley & Cowtan, 2004). Waters were built using Coot and validated by hand. The structure was refined to good stereochemistry at a resolution of 2.2 Å to a final R_{work} of 21.0% and an R_{free} value of 25.2%. Owing to structural differences in the individual monomers, noncrystallographic symmetries (NCSs) were excluded from the refinement process. The Ramachandran plot of the refined structure model of human TGS1_{653–853} generated with SFCHECK (Vaguine *et al.*, 1999) shows that 92% of the refined residues are located within the most favoured regions and 8% in additionally allowed regions; no residues lie in the generously allowed or disallowed regions.

The four monomers in the asymmetric unit show D_2 symmetry with three twofold axes perpendicular to each other and consist of the following residues. Monomer 1 is defined by residues 649–848; residues 767–771 were not built owing to missing electron density. Molecules 2 and 3 are represented by residues 649–847, but residues 767–773 are missing from the model. Monomer 4 consists of amino acids 649–847; there was no electron density for residues 662–665 and 768–773 and thus they were not built. Figures were generated using PyMOL (DeLano, 2002). The statistics of the X-ray diffraction data sets and structure refinement are summarized in Table 1.

3. Results and discussion

The conserved methyltransferase (MTase) domain of human TGS1 corresponds to the C-terminal 200 residues of the protein, while the function of the N-terminal 652 residues in snRNP biogenesis is yet unclear. We generated a truncated human TGS1 containing only the minimal MTase domain according to the homology model predicted for yeast TGS1 (Mouaikel, Bujnicki *et al.*, 2003), which

comprises residues 653–853 (TGS1_{653–853}). This truncated hTGS1 was expressed in *E. coli*, purified and crystallized as described in §2. Crystals were only obtained when at least one of the reaction partners AdoMet, AdoHcy or m⁷GpppA was present in the crystallization buffer; all attempts to crystallize the apoenzyme failed.

Cocrystallization with m⁷GpppA yielded crystals that belonged to space group $P4_32_12$, whereas cocrystallization with AdoMet led to trigonal crystals which turned out to be almost perfectly twinned. Since the crystallographic phase problem could not be solved by means of molecular replacement, a selenomethionine (SeMet) derivative of TGS1_{653–853} was produced and crystallized in the presence of the cap analogue m⁷GpppA. A two-wavelength MAD experiment provided an interpretable electron-density map and the resulting crystal structure was refined at 2.2 Å resolution (Table 1). The overall structure of TGS1_{653–853} closely resembles the canonical fold of class I methyltransferases, which is characterized by a central seven-stranded β -sheet flanked by several α -helices on both sites. However, TGS1_{653–853} contains three additional N-terminal β -strands, of which the first extends the central β -sheet to eight strands with topology $\beta 1 \uparrow \beta 9 \uparrow \beta 10 \downarrow \beta 8 \uparrow \beta 7 \uparrow \beta 4 \uparrow \beta 5 \uparrow \beta 6 \uparrow$ (Fig. 1). This N-terminal β -strand 1 is located next to $\beta 9$ and mediates important crystal-packing contacts. Strand $\beta 1$ of each monomer in the asymmetric unit packs against strand $\beta 1$ of a monomer of the adjacent asymmetric unit in an antiparallel fashion, thus forming an extended β -sheet.

The two additional short β -strands $\beta 2$ and $\beta 3$ connecting $\beta 1$ and helix $\alpha 1$ are only present in two of the four monomers in the asymmetric unit of the crystal (Fig. 2). These structural deviations are caused by different crystal-packing contacts, suggesting conformational flexibility of this region.

Interestingly, the substrate m⁷GpppA is bound *via* its adenosine diphosphate moiety in the structurally conserved AdoMet-binding cleft, while m⁷G is fully disordered (Fig. 3).

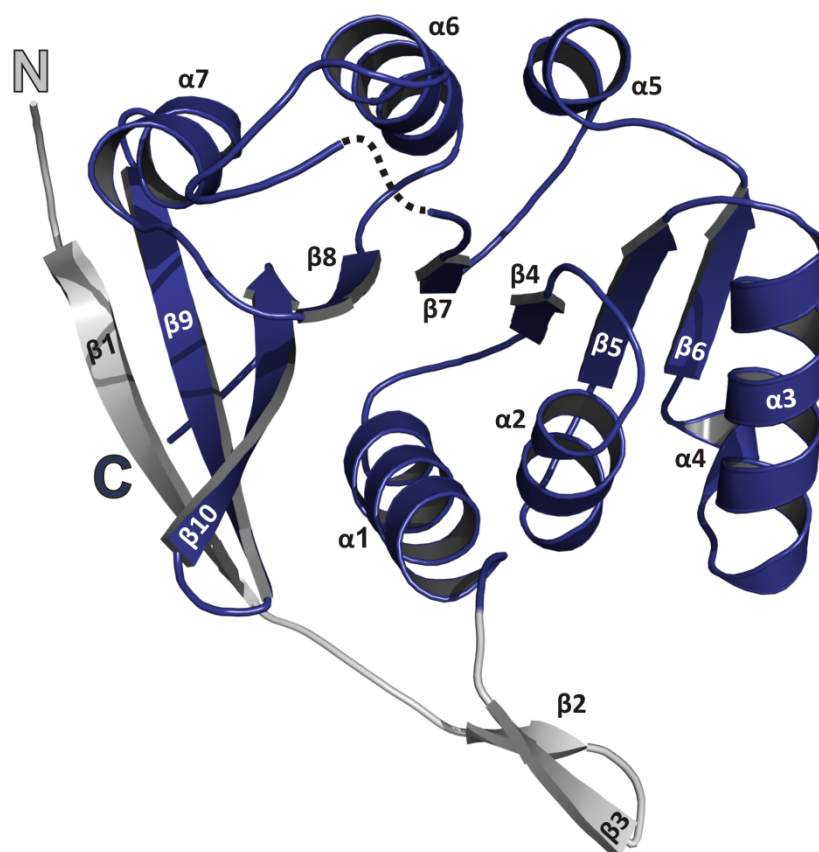


Figure 1. Overall structure of human TGS1 (amino acids 653–853) in cartoon representation. The canonical class I methyltransferase domain fold ($\alpha\beta\alpha$ -sandwich) is coloured blue, while the additional N-terminal extension (β -strands 1–3) is depicted in grey. The secondary-structure elements as well as the N- and C-termini are labelled. The missing connection between residues Trp766 and Ala774, which is not defined in the electron-density map, is shown as a dotted line.

The adenine base is sandwiched between the hydrophobic side chains of Phe698 and Ile720 and its N6 atom forms a hydrogen bond to the carboxylate of Asp747. Both ribose hydroxyls are hydrogen bonded by the side chain of Asp719 and the β -phosphate is bound by the side chain of Lys724. This binding mode of the adenosine moiety of m^7 GpppA closely resembles that of bound AdoMet or *S*-adenosyl-L-homocysteine (AdoHcy) in other class I methyltransferases (not shown).

The results of these crystallographic studies suggest that the crystallized TGS1_{653–853} is not capable of binding the m^7 G cap in the correct way. The observed binding of m^7 GpppA to the AdoMet pocket occurs owing to a defective m^7 G cap-binding pocket and only in absence of AdoMet in the crystallization buffer. In order to confirm this interpretation of the structural data,

the enzymatic activity of TGS1_{653–853} was measured using a newly established HPLC-based assay. The purified protein was incubated with both substrates and the reaction was stopped by precipitation of the protein. After removal of the precipitated protein by centrifugation, the reaction substrates and products were separated by reversed-phase HPLC and quantified. The reversed-phase HPLC column was calibrated using commercially available standards of all substrates and products (Fig. 4a).

As expected, on testing the crystallized TGS1_{653–853} neither formation of $m^{2,2,7}$ GpppA nor of AdoHcy could be observed (Fig. 4b), confirming the data derived from the crystal structure. In contrast, a larger TGS1 fragment comprising residues 636–853 (TGS1_{636–853}) and thus containing 17 additional N-terminal

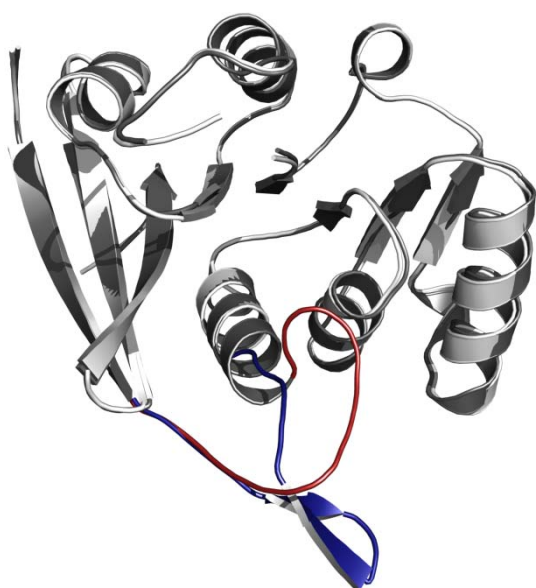


Figure 2. Superposition of the two different conformations present in the asymmetric unit. The four molecules in the asymmetric unit are evenly split into two conformational states. For each state only one molecule is depicted in the overlay, since the respective second molecule superposes almost perfectly. Within both conformations the major structural differences are restricted to the region connecting β -strand 1 to α -helix 1, as it harbours two β -strands in one state and no secondary-structure elements in the other. Therefore, this region is highlighted in blue and red for molecules 1 and 2, respectively, whereas the major parts of both molecules, coloured light and dark grey, respectively, are nearly identical.

residues was capable of converting m^7 GpppA to $m^{2,2,7}$ GpppA accompanied by the conversion of AdoMet to AdoHcy (Fig. 4c).

A structural model of yeast TGS1 (yTGS1) has previously been generated using a methyltransferase from *Methanococcus jannaschii* (MJ0882; PDB code 1dus; Huang *et al.*, 2002) as a template structure. This homology model contains yTGS1 residues Met70–Glu267, which correspond to amino acids Leu664–Ala850 in hTGS1 (Mouaikel, Bujnicki *et al.*, 2003). From this model it was predicted that residues Phe60–Cys262 form the minimal globular MTase domain, which in turn correspond to residues Phe655–Leu845 in hTGS1. The m^7 G cap was thought to be sandwiched between residues Trp178 (Trp766 in hTGS1) and Trp75 (Trp669 in hTGS1) (Bujnicki & Rychlewski, 2002; Mouaikel, Bujnicki *et al.*, 2003); residues Asp103 and Asp126 (Asp696

and Asp719 in hTGS1, respectively) participate in formation of the AdoMet-binding pocket. Surprisingly, all these residues are present in the crystallized but inactive TGS1_{653–853} fragment, raising questions regarding the molecular basis for the lack of its activity. In particular, the question arises whether the additional 17 residues are sufficient to fulfil the following two functions: to span the distance to the active site and to participate in the formation of a functional substrate-binding pocket.

Bridging the distance of approximately 30 Å from the N-terminus to the catalytic site would require approximately eight amino acids in an extended conformation, assuming a path leading directly through the protein. It is more plausible that the main chain would have to circumvent the protein with its globular shape, thus requiring even more residues. The consideration mentioned above together with the fact that an eight-stranded β -sheet has not been observed to date in any structure of a methyltransferase domain implies that a structural rearrangement of the N-terminal β -strands is more likely to occur. This hypothesis is further supported by secondary-structure predictions, which reveal an α -helical conformation of the region containing β -strands 1–3 (data not shown). An alternative explanation for the difference in catalytic

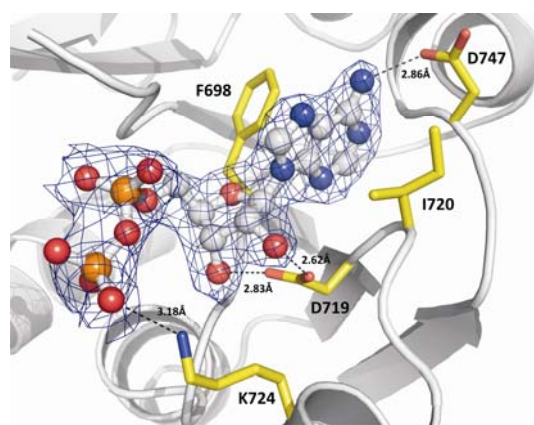


Figure 3. Detailed view of the bound adenosine moiety (ppA) from the cocrystallized m^7 GpppA cap dinucleotide. TGS1_{653–853} is shown in cartoon representation (light grey) and the residues involved in binding are labelled and highlighted as yellow sticks. The bound adenosine diphosphate (ADP) is surrounded by an $|F_o| - |F_c|$ OMIT map contoured at a level of 3σ (ADP was omitted).

activity between hTGS1_{636–853} and hTGS1_{653–853} is that the additional 17 residues may contribute to or facilitate protein dimerization or oligomerization and therefore lead to activation. However, this possibility can be excluded since both tested hTGS1 constructs exist as monomers

in solution as judged by gel-filtration chromatography analyses.

In summary, the structure of the human TGS1 methyltransferase domain presented here corresponds to an inactive truncated domain with a maximum of 17 amino-acid residues missing that would be required to gain catalytic activity. Whether this is achieved by a large structural rearrangement of the N-terminus or simply by reaching the active site within these 17 residues remains to be clarified.

4. Conclusions

The crystallized conserved methyltransferase domain of hTGS1 is catalytically inactive, even though it contains all the residues that have been predicted to be involved in substrate binding and catalysis by means of structural homology modelling. The presence of an additional 17 residues extending the N-terminus leads to an active methyltransferase. Crystal structure analysis reveals that the AdoMet-binding pocket of hTGS1_{653–853} is functional, while no complex with bound m⁷G could be obtained. It remains unclear how the additional N-terminal residues complete the active site; it is most likely that they contribute to the binding of the m⁷G cap. Hence, further structural and mutational studies are required in order to understand the substrate specificity and catalytic mechanism of hTGS1.

We are very grateful to Rémy Bordonné and Utz Fischer for providing the full-length clone of hTGS1. We thank the staff of EMBL beamline BW7A at DESY (Hamburg) for excellent support during data collection, as well as Stephanie Schell for assistance in protein preparation and crystallization. Furthermore, we are very thankful to Markus Rudolph for help with the activity assay. This work was supported by the DFG (SFB523).

References

Bujnicki, J. M. & Rychlewski, L. (2002). *BMC Bioinformatics*, **3**, 10.

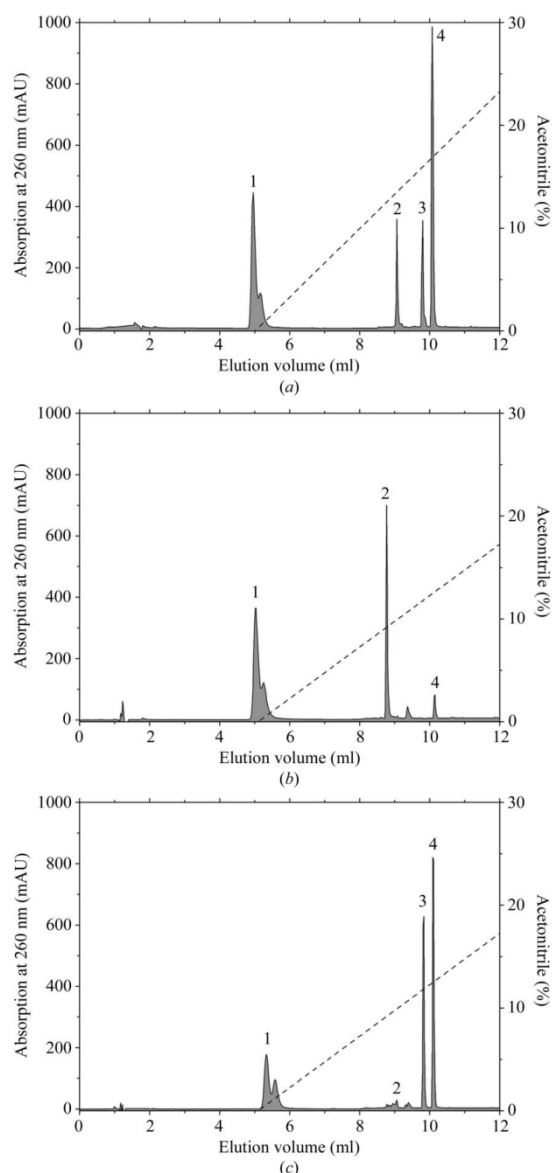


Figure 4. HPLC-based activity test of hTGS1 fragments. (a) Calibration run using a mixture of the substrates *S*-adenosyl-L-methionine (AdoMet; 1) and m⁷GpppA (cap analogue; 2) as well as both reaction products m^{2,2,7}GpppA (3) and *S*-adenosyl-L-homocysteine (AdoHcy; 4) as references. (b) Activity test for the crystallized TGS1 comprising residues 653–853. 25 μ M purified protein was mixed and incubated with 0.5 mM cap analogue m⁷GpppA and 2 mM AdoMet for 120 min. (c) The 17-amino-acid longer fragment (amino acids 636–853) shows catalytic activity as judged by the utilized substrates after 120 min incubation.

- Colau, G., Thiry, M., Leduc, V., Bordonne, R. & Lafontaine, D. L. (2004). *Mol. Cell. Biol.* **24**, 7976–7986.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, California, USA.
- Dickmanns, A. & Ficner, R. (2005). *Top. Curr. Genet.* **12**, 179–204.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Enunlu, I., Papai, G., Cserpan, I., Udvardy, A., Jeang, K. T. & Boros, I. (2003). *Biochem. Biophys. Res. Commun.* **309**, 44–51.
- Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., Bertrand, E. & Bordonne, R. (2008). *J. Biol. Chem.* **283**, 2060–2069.
- Gunzl, A., Bindereif, A., Ullu, E. & Tschudi, C. (2000). *Nucleic Acids Res.* **28**, 3702–3709.
- Hamm, J., Darzynkiewicz, E., Tahara, S. M. & Mattaj, I. W. (1990). *Cell*, **62**, 569–577.
- Hausmann, S., Ramirez, A., Schneider, S., Schwer, B. & Shuman, S. (2007). *Nucleic Acids Res.* **35**, 1411–1420.
- Hausmann, S. & Shuman, S. (2005a). *J. Biol. Chem.* **280**, 4021–4024.
- Hausmann, S. & Shuman, S. (2005b). *J. Biol. Chem.* **280**, 32101–32106.
- Hausmann, S., Zheng, S., Costanzo, M., Brost, R. L., Garcin, D., Boone, C., Shuman, S. & Schwer, B. (2008). *J. Biol. Chem.* **283**, 31706–31718.
- Huang, L., Hung, L., Odell, M., Yokota, H., Kim, R. & Kim, S.-H. (2002). *J. Struct. Funct. Genomics*, **2**, 121–127.
- Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. & Luhrmann, R. (1998). *EMBO J.* **17**, 4114–4126.
- Komonyi, O., Papai, G., Enunlu, I., Muratoglu, S., Pankotai, T., Kopitova, D., Maroy, P., Udvardy, A. & Boros, I. (2005). *J. Biol. Chem.* **280**, 12397–12404.
- Limbach, P. A., Crain, P. F. & McCloskey, J. A. (1994). *Nucleic Acids Res.* **22**, 2183–2196.
- Mattaj, I. W. (1986). *Cell*, **46**, 905–911.
- Maxwell, E. S. & Fournier, M. J. (1995). *Annu. Rev. Biochem.* **64**, 897–934.
- Misra, P., Qi, C., Yu, S., Shah, S. H., Cao, W. Q., Rao, M. S., Thimmapaya, B., Zhu, Y. & Reddy, J. K. (2002). *J. Biol. Chem.* **277**, 20011–20019.
- Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
- Mouaikel, J., Bujnicki, J. M., Tazi, J. & Bordonne, R. (2003). *Nucleic Acids Res.* **31**, 4899–4909.
- Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A. G., Bertrand, E., Tazi, J. & Bordonne, R. (2003). *EMBO Rep.* **4**, 616–622.
- Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonne, R. (2002). *Mol. Cell*, **9**, 891–901.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Neuenkirchen, N., Chari, A. & Fischer, U. (2008). *FEBS Lett.* **582**, 1997–2003.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Plessel, G., Fischer, U. & Luhrmann, R. (1994). *Mol. Cell. Biol.* **14**, 4160–4172.
- Reuter, K. & Ficner, R. (1999). *Acta Cryst.* **D55**, 888–890.
- Ruan, J. P., Ullu, E. & Tschudi, C. (2007). *Mol. Biochem. Parasitol.* **155**, 66–69.
- Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). *Trends Biochem. Sci.* **28**, 329–335.
- Strasser, A., Dickmanns, A., Luhrmann, R. & Ficner, R. (2005). *EMBO J.* **24**, 2235–2243.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Verheggen, C., Lafontaine, D. L., Samarsky, D., Mouaikel, J., Blanchard, J. M., Bordonne, R. & Bertrand, E. (2002). *EMBO J.* **21**, 2736–2745.
- Watkins, N. J., Lemm, I., Ingelfinger, D., Schneider, C., Hossbach, M., Urlaub, H. & Luhrmann, R. (2004). *Mol. Cell*, **16**, 789–798.
- Zhu, Y., Qi, C., Cao, W. Q., Yeldandi, A. V., Rao, M. S. & Reddy, J. K. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10380–10385.

Chapter 5 • Structural basis for m^7G -cap dimethylation by human TGS1

This manuscript has originally been published in *Nucleic Acids Research*.

'Structural basis for m^7G -cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1'

Thomas Monecke, Achim Dickmanns and Ralf Ficner[‡]

Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik, Abteilung für Molekulare Strukturbiologie (Göttinger Zentrum für Molekulare Biowissenschaften (GZMB), Ernst-Caspari-Haus), Justus-von-Liebig-Weg 11, D-37077, Göttingen

[‡] To whom correspondence should be addressed: E-mail: rficner@uni-goettingen.de
Phone: 0551 – 39 14071
Fax: 0551 – 39 14082

Keywords: m^7G ; dimethyltransferase; UsnRNA; snoRNA; telomerase

Nucleic Acids Research; Epub ahead of print, Advance Access published on April 22, 2009
© 2009 by the Authors.

Received February 24, 2009; Accepted April 2, 2009; published online April 22, 2009

Preface – About the Manuscript

The crystal structure of a truncated but inactive form of the human Trimethylguanosine Synthase 1 (TGS1) was determined showing an incomplete binding pocket for the m⁷G-cap ([Chapter 4](#)). The following publication describes the structural, functional and catalytic characterization of an active form of the human snRNA/snoRNA/telomerase RNA hypermethylase. Interestingly, an N-terminal extension, which is missing in the inactive structure, forms a small separate domain, which is essential for the correct binding of both substrates and for catalysis. In context of the inactive structure, the additional N-terminal residues of the active fragment most likely contribute to the stabilization of an active and productive form of the N-terminal domain.

The coordinates and structure factors of the crystal structure described in the following publication have been deposited in the protein data bank (PDB) under the PDB ID 3GDH.

Author contributions:

The project was initiated by Prof. Dr. Ralf Ficner, Dr. Achim Dickmanns and me. My contribution under the supervision of Prof. Dr. Ralf Ficner included molecular cloning, protein purification and crystallization. I performed the structure determination and analysis as well as the biochemical characterization by means of site-directed mutagenesis studies and HPLC-based activity assays. Bernhard Kuhle contributed to this work in terms of protein purification, crystallization and activity assays within a laboratory practical course.

ABSTRACT

The 5'-cap of spliceosomal small nuclear RNAs, some small nucleolar RNAs and of telomerase RNA was found to be hypermethylated *in vivo*. The Trimethylguanosine Synthase 1 (TGS1) mediates this conversion of the 7-methylguanosine-cap to the 2,2,7-trimethylguanosine (m₃G)-cap during maturation of the RNPs. For mammalian UsnRNAs the generated m^{2,2,7}G-cap is one part of a bipartite import signal mediating the transport of the UsnRNP-core complex into the nucleus. In order to understand the structural organization of human TGS1 as well as substrate binding and recognition we solved the crystal structure of the active TGS1 methyltransferase domain containing both, the minimal substrate m⁷GTP and the reaction product *S*-adenosyl-L-homocysteine (AdoHcy). The methyltransferase of human TGS1 harbors the canonical class 1 methyltransferase fold as well as an unique N-terminal, α -helical domain of 40 amino acids, which is essential for m⁷G-cap binding and catalysis. The crystal structure of the substrate bound methyltransferase domain as well as mutagenesis studies provide insight into the catalytic mechanism of TGS1.

INTRODUCTION

The m⁷G-cap of small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) as well as the telomerase RNA TLC1 is known to be hypermethylated by the Trimethylguanosine Synthase 1 (TGS1) during maturation of the respective ribonucleoprotein particles (RNPs). Especially the biogenesis of the human uridyl-rich snRNPs U1, U2, U4, U5 and U6, the major components of the spliceosome, is well characterized (1,2). Except for U6snRNP, in vertebrates the biogenesis of UsnRNPs involves a nucleocytoplasmic transport cycle. After transcription by RNA-polymerase II, the snRNAs U1, U2, U4 and U5 acquire an m⁷G-cap and are subsequently exported to the cytoplasm in a cap dependent manner (3,4). In the cytoplasm seven Sm-proteins (B/B', D1, D2, D3, E, F and G) are assembled in a doughnut-like structure around the highly conserved Sm-site (5'-PuA(U)_nGPu-3') of the UsnRNA (5-7). The resulting Sm-core UsnRNP was shown to be a prerequisite for m⁷G-cap hypermethylation (7,8) by a cap dimethyltransferase, catalyzing the addition of two methyl groups to the exocyclic NH₂ group of the N7-methylguanine (m⁷G) (7,9,10). The m₃G-cap and the Sm-core serve as bipartite import signal, which mediates the import of the assembled UsnRNP-core into the nucleus (11-17). The import occurs by the concerted action of snurportin1 (SPN1), the

survival of motor neuron (SMN)-complex and importin β (Imp β). SPN1 binds the hypermethylated UsnRNA cap-guanine and the first nucleotide of the RNA in a stacked conformation and bridges the interaction to the import factor Imp β (13,18). After translocation Imp β dissociates from the complex due to the binding of Ran-GTP. A subcomplex containing at least the assembled UsnRNP and SPN1, which is bound to the m₃G-cap, remains (19) and its disassembly is not yet understood. The released UsnRNPs individually associate with specific sets of proteins, resulting in the particular, mature UsnRNPs U1, U2, U4 and U5 (20). However, the U6snRNP is generated by a different pathway that does not include such cytoplasmic maturation steps. In contrast to all other UsnRNPs it is transcribed by RNA polymerase III (21-23) and acquires a γ -monomethyl-phosphate-cap. Together with the other UsnRNPs it accumulates subsequent to maturation in nuclear (interchromatin) speckles and in cajal bodies (24,25).

The enzyme responsible for hypermethylation of the UsnRNA-m⁷G-cap in yeast was first identified by Mouaikel and co-workers in 2002 and it was named TGS1 in accordance to its function (26). As a *bona fide* dimethyltransferase it catalyzes two successive methyl group transfers, each from one *S*-adenosyl-L-methionine (AdoMet) to the exocyclic nitrogen

N2 of the cap-guanine, generating two *S*-adenosyl-L-homocysteine molecules (AdoHcy) and the modified m^{2,2,7}G (m₃G)-cap (27). The two Sm-proteins SmB and SmD1 of the UsnRNP-core bind to TGS1 in yeast and in human cells suggesting an additional interaction between the dimethyltransferase and the RNP besides the cap (10,26). Although the yeast TGS1 (yTGS1) is not essential for cell viability its deletion results in a cold-sensitive phenotype, as the mutants show reduced growth at lower temperatures as well as missing m₃G-caps on certain UsnRNAs and snoRNAs (26). As a consequence of this deletion a splicing defect at restrictive temperatures is observed and linked to the appearance of nucleolar retention of U1snRNAs. Furthermore, the nucleolar morphology is disturbed and the cells show impairment in ribosome biogenesis. However, it turned out that it is the TGS1 protein *per se*, rather than its catalytic activity that is required for efficient ribosome synthesis (28).

In contrast to yTGS1, the orthologs from *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus* and *Homo sapiens* additionally contain large N-terminal domains.

The human ortholog of yTGS1 was originally designated PIMT (PRIP-interacting protein with methyltransferase domain) because of its binding capability to PRIP (PPAR-interacting protein) and the presence of the C-terminal methyltransferase domain (29). As shown recently, the purified C-terminal methyltransferase domain of hTGS1 containing the amino acids 576–853 is able to complement the cold-sensitive phenotype of a yeast Δ tgs1-strain demonstrating that the human TGS1 methyltransferase domain is a true functional ortholog of the yeast enzyme (30). Hence, *H. sapiens* PIMT is referred to as hTGS1 with respect to its function and relatedness to yTGS1. It was found that hTGS1 is not only able to bind *via* its large N-terminus to PRIP, but also to enhance its action. PRIP was shown to interact with the peroxisome proliferator-activated receptor (PPAR). This suggests that the human

TGS1 is a component of the nuclear receptor signal transduction cascade acting through PRIP, and at least in human cells is involved in transcriptional regulation (29). Besides that, hTGS1 was shown to interact with the transcriptional coactivators CREB-binding protein (CBP), p300 and PPAR-binding protein (PBP) *in vitro* and *in vivo* (31). hTGS1 exists in two isoforms, a full-length cytoplasmic and an N-terminally shortened nuclear form, which might be involved in snoRNP maturation (32,33). In addition it was found *in vivo* and *in vitro*, that the C-terminal part of the human cap dimethyltransferase interacts directly with the survival of motor neuron protein (SMNp) (34). This interaction is greatly impaired *in vivo* as well as *in vitro* by a SMN mutation (SMN Δ Ex7) that is most commonly present in patients who are adversely affected by the neuromuscular disease spinal muscular atrophy (SMA). Both proteins are localized in the cytoplasm as well as in the nuclear cajal bodies, where the assembled UsnRNPs and snoRNPs reside (34).

Contrary to the comparative mild effects upon deletion of yTGS1, the deletion of its counterpart in *D. melanogaster* is lethal to the larvae in the early pupal stages indicating an essential function in development (35). The fruit fly ortholog was named drosophila-tat-like (DTL) because of its capability to bind the TAR-RNA of the human immunodeficiency virus (HIV). Furthermore, by means of mutagenesis studies it could be shown that this lethal effect is causally linked to the observed absence of m₃G-caps on UsnRNPs, which is a direct result of the TGS1-knockdown (35).

The m₃G-cap does not solely occur on spliceosomal UsnRNPs, but also on a subset of small nucleolar RNAs (snoRNAs). The modification of these snoRNAs such as box C/D (U3) and box H/ACA (snR10/30), which are, at least in yeast, also hypermethylated by TGS1 is less well understood (36,37). The cap dimethyltransferase interacts with specific proteins of particular snoRNAs such as Cbf5 and Nop58, thus giving a possibility to distinguish between UsnRNA and different snoRNAs with

respect to their state of methylation (26). As for yeast UsnRNAs, it was shown for the snoRNAs that they do not cycle through the cytoplasmic compartment during maturation (38). Interestingly, it is known that yTGS1 is localized exclusively in the nucleus and that this organism lacks the orthologs for important proteins required for m⁷G-cap dependent UsnRNA export (phosphorylated adapter for RNA export, PHAX) and m₃G-dependent UsnRNP import (SPN1). Therefore it is likely that, in contrast to vertebrates, the maturation of UsnRNPs and snoRNPs in *Saccharomyces cerevisiae* occurs exclusively in the nucleus. This assumption is further confirmed by the fact, that the absence of m₃G-caps in yeast is not lethal and instead produces only a cold-sensitive phenotype and a mild splicing defect (26).

Quite recently, it could be demonstrated that in yeast the m₃G-cap at the 5'-end of the telomerase RNA TLC1 results from the catalytic activity of TGS1 as well (39). TLC1 shares several features with UsnRNAs and snoRNAs, among them the transcription by RNA polymerase II, a high content of uridines, the ability to bind Sm-proteins and the presence of a 5'-m₃G-cap. Deletion of TGS1 in yeast cells influences the length, structure and function of telomeres and they show premature aging (39).

Although there are numerous biochemical data characterizing TGS1 from *H. sapiens*, *S. cerevisiae*, *D. melanogaster*, *Schizosaccharomyces pombe*, *Giardia lamblia* and *Trypanosoma brucei* (27,28,40–43) the three-dimensional structure of an active form of this enzyme is unknown so far. However, the crystal structure of a truncated, but catalytically inactive fragment of the human TGS1 was recently determined. This fragment contains the canonical methyltransferase domain, which on its own is not sufficient for m⁷G-cap hypermethylation (44).

Here we present the 2 Å crystal structure of the catalytically active methyltransferase domain of hTGS1. The results provide insight into the structural organization and flexibility as well as a detailed view on substrate binding and

recognition by TGS1. Furthermore, on the basis of the crystal structure and biochemical data including mutagenesis studies, a mechanism for the dimethylation reaction catalyzed by TGS1 is proposed.

MATERIALS AND METHODS

Protein expression and purification

The human TGS1 fragment comprising the residues 618–853 (hTGS1_{618–853}) was subcloned from pGEX-6P-1 full-length TGS1 (residues 1–853; accession number Q96RS0) into BamHI/XhoI-sites of pGEX-6P-1 (GE Healthcare, Germany) and verified by sequencing. The GST-fusion constructs were expressed in *Escherichia coli* BL21(DE3) (Invitrogen, USA) at 16°C in ampicillin containing 2YT-medium, which was supplemented with 2% (w/v) α-D-glucose in order to suppress basal transcription. Expression of GST-hTGS1_{618–853} was induced at OD₆₀₀ = 0.8, adding IPTG to a final concentration of 0.5 mM. Directly after induction, 2% (v/v) ethanol and 50 mM K₂HPO₄ were added to the growing culture. The cells were harvested after 18 h of expression (5000 × g, 20 min, 4°C) and resuspended in lysis buffer containing 50 mM Tris/HCl pH 7.5, 500 mM NaCl, 2 mM EDTA and 2 mM DTT. All subsequent steps were carried out at 4°C unless stated otherwise. Cells were disrupted using a microfluidizer 110S (Microfluidics, USA). The clarified lysate (30 000 × g, 30 min, 4°C) was subsequently loaded onto a GSH-Sepharose column (GE Healthcare, Germany) equilibrated with lysis buffer. Unbound proteins were removed by washing with 2 column volumes (CV) of lysis buffer. In order to eliminate RNA contaminations the loaded column was washed with 1 CV of a high salt buffer containing 50 mM Tris/HCl pH 7.5, 1 M NaCl and 2 mM DTT. After re-equilibration in lysis buffer the bound fusion protein was eluted with lysis buffer containing additionally 25 mM reduced glutathione. For cleavage the fusion protein (GST-hTGS1_{618–853}) was incubated with PreScission protease (GE Healthcare, Germany) at 4°C overnight in a 1:100 molar ratio of

protease:fusion protein. Further purification was achieved using a Superdex S75 (26/60) gel filtration column (GE Healthcare, Germany) in a buffer containing 20 mM Tris/HCl pH 7.5, 200 mM NaCl and 2 mM DTT. In order to remove small amounts of remaining GST, the pooled gel filtration fractions were finally purified using a second GSH-sepharose equilibrated in gel filtration buffer. The pure protein was concentrated to 8 mg/ml using vivaspin concentrators with a molecular weight cut-off of 10 000 Da (Sartorius, Germany) and aliquots were frozen in liquid nitrogen and stored at -80°C.

Site-directed mutagenesis

Single amino-acid mutants of hTGS1_{618–853} (Ser763Ala, Ser763Asp, Trp766Ala, Asp696Ala, Asp696Asn and Phe804Lys) were generated from the wild-type clone using the QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene, USA) following the manufacturer's protocol. All mutations were confirmed by DNA sequencing and the mutants were expressed and purified in analogy to the procedures for the wild-type TGS1-fragment.

HPLC-based activity assay

In order to analyze the activity of the purified hTGS1 fragment as well as of the described mutants, an HPLC-based activity assay was applied (44). Total 0.25 nmol of purified hTGS1_{618–853} were mixed with 5 nmol of the minimal substrate m⁷GTP (Sigma-Aldrich, Germany) and 20 nmoles of the cofactor AdoMet (Sigma-Aldrich, Germany) in 1 × PBS. The mixture with a total volume of 10 µl was incubated at 37°C, the reaction was stopped by the addition of 1 µl 1 M HClO₄ followed by subsequent incubation on ice for 1 min. The solution was neutralized adding 20 µl 2 M Na-acetate. Precipitated protein was pelleted by centrifugation (16 000 × g, 10 min, 20°C) and the supernatant was loaded onto a reversed phase HPLC-column (Prontosil C18-AQ, Bischoff Chromatography, Germany), which was equilibrated in buffer A, containing 100 mM K₂HPO₄/KH₂PO₄ pH 6.5.

The substrates and products of the reaction were eluted from the column applying a linear gradient from 0% to 60% of buffer B, containing buffer A and 50% (v/v) acetonitrile. Commercially available m⁷GTP, AdoMet, AdoHcy (Sigma-Aldrich, Germany) and m^{2,2,7}(₃)GTP (KEDAR, Poland) served as reference for column calibration.

Crystallization and structure determination

hTGS1_{618–853} comprising the C-terminal 235 residues was crystallized by the vapor diffusion method in sitting drop 24-well ChrysChem-plates (Hampton Research, USA) and in presence of an 8-fold molar excess of both, the minimal substrate m⁷GTP and the reaction product *S*-adenosyl-L-homocysteine (AdoHcy). The appropriate amount of AdoHcy was desiccated, mixed with the protein solution and incubated for 10 min at 20°C. Afterwards the cap-analog m⁷GTP was added followed by an additional incubation for 10 min at 20°C. One microliter of a reservoir solution containing 13–16% PEG 8000 and 0.1 M MES pH 6.2–6.8 was mixed with 1 µl of the prepared protein-substrate solution (4–6 mg/ml). Rhombohedral shaped single crystals with dimensions of 70 × 70 × 60 µm³ grew at 20°C within 2 weeks. Crystals belong to the space group *R*3 with the cell dimensions *a* = *b* = 156.2 Å, *c* = 100.3 Å and angles $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$. The crystals were flash frozen in liquid nitrogen after soaking in reservoir solution containing an additional 15% (v/v) 1,2-propanediol as cryo protectant. An X-ray diffraction dataset of a crystal containing hTGS1_{618–853} as well as m⁷GTP and AdoHcy was collected at beamline BL14.2 of the electron synchrotron in Berlin (BESSY). Data were integrated, scaled and reduced with the HKL2000 suite (HKL Research, USA) and the structure was solved by means of molecular replacement. Therefore a starting model was designed, containing the canonical methyltransferase fold of the previously determined structure of SeMet-containing inactive hTGS1 (44). The structure of SeMet-

Table 1: Statistics of the data set of hTGS1 methyltransferase domain (aa618-853) bound to m⁷GTP and AdoHcy

Crystal	hTGS1 ₆₁₈₋₈₅₃ +m ⁷ GTP+AdoHcy
Data collection	
Space group	R3
Cell dimensions	
a, b, c (Å)	156.2, 156.2, 100.3
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.9184
X-ray source	BL14.2, BESSY (Berlin)
Resolution range (Å)	30.00-2.00 (2.07-2.00)
Number of reflections	61489
Completeness (%)	99.8 (100.0)
R _{merge} ^a (%)	5.4 (35.9)
Average I/σ	16.5 (2.6)
Redundancy	2.4 (2.4)
Mosaicity (°)	0.20
Refinement	
Resolution (Å)	30.0-2.00
Molecules per AU	3
Number of atoms	
Protein	4973
Ligand	202
Waters	415
R _{work} ^b (%)	18.0
R _{free} ^c (%)	21.3
Figure of merit	0.86
Average B factors (Å²)	
Protein	33.0
Ligand	32.9
Waters	38.9
RMS deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.914
Ramachandran statistics (%)	
Most	92.8
Allowed	7.2
Generous	0.0
Disallowed	0.0

Values in parentheses indicate the specific values in the particular highest resolution shell.

^(a)R_{merge} = $\sum_{hkl} \sum_i |I_i(hkl) - \langle I_i(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where the sum *i* is over all separate measurements of the unique reflection *hkl*.

^(b)R_{work} = $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$.

^(c)R_{free} as R_{work}, but summed over a 5 % test set of reflections.

hTGS1 comprising the residues Pro674-Arg848 served as search model in PHASER (45), which placed three molecules in the asymmetric unit. Missing residues, as well as 415 water molecules were added manually in COOT (46) and the structure was refined using REFMAC5 (47) to a reasonable R_{work} of 18.0% and a R_{free}-value of 21.3%. In the final model, 92.8% of the residues are located within the most favoured regions and 7.2% in the additionally allowed ones, whereas no residues lie in the generously allowed or disallowed regions of the Ramachandran plot (see Table 1). Figures were generated using PyMOL [DeLano, W.L. The PyMOL Molecular Graphics System (2002), DeLanoScientific, USA].

RESULTS

The globular core of yTGS1 (aa 58–266) was predicted to represent the catalytic domain, which corresponds to the C-terminal residues 653–845 in the human TGS1 (48) (sequence alignment in Supplementary Data). Recently, the crystal structure of this canonical methyltransferase domain of human TGS1 (aa 653–853) was solved, but this domain lacks catalytic activity suggesting an incomplete cap binding pocket (44). Interestingly, a fragment containing only 17 additional amino acids N-terminally and therefore comprising the residues 636–853 gains catalytic activity (44). Since all attempts to crystallize the active TGS1₆₃₆₋₈₅₃ failed, several further elongated fragments of hTGS1 were tested in crystallization. Finally, the crystallization of a fragment starting at residue 618 (hTGS1₆₁₈₋₈₅₃) in the presence of substrate-analogs succeeded.

Purification and characterization of hTGS1₆₁₈₋₈₅₃

hTGS1₆₁₈₋₈₅₃ was cloned and purified as described in ‘Materials and Methods’ section. An HPLC-based activity assay was applied to determine the catalytic activity of this hTGS1 fragment (Figure 1). After the incubation of purified hTGS1₆₁₈₋₈₅₃ in the presence of m⁷GTP and AdoMet, their significant decrease and

conversely increasing amounts of reaction products $m^{2,7}$ GTP and AdoHcy indicate the catalytic activity of the analyzed fragment (Figure 1B). The additional peak, for which no commercially available standard exists, elutes at 8.5 ml and is most likely the reaction intermediate $m^{2,7}$ GTP that has previously been reported (27,30,42).

Overall structure of the active hTGS1 methyltransferase domain

The active fragment (hTGS1_{618–853}) was crystallized in the presence of m^7 GTP and *S*-adenosyl-L-homocysteine (AdoHcy), the crystal structure was solved by means of molecular replacement and refined at a resolution of 2.0 Å. The statistics of X-ray diffraction data and structure refinement are summarized in Table 1. Crystals belong to the space group *R*3 with cell dimensions of $a = b = 156.2$ Å, $c = 100.3$ Å and $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$, containing three TGS1 monomers in the asymmetric unit. Residues 634–844 of monomer 1 are well defined in the electron density map, whereas for monomers 2 and 3 only the residues 637–847 and 641–847 could be placed.

All three monomers superpose well with root mean square deviations of 0.47 Å, 0.60 Å and 0.28 Å for all three possible combinations (molecule 1/2, 2/3 and 1/3) and with respect to all common C_α atoms. Due to the high similarity of the three monomers only the structure of monomer 1 is discussed hereafter. One monomer consists of 11 α -helices and seven β -strands (Figure 2). It is divided into the core domain (residues Glu675–Asp844) forming the classical class I methyltransferase fold (green part) and an α -helical N-terminal extension (NTE) encompassing the residues Leu634–Ser671 (yellow part). Both domains are separated by a short loop comprising the three amino acids Val672, Thr673 and Pro674. The core domain of the monomer consists of a central seven-stranded β -sheet, which is flanked by α -helices on both sides, leading to the typical $\alpha\beta\alpha$ sandwich and resembles the structure of the classical Rossmann-fold AdoMet-dependent

methyltransferase superfamily. This fold is generally referred to as α/β twist with the first strand located in the middle of the β -sheet whereupon the following strands are placed consecutively outward to one edge. Thereafter, the chain returns to the middle of the β -sheet and continues out to the other edge. Thus, the β -sheet has a $\beta 6 \uparrow \beta 7 \downarrow \beta 5 \uparrow \beta 4 \uparrow \beta 1 \uparrow \beta 2 \uparrow \beta 3 \uparrow$ topology with six strands in a parallel and the last one (β -strand 7) in an antiparallel orientation but positioned in between β -strand 5 and 6 (Figure 2). The β -sheet is twisted, thus the two outermost strands lie in an almost perpendicular orientation with respect to each other. Between two consecutive β -strands one or more α -helices build a right-handed crossover connection. Thus altogether three α -helices ($\alpha 5$, $\alpha 6$, $\alpha 7$) pack against one side of the β -sheet and another three helices ($\alpha 8$, $\alpha 10$, $\alpha 11$) on the opposing side. The α -helices adjacent to the β -sheet are oriented almost parallel to the strands, following the rotation impaired by this β -sheet. Using this conserved methyltransferase core as search model, the program DALI (49) found nearly 600 methyltransferases in the protein databank (PDB) with Z-scores greater than 12, indicating a closely related fold as discussed below.

The remaining four N-terminal α -helices of the structure ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$) form a separate overall globular subdomain (Figure 2, yellow part) with a small hydrophobic core. This small NTE packs against the core domain (Figure 2) and is involved in binding and recognition of both ligands. Interestingly, the residues Pro769 and Phe670 as well as Pro765 and Glu667 are in close proximity to each other and hence close the binding clefts on the top over both substrates as shown in the surface representation of Figure 2. A DALI-search using the NTE (aa 634–674) revealed that there are no structural related proteins or domains present in the protein data bank (PDB).

Substrate binding sites

The methyltransferase domain of human TGS1 harbors both, m^7 GTP and AdoHcy bound in a pocket and they are held in close proximity to

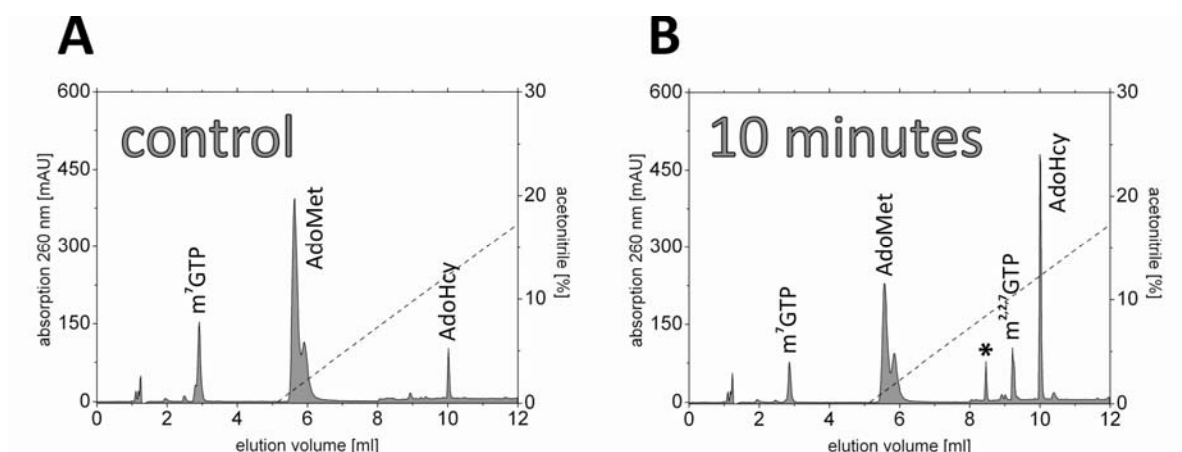


Figure 1. HPLC-based methyltransferase assay showing the catalytic activity of the crystallized hTGS1_{618–853}. **(A)** Chromatogram of the control reaction containing the substrates but no purified human hTGS1_{618–853} and **(B)** chromatogram after 10 min of incubation at 37°C including the purified enzyme. The elution volume is plotted against the absorption at 260nm (left y-axis) and the acetonitrile concentration (right y-axis), respectively. The corresponding peaks are labeled and the putative reaction intermediate $m^{2,7}$ GTP is marked by an asterisk.

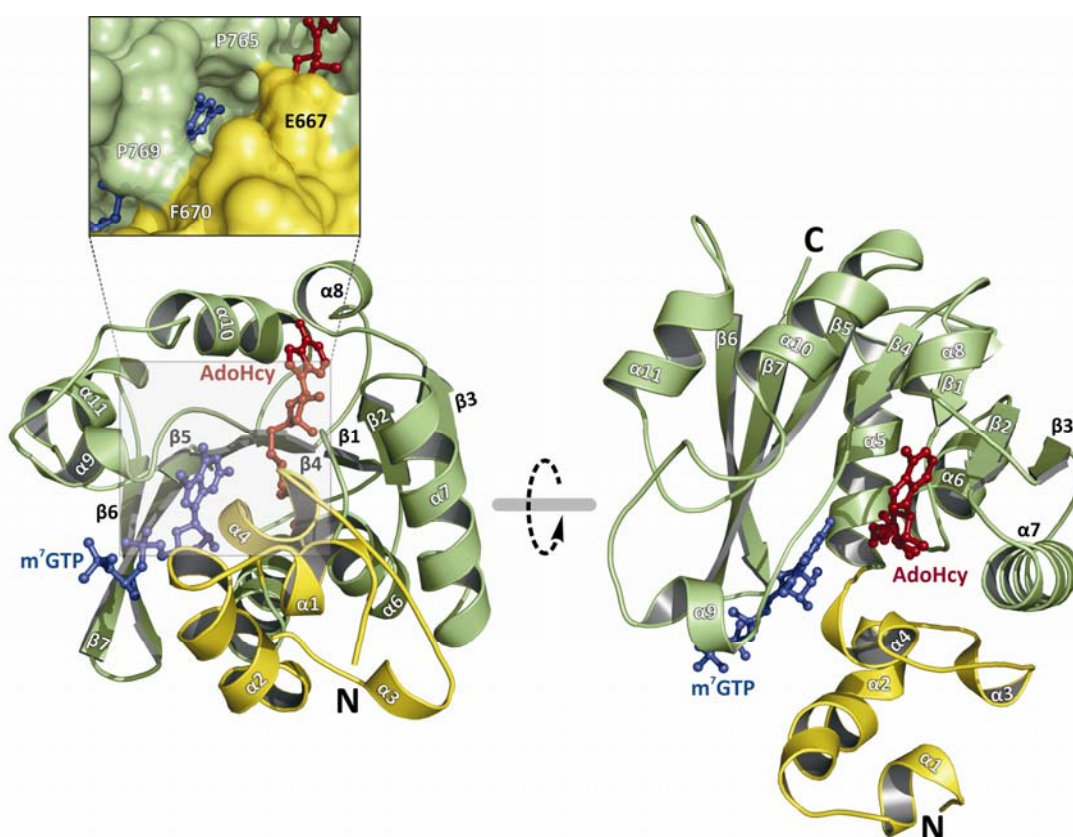


Figure 2. Overall structure of the human TGS1 methyltransferase domain. The crystallized methyltransferase domain comprising the residues Leu634–Asp844 is shown in cartoon representation and the bound substrate-analogs *S*-adenosyl-L-homocysteine (AdoHcy, red) and m^7 GTP (blue) are depicted in ball-and-stick mode. The methyltransferase core (Glu675–Asp844) is colored in light green, whereas the N-terminal, α -helical extension (NTE; Leu634–Ser671) is shown in yellow. Both views are rotated by 65° with respect to each other around the indicated rotation axis. The secondary structure elements are labeled. The upper panel shows a surface representation of the active site cleft highlighting the enclosure of both ligands by the enzyme.

each other by an intricate pattern of hydrogen bonds. Either substrate is well defined in the $|\text{Fo}| - |\text{Fc}|$ electron density omit map as shown in Figure 3A. The binding site for AdoHcy is mainly built up by residues belonging to the loops connecting β -strand 1 with α -helix 6, β -strand 2 with α -helix 7 and β -strand 3 with α -helix 8, respectively (Figures 2, 3A and C). The loop joining the methyltransferase core and the α -helix 4 of the NTE, as well as residues from the helix itself make additional contacts to AdoHcy. The adenine-moiety of AdoHcy is directly bound between the two side chains of

Phe698 and Ile720 generating a tight hydrophobic interaction (Figure 3C). Furthermore, N6 of this adenine forms a hydrogen bond with Asp747 and a water molecule, respectively. The ribose of AdoHcy adopts a C1'-exo conformation in the crystal structure and its hydroxyls are bound by two hydrogen bonds each. One hydrogen bond is formed from both hydroxyls to each of the two oxygen atoms of Asp719. The other hydrogen bonds are mediated by two water molecules, while the water bound to the 2'-OH is further coordinated by Glu667.

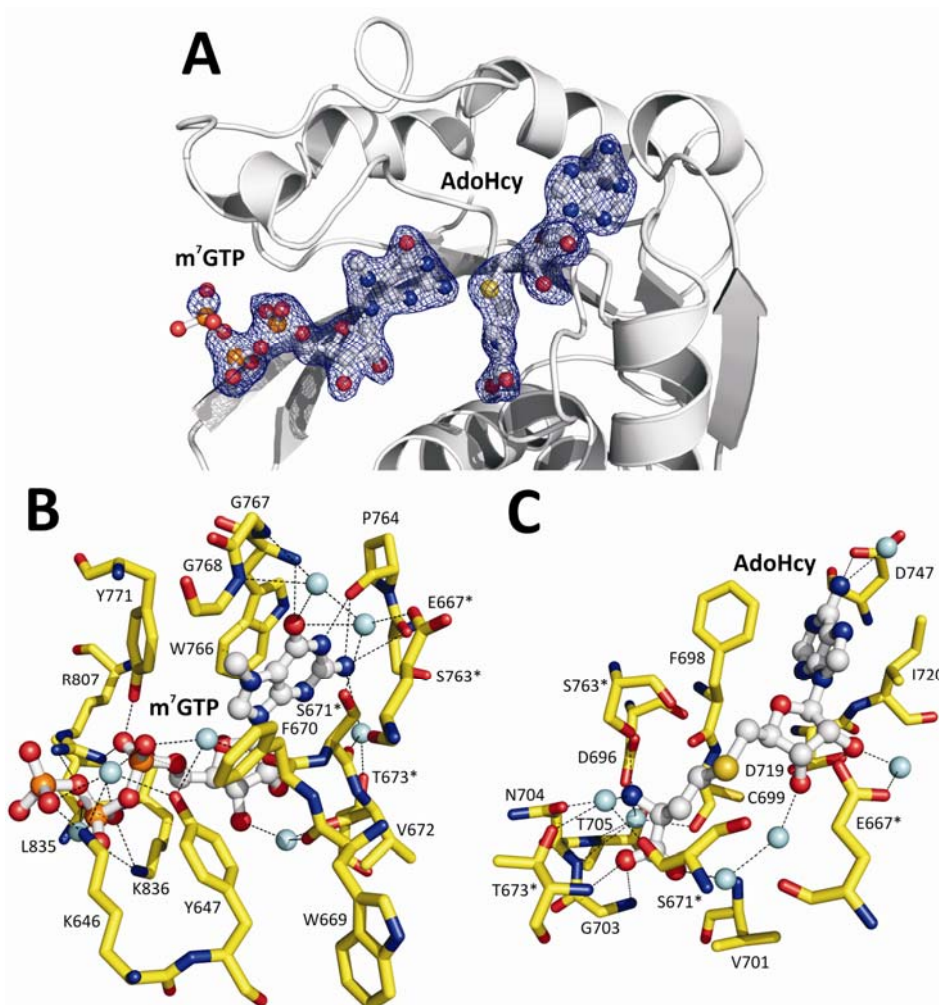


Figure 3. Binding of the minimal substrate 7-methylguanosine-triphosphate ($m^7\text{GTP}$) and the cofactor-analog *S*-adenosyl-L-homocysteine (AdoHcy) by hTGS1₆₁₈₋₈₅₃. **(A)** Overall binding of $m^7\text{GTP}$ and AdoHcy (ball-and-stick mode) and relative orientation of the substrates to each other. The substrate molecules are surrounded by a $|\text{Fo}| - |\text{Fc}|$ electron density omit map contoured at 2.9σ ($m^7\text{GTP}$ and AdoHcy were omitted). The helices $\alpha 1 - \alpha 4$ were removed in this view for clarity reasons. **(B)** Detailed view of the binding-pockets for $m^7\text{GTP}$ and **(C)** for AdoHcy. Protein residues involved in binding are labeled and drawn in stick mode (yellow) whereas the substrate-analogs are shown in ball-and-stick representation. Interactions of protein residues with water molecules (light blue) or substrate atoms (e.g. hydrogen bonds) are indicated by dashed lines. Amino acids participating in the binding of both ligands are marked by an asterisk.

The homocysteine-moiety of AdoHcy, more precisely the amino and the carboxyl group, are also bound by an extended network of hydrogen bonds to Ser671, Thr673, Asp696, Cys699, Val701, Gly703, Asn704, Thr705 and Ser763, respectively (Figure 3C). Three water molecules additionally contribute to this interaction. Within this array of interactions, the AdoHcy is tightly bound in its binding pocket and suggests that the site of the methyl group present in AdoMet is oriented towards the N2 of the bound cap-guanine. The distance between the AdoHcy-sulfur and the methyl group acceptor N2 of the cap-guanine averages 4 Å, representing an ideal spacing for a nucleophilic attack of N2 on the AdoMet-methyl group.

The cap-binding pocket is located next to the AdoHcy binding site and in contrast to AdoHcy, the m⁷GTP-ribose adopts a clear 2'-endo conformation (Figure 3A and B). The guanine of m⁷GTP, which is located next to the homocysteine-moiety of AdoHcy, is perfectly stacked on Trp766. Both aromatic rings are oriented almost coplanar with a distance of 3.2 Å leading to a tight π - π interaction between them (Figure 3B). This base stacking is further stabilized by a cation- π interaction, due to the net positive electric charge of the guanine caused by the methylation of N7. The side chain of Trp766 is fixed by the side chain of Pro806 leading to a higher rigidity of its aromatic side chain (data not shown). On the opposing side of the m⁷guanine the polar side chain of Ser671 limits the binding pocket. Further important contacts between the protein and the cap-analog involve all three phosphates α , β and γ , as well as the guanine nitrogens N2, N1 and the oxygen O6, generating the specificity for a guanine base. The carbonyl oxygen of Pro764 forms hydrogen bonds with N1 and N2, while the latter is further coordinated by a water molecule and the main chain carbonyl of Ser763. The O6 of m⁷GTP is bound by the main chain amide of the stacking Trp766 as well as by two water molecules, which are held in position by a complex interaction network with Ser671, Gly767 and Gly768. Moreover, the γ -phosphate of the cap-

analog is bound by Lys646 and a water molecule, whereas the β -phosphate is fixed by the same lysine residue and two additional interactions to atoms NH1 and NH2 of Arg807 (Figure 3B). The most pronounced interactions occur with the α -phosphate, which is positioned by Tyr771, Arg807 and two water molecules, respectively.

Besides the two substrate-analogs, two molecules of the cryo protectant 1,2-propanediol are bound to the cap-analog m⁷GTP and to the surface of the protein, respectively (data not shown). The first propylene glycol molecule binds *via* its hydroxyls to the β -phosphate and the 3'OH of the cap-nucleotide of all three molecules, whereas the second one is bound on the surface of the protein molecules 1 and 3 involving the residues Asp739 and Ala738, respectively. However, the interaction of 1,2-propanediol with m⁷GTP does not interfere with its interaction to the protein. The two hydroxyls of propylene glycol forming the hydrogen bonds are expected to be replaced by water molecules in an aqueous solution lacking the cryo protectant.

Mutagenesis studies

The impact of residues involved in substrate binding was analyzed for the crystallized fragment (hTGS1₆₁₈₋₈₅₃). The residues Trp766, Ser763, Asp696 and Phe804, which participate in substrate binding or facilitate catalysis, were mutated and their influence on catalytic activity was measured using the HPLC-based activity assay. The effect of individual residues on catalytic activity of TGS1 is summarized in Table 2. For this purpose, the area of the m⁷GTP-peak after 20 min of incubation with the accordant mutant protein was compared to the corresponding peak areas of a control reaction without enzyme (0% activity) and in the presence of the wild type enzyme (100% activity), respectively.

As expected, substitution of the stacking tryptophan by alanine (Trp766Ala) results in a complete loss of methyltransferase activity (<1% of wild-type activity).

Table 2. Effect of selected amino acid mutations on catalytic activity of hTGS1₆₁₈₋₈₅₃

Mutation	Activity
wt (hTGS1 ₆₁₈₋₈₅₃)	100 %
W766A	< 1 %
S763A	18 %
S763D	97 %
D696A	n/a
D696N	n/a
F804A	n/a

This observation underlines the ultimate importance of the aromatic indole ring being the exclusive cation- π and π - π stacking partner for the m⁷G-cap in the binding pocket. Further on, replacement of serine 763, which was postulated to be a key residue in catalysis by alanine (Ser763Ala) roughly reduces the catalytic activity of the enzyme by a factor of five. A mutant containing an aspartate instead of serine 763 (Ser763Asp), as present in some other methyltransferases and the *G. lamblia* TGS2, shows almost wild-type activity confirming the hypothesis that a proton acceptor in this position enforces catalysis. When we tried to purify the mutant containing an alanine instead of aspartate 696 (Asp696Ala), the protein immediately precipitated subsequent to elution from the first GSH-sepharose-affinity column. Nevertheless, in order to test the effect of the lacking carboxyl group on the side chain of Asp696 without losing the structural integrity, we generated the more conservative mutant Asp696Asn. Surprisingly, this mutation showed the same effect on protein stability as the exchange against alanine, revealing the strict requirement of aspartate 696. The obvious structural importance therefore prevents the detailed analysis of the catalytic impact of Asp696. The analysis of replacement of Phe804 by an alanine or by a positively charged residue (e.g. lysine) was impeded by a related problem. Phe804 in combination with other residues is thought to form a hydrophobic pocket harboring the first transferred methyl group and therefore supporting the second methylation reaction (see

below). However, replacement of this residue results in a significant reduction of bacterial cell growth during expression and mostly insoluble protein after cell lysis.

Potential protein and RNA interaction sites

In order to find potential binding sites of the hTGS1-methyltransferase domain for RNA nucleotides adjacent to the 5'-cap as well as for interacting proteins of the cognate RNPs, an electrostatic surface potential of the protein was calculated (Figure 4). The front view (left-hand side) shows TGS1 in the same orientation as in Figure 2 (left-hand side) with the bound substrates. In the side view (middle) basic patches in close proximity to the m⁷G-cap-binding pocket indicate the binding site for further RNA nucleotides. In fact, the first some ten nucleotides of the U1snRNP are accessible for interaction partners as demonstrated by RNase H digestion experiments (13). The back view (right-hand side) reveals acidic patches of hTGS1 methyltransferase domain, which are likely to bind the positively charged C-terminal tails of Sm-proteins B and D1. The interaction of RNP specific proteins and TGS1 in human and yeast was investigated *in extenso* (26,34,48).

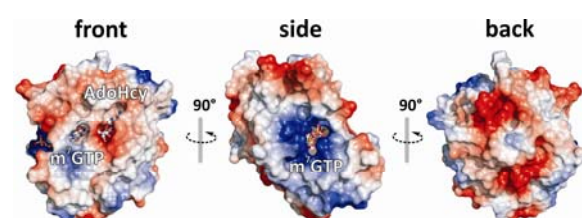


Figure 4. Charge distribution on the hTGS1 surface. Basic patches are colored in blue, while acidic ones are depicted in red. The three views are rotated by 90° with respect to each other. The front view has the same orientation as the left panel in Figure 2, while the side view shows the putative binding region for the ongoing UsnRNA backbone (blue patches). The back view reveals two acidic regions, representing potential interaction sites for the Sm-proteins B and D1 or other proteins.

DISCUSSION

The RNA 5'-cap dimethyltransferases from diverse organisms have been biochemically characterized, however no crystal structure of an

active enzyme has been available allowing an investigation of the structure-function relationship. In the present study we report the crystal structure of the active C-terminal methyltransferase domain of human m⁷G-cap-specific dimethyltransferase TGS1 bound to the minimal substrate m⁷GTP as well as the reaction product *S*-adenosyl-L-homocysteine (AdoHcy). Recently, it was reported that the canonical and structurally conserved methyltransferase domain of hTGS1 is not sufficient for catalytic activity, but requires 17 additional residues to gain catalytic activity (44). Figure 5A shows the superposition of the active (hTGS1_{618–853}) and inactive (hTGS1_{653–853}) form of the hTGS1

methyltransferase domain. The structural reason for the loss of function of the shorter hTGS1 fragment is a significant change in the conformation of the N-terminal extension (NTE), which forms a small α -helical subdomain in the active state (blue part in Figure 5), while it contains three β -strands in the inactive state (red part). The NTE is essential for m⁷G-cap binding and as it completes the substrate binding pocket. Lys646, Tyr647, Glu667, Phe670, Ser671, Val672 and Thr673 of this NTE are involved in m⁷G-cap binding (Figure 3). The NTE is mainly stabilized by its hydrophobic core consisting of the residues Ile638, Leu644, Trp648, Leu654, Phe655, Leu664, Gly668 and Trp669 (Figure 5B). The truncated NTE of hTGS1_{653–853} adopts a different fold, since important residues of the hydrophobic core are missing, resulting in the formation of three β -strands instead of α -helices present in the NTE of the active form.

m⁷G-cap-binding mode

To date several three-dimensional structures of m⁷G-cap-binding proteins are known, clearly defining their binding characteristics for the positively charged, N7-methylated guanine of the 5'-cap structure. Among them are the eukaryotic initiation factor 4E (eIF4E) (50) the viral protein 39 (VP39) (51) and the small subunit of the cap-binding complex CBP20 (52) all of which show a tight stacking of the cap-N7-guanine between two aromatic side chains (eIF4E, Trp56/Trp102; VP39, Tyr22/Phe180; CBP20, Tyr20/Tyr43) (Figure 6). Differing thereof, the scavenger decapping enzyme DcpS (53) and the influenza virus polymerase subunit PB2 (54) stack the m⁷G-cap only on one side by an aromatic and by a non-aromatic side chain on the other (DcpS, Trp175/Leu206; PB2, Phe404/His357). Recently, it was shown that in the poly(A)-specific ribonuclease (PARN) a second stacking residue is completely missing and the N7-methylguanine is mainly bound by the aromatic Trp475 (Figure 6) (55–57).

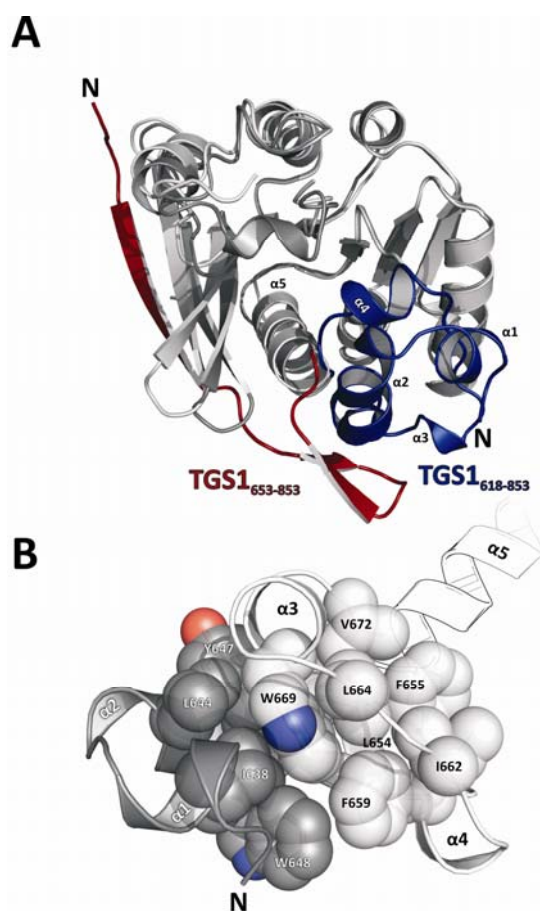


Figure 5. Superposition of the active (aa 618–853; NTE in blue) and inactive (aa 653–853; NTE in red; PDB ID 3EGI) hTGS1 methyltransferase domain (A) and detailed view on the hydrophobic core of the NTE of the active conformation (B). Hydrophobic residues of the NTE are drawn in space-filling spheres to clarify the core. Residues that are not present in the inactive structure (Leu634–Arg651) and which contribute to the hydrophobic core are drawn in dark grey and labeled in white, while amino acids from 652 on are colored in white and labeled in black.

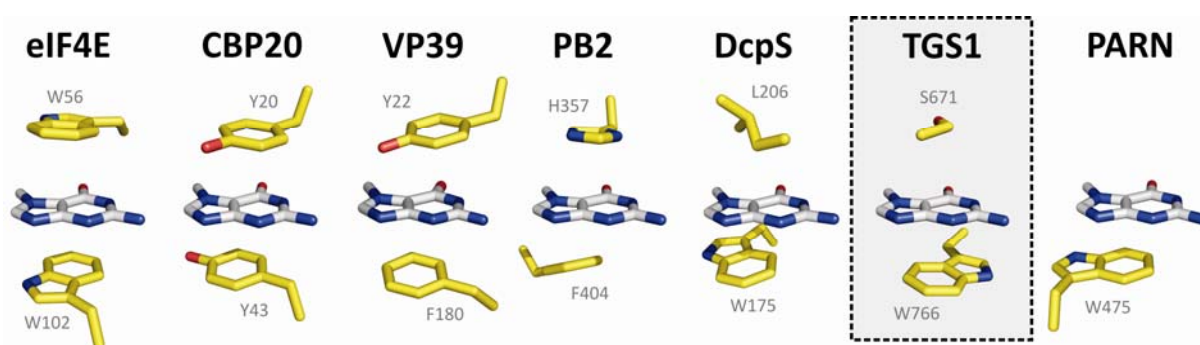


Figure 6. Comparison of the cap binding pockets of several cap binding proteins. The positively charged, N7-methylated guanine-moiety (light grey) is stacked in between or flanked by the individual, labeled protein residues (yellow). The interaction of m⁷guanine with the eukaryotic initiation factor 4E (eIF4E; PDB ID 1L8B), the small subunit of the cap binding complex (CBP20; PDB ID 1H2T), the vaccinia virus protein 39 (VP39; PDB ID 1AV6), the influenza virus polymerase subunit PB2 (PDB ID 2VQZ), the scavenger decapping enzyme (DcpS; PDB ID 1ST0), hTGS1_{618–853} and the poly(A)-specific ribonuclease (PARN; PDB ID 3CTR) is shown.

In order to gain insight into the cap binding mode of TGS1 the structural information of the protein Mj0882 (PDB ID 1DUS), which was shown to be a protein related to a family comprised of bacterial rRNA methyltransferases (RsmC /RsmD) was used to generate a homology model of yTGS1 (48). Based on this homology model of yTGS1, it was suggested that the m⁷G-cap is stacked between the two side chains of Trp178 and Trp75 (which correspond to Trp766 and Trp669 in hTGS1). However, this cap binding mode remained questionable, as mutation of Trp75 in yTGS1 to an alanine did not decrease catalytic activity (48). Indeed the crystal structure of hTGS1_{618–853} reveals that only Trp766 (yTGS1 Trp178) stacks the m⁷G-cap, while Trp669 (yTGS1 Trp75) only makes a water-mediated contact with the 3'OH of m⁷GTP *via* the main chain carbonyl (Figure 3B). Moreover, the side chain of Trp669 is part of the hydrophobic core of the NTE, interacting for example with Leu664 and Ile638 (Figure 5B). The common second aromatic side chain sandwiching the m⁷G-cap is replaced by the side chain of Ser671 that limits the cap-binding pocket spatially on the side opposing Trp766, allowing for some flexibility of the NTE (Figures 3B and 6). The structural data now explain the mutagenesis studies mentioned above.

Mutagenesis studies

The functional relevance of residues that are involved in substrate binding was analyzed by studying several single amino-acid mutants. As expected, the replacement of the stacking Trp766 (Trp766Ala) results in a complete loss of catalytic activity, whereas the exchange of another putative key residue in the active site, Ser763, has a milder effect. The mutation Ser763Ala decreases the activity to 18%. Contrary to our results the yTGS1 mutant Ser175Ala (corresponding to Ser763 in human TGS1) showed no defect in activity as shown indirectly by the immunoprecipitation of cellular RNA (48), while the corresponding exchange in the fruit fly TGS1 ortholog (DTL) was lethal to the larvae (35). The mutant Asp696Ala, as well as a more conservative mutation to an asparagine (Asp696Asn) resulted in insoluble and precipitating hTGS1, demonstrating the importance of this residue to maintain a structurally stable enzyme and an intact catalytic pocket. In contrast, Hausmann *et al.* reported the purification of this mutant and its activity to be reduced to below 1% (30). In this previous study additional six residues (Phe655, Thr673, Asn704, Asp719, Asn731, Arg807 and Asn808) predicted to be important for activity were mutated to alanine and characterized using a radioactive methylation assay (30). In summary, the results from this mutational analysis can be

explained by the structure analysis of hTGS1_{618–853}. The residues Thr673 and Asn704 establish water mediated hydrogen bonds to the reaction product AdoHcy in the structure and their mutation to alanine reduces the activity to 13% and 5%, respectively. Asp719, which makes contacts to both ribose hydroxyls and therefore significantly contributes to a binding of AdoHcy, reduces the catalytic activity to <1% when mutated to alanine. Both amines NH1 and NH2 of the guanidinium group of Arg807 form hydrogen bonds to the α - and the β -phosphate of the cap-analog, thus the mutation Arg807Ala results in a decrease of activity to 8% compared to the wild type protein. Replacement of Phe655, which is part of the hydrophobic core (Figure 5B), with alanine reduced the measured activity to less than 1% indicating its importance in the stabilization of the NTE and thus its connection to the methyltransferase core. Another important residue Asn731 is located in α -helix 7 and makes two contacts to the main chain carbonyl and amine of Lys663, which belongs to the NTE. The mutation Asn731Ala reduces the activity to 4% compared to the wild type protein. Located in the cleft between both domains it contributes significantly to the stability of their interaction and thus promotes the active conformation of the enzyme. A related function is fulfilled by the side chain N δ of Asn808, which forms two hydrogen bonds to the main chain carbonyls of Ala774 (2.8 Å) and Tyr771 (3.2 Å) the latter of which is directly positioned in α -helix 9. Mutation of this asparagine to an alanine decreases the TGS1 activity to about 11%. This region encompassing the residues (Gly767 to Thr773) is not visible in the inactive structure of the hTGS1_{653–853}, which does not contain a correctly bound substrate. Thus the interactions between Asn808 and the mentioned carbonyls help to fix this region in a defined conformation competent to bind the cap-analog m⁷GTP.

Catalytic mechanism

Catalysis by most methyltransferases is known to proceed by an S_N2 substitution reaction (58). On the basis of the presented structural and

functional characterization of the human TGS1 methyltransferase domain in complex with substrate-analogs the following model for the catalytic mechanism of dimethylation by TGS1 is proposed. The exocyclic N2 of the m⁷G performs a nucleophilic attack on the activated methyl group of the AdoMet. The distance of 4 Å between the AdoHcy sulfur and the cap N2 atoms observed in our crystal structure is consistent with that mechanism (59). In order to increase the nucleophilicity of the N2 its hybridization state has to be changed from sp² to sp³. This transition might be stabilized by the main chain carbonyl oxygens of Ser763 and Pro764, which are in reasonable distance and a productive orientation (Figure 7). Such a mechanism was previously proposed for N6-adenine DNA methyltransferase MTaqI (60). Accordingly, the carbonyl oxygens of Pro764 and Ser763, which are both located out of the planar purine ring pull the hydrogens of N2 out of the conjugated system. In TGS1 the positive charge of the methylated guanine leads to an increased electron pull on N2 facilitating its deprotonation. This generates a lone pair of electrons that is no longer part of the conjugated system leading to an enhanced nucleophilicity of N2, which is required to attack the methyl group. During this process the exocyclic NH₂ has to release a proton which most likely is accepted by a protein residue. A putative proton acceptor is the hydroxyl group of Ser763, which is, however, in a distance of 4.2 Å. Interestingly, a different rotamer of this side chain would reduce the distance to 2.5 Å, which would allow the direct proton transfer. In the observed conformation the serine hydroxyl group forms a hydrogen bond to a water molecule, which in turn is additionally bound to the carboxyl group of Asp696 and the main chain carbonyl group of Phe761. Therefore it appears likely that the proton is transferred *via* the serine side chain to the water molecule. However, this role of Ser763 is not essential, since its mutation to alanine reduces the activity to 18%. Furthermore, the exchange of Ser763 with an aspartate residue nearly retains catalytic activity compared to the

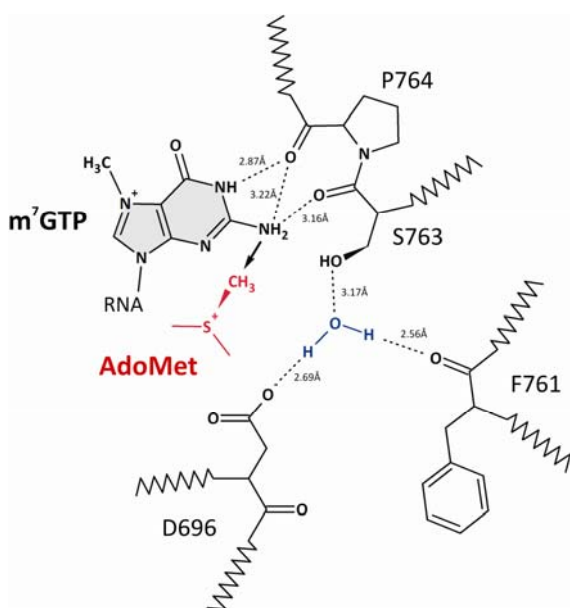


Figure 7. Proposed catalytic mechanism of methyltransfer by hTGS1. The methylation proceeds by an S_N2 substitution reaction. The exocyclic N2-amino group forms hydrogen bonds with the carbonyls of Ser763 and Pro764 of the motif $^{763}\text{SPPW}^{766}$ and the release of the hydrogen is further promoted by the net positive charge distributed over the whole methylated purine ring. Once shifted towards the tetrahedral state (sp^3) the N2 attacks the methyl group. After exchange of AdoHcy with a new AdoMet molecule the N2 methyl group is located outside of the ring plane and stabilized by a hydrophobic pocket formed by Trp766, Pro764 and Phe804. The second methylation then applies to the same mechanism as the first one with an additional activation of N2 by the methyl group. The proton release is facilitated further by the side chain of Ser763 connected to a water molecule, which on its own forms a hydrogen bond with the side chain carboxyl of Asp696 and the Phe761 carbonyl, respectively.

wild type enzyme (97%). This mutation is observed in some other methyltransferases like in the *G. lamblia* TGS2 and a mimivirus methyltransferase (mimiTGS), altering the main catalytic motif from $^{763}\text{SPPW}^{766}$ to DPPW (42,61). In fact, for the tRNA N2, N2-guanosine dimethyltransferase Trm1 it was postulated that the negatively charged carboxyl group of the aspartate in this motif is the general base deprotonating the guanine N2 (62). After the first methylation the reaction products AdoHcy and $m^{2,7}\text{G-cap}$ dissociate from the active site, allowing the binding of two new substrates. The assumption that the dimethylation is not a processive reaction is supported by fact that *S. pombe* TGS1 under $m^7\text{GDP}$ -excess conditions

converts almost exclusively $m^7\text{GDP}$ to $m^{2,7}\text{GDP}$. Only when the educt $m^7\text{GDP}$ was limiting, the enzyme converted the reaction intermediate further to the end product $m^{2,2,7}(\text{G})\text{GDP}$ in a pulse-chase experiment (27). After binding of the intermediate $m^{2,7}\text{GTP}$, the second methylation presumably applies to the same mechanism. The N2 is activated additionally by the first bound methyl group located outside of the plane of the guanine ring. This methyl group would fit in an adjacent hydrophobic pocket formed by Trp766, Pro764 and Phe804 (data not shown). The activated N2 attacks the newly bound AdoMet-methyl group, leading to the second methylation reaction, which results in the fully modified $m^{2,2,7}\text{G-cap}$.

Comparison with structurally and functionally related methyltransferases

The three-dimensional structures of several other RNA-modifying methyltransferases targeting exocyclic amino groups of nucleobases are known. Among them are the rRNA-adenine dimethyltransferase KsgA (63), the rRNA guanine methyltransferase RsmC (64) and the tRNA-guanine-N2 dimethyltransferase Trm1 (62). The methyltransferase core domains of KsgA, RsmC, Trm1 and hTGS1, encompassing the canonical $\alpha\beta$ sandwich, are structurally very similar, whereas the remaining parts of these enzymes differ significantly.

RsmC contains two separate methyltransferase domains, with only the C-terminal one binding the substrate-guanosine and AdoMet in the crystal structure (64). Unlike hTGS1, RsmC contains the four residues $^{305}\text{NPPF}^{308}$ in both the N-terminal and the C-terminal methyltransferase domains. RsmC binds the guanosine mainly by its stacking interaction with the aromatic side chain of Phe308. However, the distance of 4.7 Å between the AdoMet methyl group and the exocyclic target amino group suggests that the substrates are in a non-productive orientation. Interestingly, the *Pyrococcus horikoshii* Trm1 contains a C-terminal extension, which is proposed to be involved in tRNA-binding and composed of six α -helices and six β -strands

forming the three subdomains C1–C3 in addition to the conserved $\alpha\beta\alpha$ -sandwich. The N-terminus contains two additional β -strands generating a small β -sheet enclosing α -helix 1, which represents the first helix of the methyltransferase domain and a two β -strand insertion between α -helix 3 and β -strand 7 elongating the present β -sheet to nine strands.

The four related methyltransferases KsgA, RsmC, Trm1 and hTGS1 bind the cofactor AdoMet or its analog AdoHcy in a structurally conserved binding site, but the residues that contribute to the binding of AdoMet are not strictly conserved. The common catalytic motif of KsgA, RsmC and hTGS1, which has the sequence $^1(\text{N/D/S})\text{-}^2(\text{P/L})\text{-}^3(\text{P})\text{-}^4(\text{W/F/Y})$, is altered in Trm1 as it lacks the second residue. The first residue of this motif is involved in the activation of the methyl group acceptor, which is held in position by the two main chain carbonyl oxygens of the first and second residues. The aromatic side chain of the last motif residue is generally believed to bind the methyl-acceptor nucleobase by stacking interaction as experimentally demonstrated for RsmC and hTGS1.

In summary, our structural studies on hTGS1 revealed the necessity of an unexpected N-terminal and helical extension required for m⁷G-cap binding and catalytic activity in addition to the canonical methyltransferase domain.

ACCESSION NUMBER

Accession number: PDB ID: 3GDH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Utz Fischer and Rémy Bordonné for the donation of human pGEX-6P-1-TGS1. We thank Bernhard Kuhle for excellent technical assistance as well as our colleague Markus G. Rudolph for help with the activity

assay and many scientific discussions. Moreover, we are very thankful to the staff of beamline BL14.2 at the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY) for assistance during data measurements.

FUNDING

Deutsche Forschungsgemeinschaft (DFG) Sonderforschungsbereich 523 (TP A11); Synchrotron data collection was partially supported by the Bundesministerium für Bildung und Forschung (BMBF 05 ES3XBA/5). Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

1. Dickmanns, A. (2009) Import and Export of snRNPs. In Kehlenbach, R. (ed.), *Nuclear Transport*. Landes Bioscience, Austin, USA.
2. Dickmanns, A. and Ficner, R. (2005) Role of the 5'-cap in the biogenesis of spliceosomal snRNPs. *Top. Curr. Genet.*, **12**, 179–204.
3. Cougot, N., van Dijk, E., Babajko, S. and Seraphin, B. (2004) Cap-tabolism. *Trends Biochem. Sci.*, **29**, 436–444.
4. Hernandez, N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.*, **276**, 26733–26736.
5. Kambach, C., Walke, S. and Nagai, K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.*, **9**, 222–230.
6. Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, 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.
7. 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.

8. Mattaj, I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905–911.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Mattaj, I.W. and De Robertis, E.M. (1985) Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell*, **40**, 111–118.
15. 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.
16. 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.
17. 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.
18. Strasser, A., Dickmanns, A., Luhrmann, R. and Ficner, R. (2005) Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J.*, **24**, 2235–2243.
19. Huber, J., Dickmanns, A. and Luhrmann, R. (2002) The importin beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. *J. Cell Biol.*, **156**, 467–479.
20. Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290–301.
21. Krol, A., Carbon, P., Ebel, J.P. and Appel, B. (1987) *Xenopus tropicalis* U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes. *Nucleic Acids Res.*, **15**, 2463–2478.
22. 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.
23. 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.
24. Sleeman, J.E., Ajuh, P. and Lamond, A.I. (2001) snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *J. Cell Sci.*, **114**, 4407–4419.
25. Sleeman, J.E. and Lamond, A.I. (1999) Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.*, **9**, 1065–1074.
26. 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.
27. Hausmann, S. and Shuman, S. (2005) Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J. Biol. Chem.*, **280**, 4021–4024.
28. Colau, G., Thiry, M., Leduc, V., Bordonne, R. and Lafontaine, D.L. (2004) The small nucleolar RNA cap trimethyltransferase is required for ribosome synthesis and intact nucleolar morphology. *Mol. Cell Biol.*, **24**, 7976–7986.
29. Zhu, Y., Qi, C., Cao, W.Q., Yeldandi, A.V., Rao, M.S. and Reddy, J.K. (2001) Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc. Natl Acad. Sci. USA*, **98**, 10380–10385.
30. Hausmann, S., Zheng, S., Costanzo, M., Brost, R.L., Garcin, D., Boone, C., Shuman, S. and Schwer, B. (2008) Genetic and biochemical analysis of yeast and human cap Trimethylguanosine synthase: Functional overlap of TMG caps, snRNP

- components, pre-mRNA splicing factors, and RNA decay pathways. *J. Biol. Chem.*, **283**, 31706–31718.
31. Misra,P., Qi,C., Yu,S., Shah,S.H., Cao,W.Q., Rao,M.S., Thimmapaya,B., Zhu,Y. and Reddy,J.K. (2002) Interaction of PIMT with transcriptional coactivators CBP, p300, and PBP differential role in transcriptional regulation. *J. Biol. Chem.*, **277**, 20011–20019.
32. Enunlu,I., Papai,G., Cserpan,I., Udvardy,A., Jeang,K.T. and Boros,I. (2003) Different isoforms of PRIP-interacting protein with methyltransferase domain / trimethylguanosine synthase localizes to the cytoplasm and nucleus. *Biochem. Biophys. Res. Commun.*, **309**, 44–51.
33. Girard,C., Verheggen,C., Neel,H., Cammas,A., Vagner,S., Soret,J., Bertrand,E. and Bordonne,R. (2008) Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J. Biol. Chem.*, **283**, 2060–2069.
34. 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 atrophy protein, survival of motor neuron. *EMBO Rep.*, **4**, 616–622.
35. Komonyi,O., Papai,G., Enunlu,I., Muratoglu,S., Pankotai,T., Kopitova,D., Maroy,P., Udvardy,A. and Boros,I. (2005) DTL, the *Drosophila* homolog of PIMT/Tgs1 nuclear receptor coactivators interacting protein / RNA methyltransferase, has an essential role in development. *J. Biol. Chem.*, **280**, 12397–12404.
36. Terns,M.P. and Dahlberg,J.E. (1994) Retention and 5' cap trimethylation of U3 snRNA in the nucleus. *Science*, **264**, 959–961.
37. Watkins,N.J., Lemm,I., Ingelfinger,D., Schneider,C., Hossbach,M., Urlaub,H. and Luhrmann,R. (2004) Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell*, **16**, 789–798.
38. Terns,M.P., Grimm,C., Lund,E. and Dahlberg,J.E. (1995) A common maturation pathway for small nucleolar RNAs. *EMBO J.*, **14**, 4860–4871.
39. Franke,J., Gehlen,J. and Ehrenhofer-Murray,A.E. (2008) Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J. Cell Sci.*, **121**, 3553–3560.
40. Gunzl,A., Bindereif,A., Ullu,E. and Tschudi,C. (2000) Determinants for cap trimethylation of the U2 small nuclear RNA are not conserved between *Trypanosoma brucei* and higher eukaryotic organisms. *Nucleic Acids Res.*, **28**, 3702–3709.
41. Hausmann,S., Ramirez,A., Schneider,S., Schwer,B. and Shuman,S. (2007) Biochemical and genetic analysis of RNA cap guanine-N2 methyltransferases from *Giardia lamblia* and *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **35**, 1411–1420.
42. Hausmann,S. and Shuman,S. (2005) *Giardia lamblia* RNA cap guanine-N2 methyltransferase (Tgs2). *J. Biol. Chem.*, **280**, 32101–32106.
43. Ruan,J.P., Ullu,E. and Tschudi,C. (2007) Characterization of the *Trypanosoma brucei* cap hypermethylase Tgs1. *Mol. Biochem. Parasitol.*, **155**, 66–69.
44. Monecke,T., Dickmanns,A., Strasser,A. and Ficner,R. (2009) Structure analysis of the conserved methyltransferase domain of human trimethylguanosine synthase TGS1. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 332–338.
45. McCoy,A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 32–41.
46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
47. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
48. Mouaikel,J., Bujnicki,J.M., Tazi,J. and Bordonne,R. (2003) Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res.*, **31**, 4899–4909.
49. Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
50. Niedzwiecka,A., Marcotrigiano,J., Stepinski,J., Jankowska-Anyszka,M., Wyslouch-Cieszyńska,A., Dadlez,M., Gingras,A.C., Mak,P., Darzynkiewicz,E., Sonenberg,N. *et al.* (2002) Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 50 cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol.*, **319**, 615–635.
51. Hodel,A.E., Gershon,P.D. and Quijcho,F.A. (1998) Structural basis for sequence-nonspecific

- recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol. Cell*, **1**, 443–447.
52. Mazza, C., Segref, A., Mattaj, I.W. and Cusack, S. (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.*, **21**, 5548–5557.
53. Gu, M., Fabrega, C., Liu, S.W., Liu, H., Kiledjian, M. and Lima, C.D. (2004) Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol. Cell*, **14**, 67–80.
54. Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W., Ortin, J., Hart, D.J. *et al.* (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, **15**, 500–506.
55. Monecke, T., Schell, S., Dickmanns, A. and Ficner, R. (2008) Crystal structure of the RRM domain of poly(A)-specific ribonuclease reveals a novel m(7)G-cap-binding mode. *J. Mol. Biol.*, **382**, 827–834.
56. Nagata, T., Suzuki, S., Endo, R., Shirouzu, M., Terada, T., Inoue, M., Kigawa, T., Kobayashi, N., Guntert, P., Tanaka, A. *et al.* (2008) The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Res.*, **36**, 4754–4767.
57. Wu, M., Nilsson, P., Henriksson, N., Niedzwiecka, A., Lim, M.K., Cheng, Z., Kokkoris, K., Virtanen, A. and Song, H. (2009) Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease. *Structure*, **17**, 276–286.
58. Schubert, H.L., Blumenthal, R.M. and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.*, **28**, 329–335.
59. Zheng, Y.J. and Bruice, T.C. (1997) A theoretical examination of the factors controlling the catalytic efficiency of a transmethylation enzyme: catechol O-methyltransferase. *J. Am. Chem. Soc.*, **119**, 8137–8145.
60. Goedecke, K., Pignot, M., Goody, R.S., Scheidig, A.J. and Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Mol. Biol.*, **8**, 121–125.
61. Benarroch, D., Qiu, Z.R., Schwer, B. and Shuman, S. (2009) Characterization of a mimivirus RNA cap guanine-N2 methyltransferase. *RNA*, **15**, 666–674.
62. Ihsanawati, Nishimoto, M., Higashijima, K., Shirouzu, M., Grosjean, H., Bessho, Y. and Yokoyama, S. (2008) Crystal structure of tRNA N2,N2-guanosine dimethyltransferase Trm1 from *Pyrococcus horikoshii*. *J. Mol. Biol.*, **383**, 871–884.
63. O'Farrell, H.C., Scarsdale, J.N. and Rife, J.P. (2004) Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in *Escherichia coli*. *J. Mol. Biol.*, **339**, 337–353.
64. Demirci, H., Gregory, S.T., Dahlberg, A.E. and Jögl, G. (2008) Crystal structure of the *Thermus thermophilus* 16 S rRNA methyltransferase RsmC in complex with cofactor and substrate guanosine. *J. Biol. Chem.*, **283**, 26548–26556.

5.1 Supplementary Data

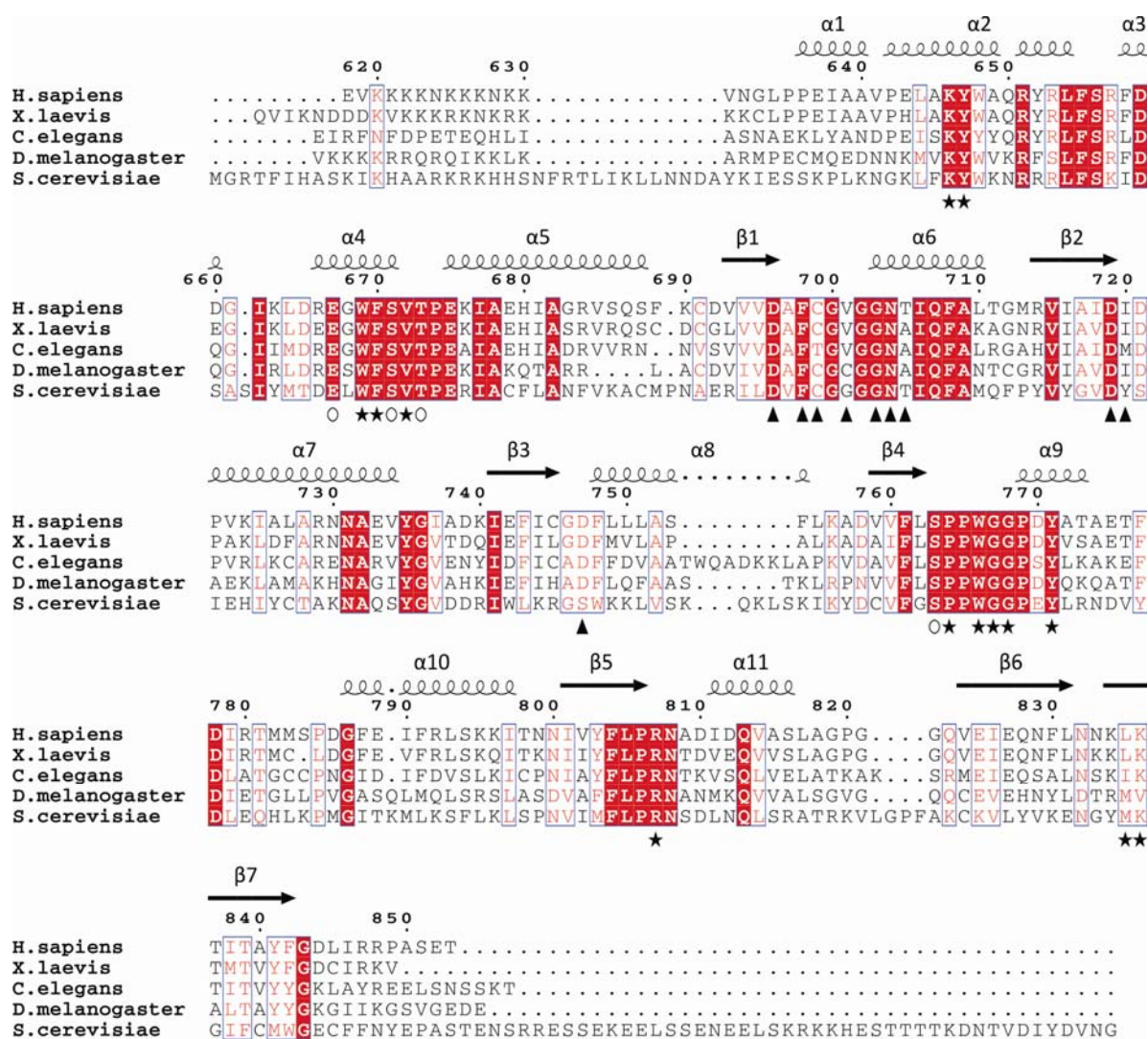


Figure S1. Alignment of TGS1 methyltransferase domain from *H. sapiens*, *X. laevis*, *C. elegans*, *D. melanogaster* and *S. cerevisiae* using CLUSTALW and ESPRIPT. Identical residues are shown in white on red background, while similar residues are drawn in red on white background. The secondary structure elements of active hTGS1 methyltransferase domain are displayed on the top of each row and the appropriate residue numbers are indicated. Residues involved in m⁷GTP-binding are marked by an asterisk, those binding to AdoHcy by a triangle. Amino acids involved in the binding of both ligands are assigned by a circle.

Chapter 6 • Crystal structure of the nuclear export complex CRM1·SPN1·RanGTP

This manuscript has originally been published in *Science*.

'Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP'

Thomas Monecke^{1,†}, Thomas Güttler^{2,†}, Piotr Neumann¹, Achim Dickmanns¹, Dirk Görlich^{2,‡} and Ralf Ficner¹

¹ Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik, Abteilung für Molekulare Strukturbiochemie (Göttinger Zentrum für Molekulare Biowissenschaften (GZMB), Ernst-Caspari-Haus), Justus-von-Liebig-Weg 11, D-37077, Göttingen

² Max-Planck-Institut für biophysikalische Chemie, Abteilung für Zelluläre Logistik, Am Faßberg 11, D-37077, Göttingen

[†] These Authors contributed equally to this work.

[‡] To whom correspondence should be addressed: E-mail: goerlich@mpibpc.mpg.de
Phone: 0551 – 201 2401
Fax: 0551 – 201 2407

Keywords: nuclear export complex; CRM1; exportin1; snurportin1; RanGTP

Science; Science express, Publication ahead of print, published online April 23, 2009
(Printed version: *Science*; VOL. 324, pp. 1087-91, May, 2009)
© 2009 by the American Association for the Advancement of Science.

Received March 11, 2009; Accepted April 9, 2009; published online April 23, 2009

Preface – About the Manuscript

The nuclear export receptor CRM1 binds the snRNP import adapter SPN1 in the nucleus and recycles it to the cytoplasm. Besides SPN1, CRM1 exports hundreds of proteins and protein complexed RNA molecules to the cytoplasm. However, the molecular mechanisms underlying export complex assembly as well as the structural organization of the complex so far remained unknown. The following publication reports the crystal structure determination and biochemical characterization of this nuclear export complex. The analysis of the crystal structure reveals that the export signature of SPN1 is tripartite and that the export receptor unexpectedly binds the cargo on its outer surface. This binding mode poses a plausible explanation for the extremely broad substrate specificity and ability of CRM1 to export even huge macromolecular assemblies such as ribosomal subunits which cannot be enwrapped by the exportin. Furthermore, a model for cooperative binding of CRM1, SPN1 and RanGTP is presented.

The coordinates and structure factors of the crystal structure described in the following publication have been deposited in the protein data bank (PDB) under the PDB ID 3GJX.

Author contributions:

Prof. Dr. Ralf Ficner, Dr. Achim Dickmanns and Prof. Dr. Dirk Görlich initiated the project. Protein purification was performed by Thomas Güttler (RanGTP and CRM1) and me (SPN1). Besides protein purification, my contribution comprised the crystallization, structure determination as well as structure analysis under supervision of Prof. Dr. Ralf Ficner. Dr. Piotr Neumann analyzed and evaluated diffraction data and made major contributions to structure solution and model building. Thomas Güttler and Prof. Dr. Dirk Görlich performed the biochemical experiments and subsequent analyses.

ABSTRACT

CRM1 mediates nuclear export of numerous unrelated cargoes, which may carry a short leucine-rich nuclear export signal or export signatures that include folded domains. How CRM1 recognizes such a variety of cargoes has been unknown. Here we present the crystal structure of the snurportin1-CRM1-RanGTP export complex at 2.5 Å resolution. Snurportin1 is a nuclear import adapter for cytoplasmically-assembled, m₃G-capped spliceosomal U snRNPs. The structure shows how CRM1 can specifically return the cargo-free form of snurportin1 to the cytoplasm. The extensive contact area includes five hydrophobic residues at the snurportin1 N terminus that dock into a hydrophobic cleft of CRM1, as well as numerous hydrophilic contacts of CRM1 to m₃G cap-binding domain and C-terminal residues of snurportin1. The structure suggests that RanGTP promotes cargo-binding to CRM1 solely through long-range conformational changes in the exportin.

Nuclear transport proceeds through nuclear pore complexes (NPCs) and supplies cell nuclei with proteins and the cytoplasm with nuclear products such as ribosomes and tRNAs. Most nuclear transport pathways are mediated by importin β -type nuclear transport receptors, which comprise nuclear export receptors (exportins) as well as importins (1, 2). These receptors bind cargoes directly or through adapter-molecules, shuttle constantly between nucleus and cytoplasm, and use the chemical potential of the nucleocytoplasmic RanGTP-gradient to act as unidirectional cargo-pumps (3). Exportins recruit cargo at high RanGTP-levels in the nucleus, traverse NPCs as ternary cargo-exportin-RanGTP complexes, and release their cargo upon GTP-hydrolysis into the cytoplasm. CRM1 (exportin1/Xpo1p) (4, 5) and CAS (Cse1p/exportin2) (6) are the prototypical exportins. While CAS is specialized to retrieve the nuclear import adapter importin α back to the cytoplasm (6), CRM1 exports a very broad range of substrates from nuclei (4, 5, 7–11), including ribosomes and many regulatory proteins. It also depletes translation factors from nuclei and is essential for the replication of viruses like HIV. CRM1 has a dual function during biogenesis of spliceosomal U snRNPs. It exports m⁷G-capped U snRNAs to the cytoplasm (4, 12), where they recruit Sm-core proteins and receive a 2,2,7-trimethyl (m₃G) cap structure. The import adapter snurportin 1 (SPN1) and importin β then

transport the mature m₃G-capped U snRNPs into nuclei (13). To mediate another import cycle, SPN1 is returned to the cytoplasm by CRM1 (14). Many CRM1-cargoes harbor a leucine-rich nuclear export signal (NES), which typically comprises four characteristically-spaced hydrophobic residues (7). Examples are the HIV-Rev protein (15) or the protein kinase A inhibitor PKI (16). In other cases, however, CRM1 recognizes not just a short peptide, but instead a large portion of the export cargo – here, SPN1 is the prototypical example (14). CRM1 binds SPN1 tighter than other export substrates, apparently because CRM1 must displace the imported U snRNP from SPN1 before export may occur. The cytoplasmic dissociation of CRM1 from SPN1 is essential for multi-round import of U snRNPs. Hydrolysis of the Ran-bound GTP alone is insufficient to fully disrupt the interaction (Fig. 1, A to C) (14), but importin β can displace CRM1 from SPN1 (Fig. 1A). Thus, either the binding sites of SPN1 for CRM1 and importin β overlap, or importin β forces SPN1 into a conformation, that is incompatible with CRM1-binding. Two functional domains in SPN1 have been described, the m₃G cap-binding domain (SPN1^{97–300}) (17) and the Nterminal IBB-domain (SPN1^{40–65}) (14, 18, 19), which confers binding to and import by importin β (20). A multiple alignment of SPN1 from various species revealed another conserved region that precedes the IBB-domain and includes the hydrophobic residues Leu⁴,

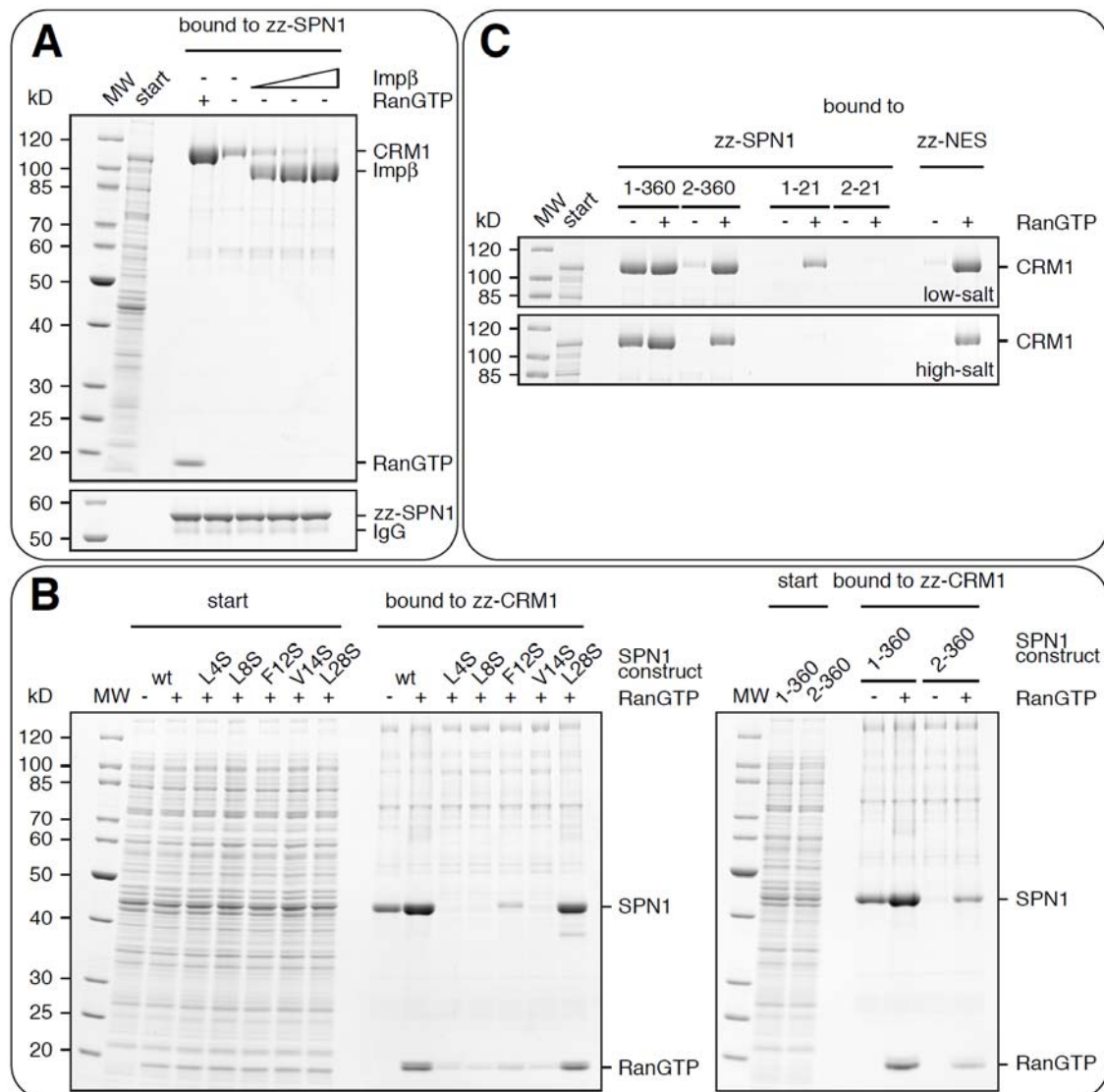


Fig. 1. (A) Effects of RanGTP and importin β on the SPN1-CRM1 interaction. 1 μ M SPN1 was mixed with an *E. coli* lysate containing 200 mM NaCl, 1 μ M CRM1 and the indicated combinations of 3 μ M RanGTP and 1, 2, or 3 μ M importin β (27). Complexes were retrieved by IgG-Sepharose via the zz-tag of SPN1. SPN1-ligands were eluted with 1.5 M $MgCl_2$ (upper panel); the remaining baits (zz-SPN1; lower panels) with SDS. Analysis was by SDS-PAGE/Coomassiestaining. Note that RanGTP enhanced CRM1-binding to SPN1, however this interaction was also detectable in the absence of Ran. This residual CRM1-SPN1 interaction could be suppressed by importin β that binds the IBB-domain of SPN1. (B) Met¹, Leu⁴, Leu⁸, Phe¹² and Val¹⁴ of SPN1 are all required for high-affinity binding to the CRM1-RanGTP complex. zz-tagged CRM1 immobilized on IgG-Sepharose was incubated with an *E. coli* lysate containing 200 mM NaCl and indicated combinations of 3 μ M RanGTP and 1 μ M untagged wild-type SPN1 or the specified mutants. CRM1-ligands were eluted with $MgCl_2$ and analyzed as described in (A). Note that mutating either Leu⁴, Leu⁸, Phe¹² or Val¹⁴ to Ser (left panel) or deleting Met¹ (right panel) abolished or substantially impaired SPN1-binding to CRM1-RanGTP, while mutating Leu²⁸ did not (left panel). (C) The N terminus of SPN1 contains export determinants that allow autonomous, RanGTP-stimulated binding to CRM1. Indicated zz-tagged SPN1 derivatives or the PKI-NES immobilized on IgG Sepharose were incubated with an *E. coli* lysate containing 1 μ M CRM1 and 3 μ M RanGTP as specified. Bound ligands were eluted with $MgCl_2$ and analyzed as described in (A). At low NaCl concentration (50 mM, upper panel), SPN1¹²⁻³⁶⁰ bound CRM1-RanGTP nearly as efficiently as full-length SPN1¹⁻³⁶⁰, however, a clear decrease in binding was observed without Ran. SPN1¹⁻²¹ recruited CRM1 in a strictly RanGTP dependent manner. This CRM1-binding was lost when SPN1^{Met1} was deleted. Even though SPN1¹⁻²¹ contains 5 hydrophobic residues, it bound CRM1 considerably weaker than the classical PKI-NES with only 4 hydrophobic residues. This difference was particularly apparent at 200 mM NaCl (lower panel).

Leu⁸, Phe¹², and Val¹⁴. Strikingly, mutating any of those residues to serine or deleting Met¹ strongly impaired the interaction with CRM1, in particular at higher salt concentrations (Fig. 1B and fig. S1). Even though the SPN1 N terminus with its conserved hydrophobic residues resembles a classical NES, there are clear differences, foremost that CRM1 binds the isolated SPN1 N terminus (SPN1¹⁻²¹) considerably weaker than, e.g., the PKI-NES (Fig. 1C). In the context of full-length SPN1, this difference is, however, more than compensated by the contribution of the m₃G cap-binding domain to the CRM1 interaction. We then assembled, purified and crystallized an export complex containing full-length human SPN1¹⁻³⁶⁰, full-length mouse CRM1¹⁻¹⁰⁷¹ and GTP-RanQ69L¹⁻¹⁸⁰, a C-terminally truncated and GTPase-deficient form of human Ran (21). The resulting crystals contained two complexes per asymmetric unit. The structure was solved by molecular replacement using known structures of GTP-bound Ran⁷⁻¹⁷⁶ (22), SPN1⁹⁷⁻³⁰⁰ (17), and a short human CRM1⁷⁰⁷⁻¹⁰²⁷ fragment (23). The final model, refined at a resolution of 2.5 Å, comprises residues 12 to 1055 of CRM1, Ran⁹⁻¹⁷⁹, as well as SPN1¹⁻³⁶⁰. CRM1⁶⁷⁻⁶⁹ and four regions of SPN1 appear disordered (table S1) (21). As expected from previous sequence analysis (23, 24), CRM1 is built from so-called HEAT repeats (Fig. 2, fig. S2, and table S2), which comprise two consecutive helices (A and B) that pack in anti-parallel orientation against each other and against the adjacent repeat (25). However, previous structure prediction (23) failed to predict the correct number and exact positions of the 21 repeats. This reflects the highly degenerate nature of some of the repeats, which even leads to an inverted topology of helices at the C terminus of CRM1 (fig. S2). In contrast to importin β (26), transportin (27) and CAS/ Cselp (22, 28), the overall CRM1 structure shows remarkably little superhelical twist (Fig. 2). However, it is bent to a distorted toroid-structure, with HEAT 21 touching helices 2B and 5A, as well as the loop between HEATs 4 and 5 (Fig. 2 and fig. S2). Ran is enclosed into

this toroid and stabilizes the ring-closure by extensive contacts. In contrast to the IBB-importin β interaction (18, 19, 26), the cargo SPN1 is not enveloped by CRM1, but rests on the outside of the CRM1-toroid (Fig. 2). This different binding topology might reflect the fact that CRM1 carries cargoes, such as ribosomal subunits, that are anyway too large to be engulfed by an exportin.

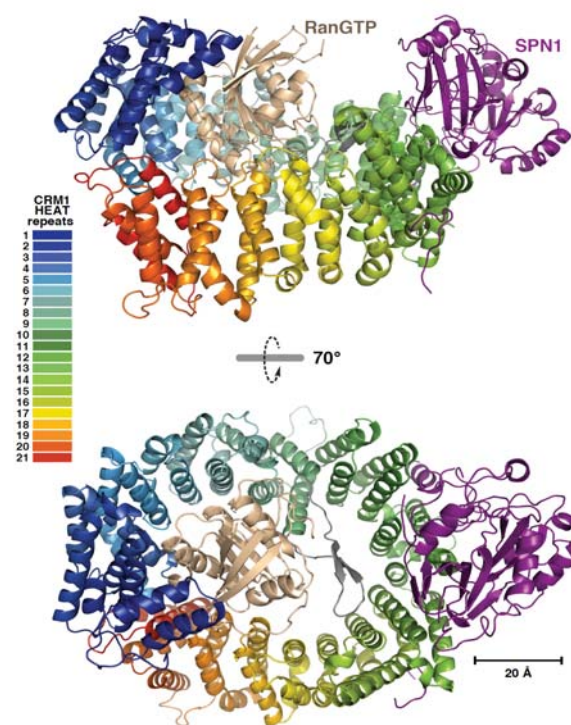


Fig. 2. Structure of the SPN1-CRM1-RanGTP nuclear export complex. Two views of the complex are depicted. Color-codes for Ran, SPN1, and the 21 consecutive HEAT repeats of CRM1 are indicated. Except for HEAT 21, A-helices of the HEAT repeats are located at the outer and B-helices at the inner surface of the CRM1-toroid (see also fig. S2). RanGTP is engulfed by the toroid-shaped structure of CRM1 and fixed by the so-called acidic loop (shown in the lower panel in gray), which is part of HEAT repeat 9. SPN1 is bound on the outer surface of CRM1, far away from the Ran molecule.

In addition, the outside of the torus provides a larger surface area and possibly also a greater variety of binding sites for cargo recognition than the inner face that already accommodates the Ran molecule. The structure of m₃G cap-bound SPN1⁹⁷⁻³⁰⁰ was previously solved (17) and remained essentially unaltered in the SPN1-CRM1-RanGTP complex (rmsd 0.67 Å).

However, several residues of SPN1 and as well as of CRM1 HEATs 12 and 13 protrude into the m₃G-cap binding pocket (fig. S3). With the physiological import cargo of SPN1 (fully assembled U snRNPs) the clashes would be even more severe, because the RNA would run into the CRM1 molecule. Thus, SPN1 cannot simultaneously bind its import cargo and its export receptor, which agrees with previous data (14). This ensures that only cargo-free SPN1 is returned to the cytoplasm and allows SPN1 to mediate uni-directional transport of m₃G-capped U snRNPs into nuclei. SPN1 binds CRM1 through an elaborate contact area (2330 Å²), which comprises three parts, the N terminus (SPN1¹⁻³⁵), the m₃G cap-binding domain (SPN1⁹⁷⁻³⁰⁰) and a C-terminal region, SPN1³⁴⁹⁻³⁶⁰ (Fig. 3A). This is consistent with biochemical data that revealed strong contributions of SPN1¹⁻²¹ and the cap-binding domain to CRM1-binding (Fig. 1, B and C, and fig. S1) (14) and a

weaker contribution of SPN1286-360 (14). All N-terminal residues that were found to be critical for CRM1-binding (SPN1^{Met1, Leu4, Leu8, Phe12, Val14}, Fig. 1B and fig. S1), dock into a hydrophobic cleft that is formed by helices 11A and 12A and the intervening helical linker between 11B and 12A of CRM1 (Fig. 3B and fig. S4). The side chain of CRM1^{Lys534}, which is positioned by a salt bridge to CRM1^{Glu575}, closes the cleft and introduces a sharp kink into the SPN1 chain between SPN1^{Val14} and SPN1^{Ser15}. There are several additional contacts in this area, such as electrostatic attraction between the negatively charged N-terminal helix of SPN1 and basic regions on the CRM1-surface, as well as hydrogen bonds between SPN1^{Ser15} and CRM1^{Glu575} and between SPN1^{Tyr35} and CRM1^{Glu529} (fig. S4). The CRM1-inhibitor leptomycin B (LMB) covalently modifies CRM1^{Cys528} (29).

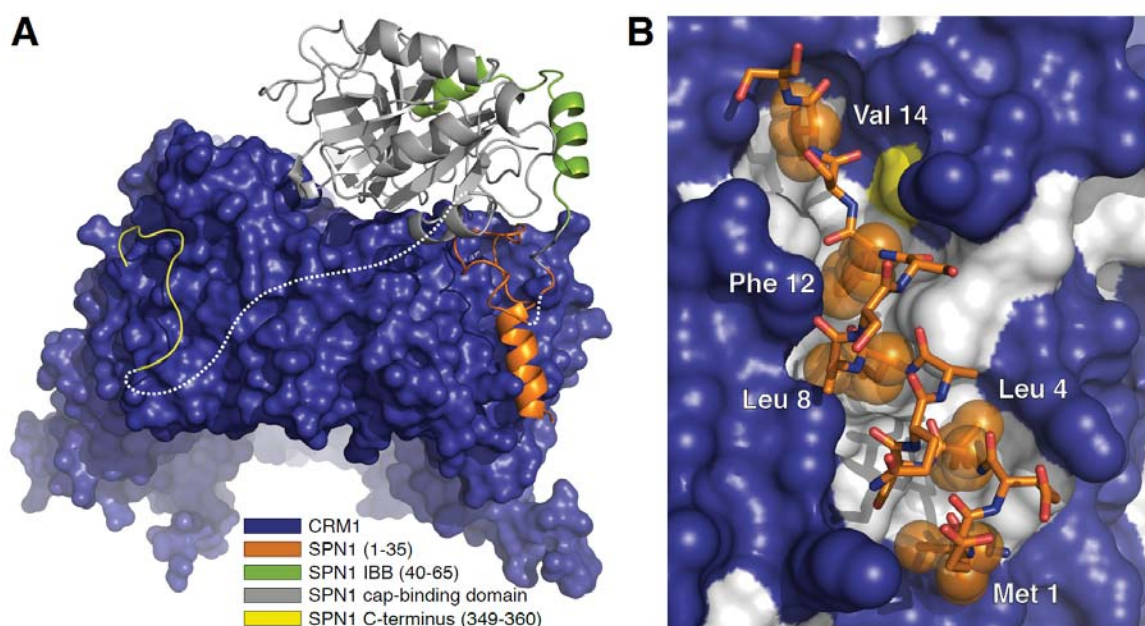


Fig. 3. The nuclear export signature of SPN1 involves a large interface formed by residues from all three domains of SPN1. **(A)** Three domains of SPN1 contact CRM1. These include N-terminal residues of SPN1 (shown in orange), the cap-binding domain (gray), and C-terminal residues (SPN1³⁴⁹⁻³⁶⁰, yellow). The IBB-domain of SPN1, which forms a straight helix within the importin β complex (18, 19), is here partially unwound and bent (depicted in green). White dashed lines mark unresolved stretches. **(B)** The N-terminal hydrophobic residues of the SPN1 (Met¹, Leu⁴, Leu⁸, Phe¹², Val¹⁴) dock into a hydrophobic cleft of CRM1. Carbons of SPN1 are shown as orange, oxygens as red and nitrogens as blue sticks. The side chains of the hydrophobic residues are depicted as spheres. CRM1 is shown as a surface representation; blue indicates hydrophilic, white denotes hydrophobic areas. The yellow patch marks the sulfur of Cys⁵²⁸, which is covalently modified by the CRM1-specific inhibitor leptomycin B (29).

Cys⁵²⁸ is located within the hydrophobic cleft (Fig. 3B), which explains plausibly why LMB-modified CRM1 cannot bind export cargoes that rely on this cleft. The N-terminal part of snurportin's export signature with its 5 critical hydrophobic residues resembles a classical NES and binds CRM1 in a conformation, where residues Met¹-Ser¹¹ form an α helix (Fig. 3A and fig. S4). The classical NES from the HIV Rev protein (15) must be recognized differently for three reasons: The spacing of the hydrophobic residues is different, the intervening prolines would not allow such a helix to form, and this classical NES contains only four critical hydrophobic residues (15).

Nevertheless, we cannot exclude the possibility that the same hydrophobic cleft also accommodates some or all of the key hydrophobic residues from classic NESs. The interaction between the SPN1 m₃G cap-binding domain and CRM1 is dominated by polar contacts. SPN1³⁴⁹⁻³⁶⁰, the third part of the export signature, binds to helices 14A, 15A, and 16A of CRM1 (Figs. 2 and 3A).

Importin β can displace CRM1 from the rather stable Ran-free SPN1-CRM1 complex (Fig. 1A) and thereby restore m₃G cap-binding of SPN1 in the cytoplasm (fig. S3) (14). This antagonism between CRM1 and importin β is not caused by an overlap of the respective binding sites, but apparently by a combination of conformational changes in SPN1 and volume extrusion. The IBB domain binds importin β as a straight helix (18, 19). However, within the CRM1-complex, the central part of this IBB-helix is unwound and the remaining helix-fragments are kinked by $\approx 80^\circ$ (shown in green in Fig. 3A). This distortion of the IBB-helix appears enforced by contacts of the 35 N-terminal residues of SPN1 with CRM1. Thus, straightening of the IBB-helix by importin β is likely to break crucial contacts between CRM1 and SPN1. The structure of Ran in the SPN1-CRM1-RanGTP complex is virtually identical to other transport receptor-RanGTP complexes (22, 30, 31). Ran is almost completely engulfed by the CRM1-toroid and contacts four distinct areas of CRM1 (Figs.

2 and 4, A to C, and movie S1). The first area is located within the region that is most conserved between nuclear transport receptors (24, 32). HEATs 1-3 bind switch II of Ran, while HEATs 4 and 5 pack against Ran helix 3 and the so-called "basic patch" (30), respectively (Fig. 4, B and C). The second Ran-binding region (HEATs 7-9) also contacts the basic patch and extends to β strand 6 of Ran. Analogous interactions occur in RanGTP-complexes with CAS, transportin and importin β (22, 30, 31). In contrast, the long "acidic loop" within HEAT 9 (region 3) binds Ran in an unprecedented manner. It forms a β hairpin, touches HEAT helices B12-15 and reaches through the entire central "hole" of CRM1-toroid (Figs. 2 and 4, A and B, and figs. S2 and S5). It locks RanGTP closely to the N- and C-terminal HEAT repeats and binds Ran37 from switch I as well as Ran^{127,129,155} from the loops involved in guanine recognition. The fourth, C-terminal Ran-binding region (HEATs 17 and 19) was not anticipated by sequence similarity or previous structures. It contacts both switch regions of Ran. In order to function as an effective, uni-directional cargo-pump, CRM1 must strongly discriminate between GTP- and GDP-bound Ran. CRM1 can sense the nucleotide state of Ran, because it contacts switches I and II, i.e., the regions that differ most between GDP- and GTP-Ran (Fig. 4, B and C, and fig. S5). Indeed, the structure of RanGDP (33, 34) is incompatible with CRM1-binding, because GDP-Ran40-42 (within switch I) would clash with CRM1³¹⁻³⁵, while GDP-Ran⁷¹⁻⁷² (of switch II) would collide with CRM1⁹³³⁻⁹³⁴. Ran switches the affinity of importin β -type transport receptors for their cargoes and thereby provides energy for the transport cycles. In the case of Cse1p, RanGTP increases the affinity of the exportin for its cargo importin α by directly interacting with both, Cse1p and importin α (22). There are, however, no direct contacts between Ran and cargo in the SPN1-CRM1-RanGTP complex (Fig. 2). The ≈ 1000 -fold increase in the affinity of CRM1 for RanGTP by SPN1 and the equally large strengthening of the CRM1-SPN1 interaction by RanGTP (14) must

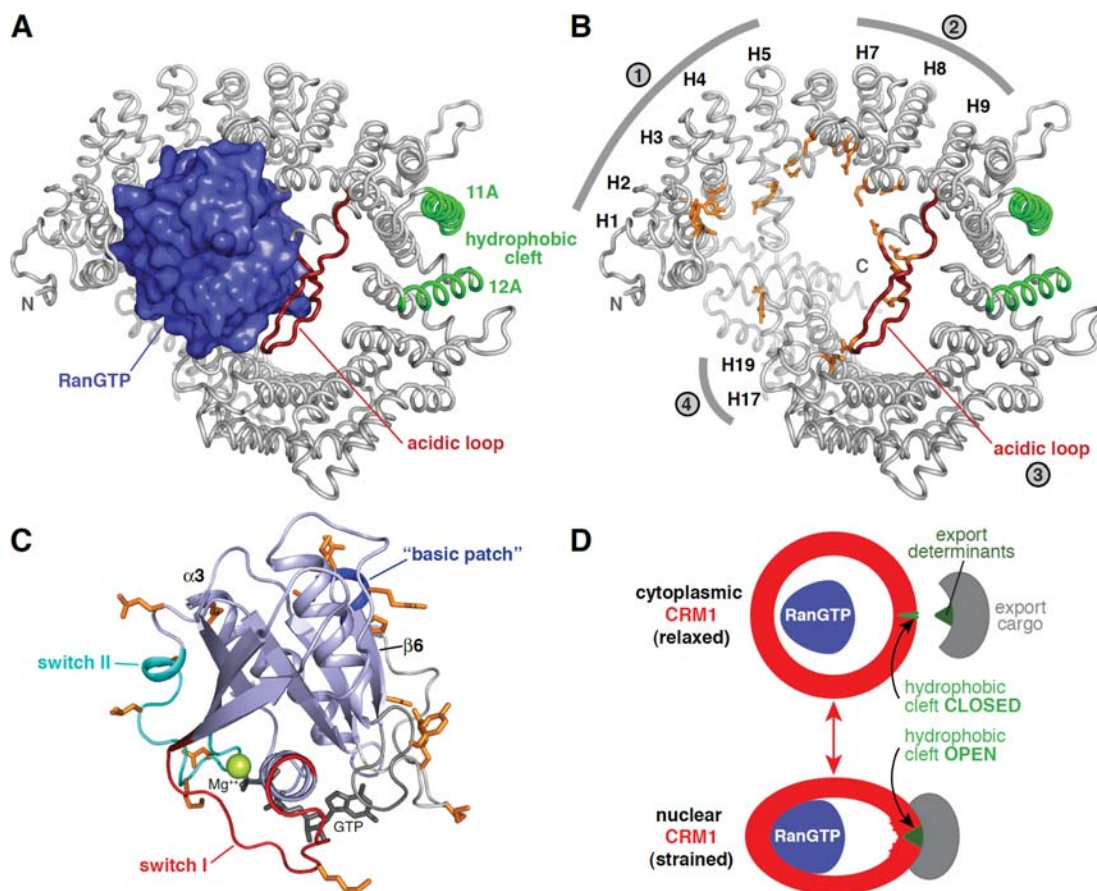


Fig. 4. Molecular details of the CRM1·RanGTP interaction. **(A)** CRM1 is shown as a gray backbone tube, Ran as surface representation. HEAT helices 11A and 12A, forming the cargo-binding hydrophobic cleft, are shown in green, the acidic loop in red. **(B)** RanGTP contact areas on CRM1. Orientation of CRM1 is as in (A) but Ran has been removed and Ran-binding residues of CRM1 (distance <3.6 Å) are shown as orange sticks. Note that the Ran-binding site comprises four distinct areas (labeled 1-4). See also [movie S1](#). **(C)** Contacts of RanGTP to CRM1. Ran is depicted as a ribbon diagram. Orientation is as in (A). CRM1-binding residues are shown in orange, switch I (Ran³⁰⁻⁴⁷) in red, switch II (Ran⁶⁵⁻⁸⁰) in cyan. CRM1 contacts both switches. The "basic patch" (Ran¹³⁹⁻¹⁴², dark blue) shows extensive contacts to CRM1 regions 1 and 2 (see panel B). Secondary structure elements are numbered as in (33). GTP is depicted as gray sticks, the Mg^{2+} ion as a green sphere. **(D)** Model for conformational states of CRM1. CRM1 switches between a relaxed cytoplasmic (top) and a strained nuclear conformation (bottom). In the hypothetical cytoplasmic conformation, the contact sites for RanGTP inside the CRM1-toroid are too far apart to bind Ran with high affinity. Also, the hydrophobic cleft on the outer side of the toroid is closed. Rigid body movements allow transition to the nuclear conformation. Here, the Ran-binding sites are close enough to bind Ran simultaneously and thus with high affinity. The conformational change also alters the curvature of the toroid near the cargo-binding site, opens the hydrophobic cleft and allows the export cargo to dock. For details, see main text.

therefore be caused solely by conformational changes in the CRM1 molecule. Both RanGTP and SPN1 obviously select the same conformation of CRM1 for high-affinity binding (here referred to as the nuclear conformation), while the free cytoplasmic conformation of CRM1 has only a low affinity for the two ligands (Fig. 1 and fig. S1). To explain this cooperativity, one must assume that the nuclear conformation is under considerable tension and that this tension is compensated by the released

binding energies of the Ran-CRM1 and SPN1-CRM1 interactions. In other words, Ran promotes SPN1-binding apparently by stabilizing the strained nuclear conformation of CRM1, and vice versa. Ran and SPN1 are ≈ 55 Å apart in the export complex (see Fig. 2). The conformational changes in CRM1 that coordinate their binding must therefore be transmitted over a considerable distance, probably through rigid body movements along the HEAT repeats. The splitting of the Ran-

binding site on CRM1 into distinct regions is probably crucial for driving these movements, since even small changes in their distances will greatly affect the binding of Ran. Ran-binding regions 2 and 4, for example, are located on opposite sides of the CRM1-toroid (Fig. 4B), and it is well possible that they are too far apart in the relaxed CRM1 conformation to contact Ran simultaneously (Fig. 4D). A low affinity for Ran would result. The transition to the nuclear conformation would bring these interfaces closer together and allow high-affinity binding of Ran. Conformational changes in CRM1, which favor the CRM1-Ran interaction, must also activate the cargo binding site(s). We therefore suggest that the hydrophobic cleft is also controlled by these transitions. The cleft might be closed in the cytoplasmic, relaxed conformation of CRM1 (Fig. 4D). Putting the CRM1-toroid under tension to bind Ran with all interfaces should also change the curvature of the CRM1-molecule around the cargo-binding site. This might stretch the contacts between the outer α -helices of HEATs 11 and 12 and thereby open the hydrophobic cleft. The observation that CRM1-binding of the isolated SPN1 N terminus and thus docking into the hydrophobic cleft is efficient only in the presence of RanGTP (Fig. 1C) strongly supports this model.

References and Notes

1. D. Görlich, U. Kutay, *Annu. Rev. Cell Dev. Biol.* **15**, 607 (1999).
2. A. Cook, F. Bono, M. Jinek, E. Conti, *Annu. Rev. Biochem.* **76**, 647 (2007).
3. D. Görlich, N. Pante, U. Kutay, U. Aebi, F. R. Bischoff, *EMBO J.* **15**, 5584 (1996).
4. M. Fornerod, M. Ohno, M. Yoshida, I. W. Mattaj, *Cell* **90**, 1051 (1997).
5. K. Stade, C. S. Ford, C. Guthrie, K. Weis, *Cell* **90**, 1041 (1997).
6. U. Kutay, F. R. Bischoff, S. Kostka, R. Kraft, D. Görlich, *Cell* **90**, 1061 (1997).
7. U. Kutay, S. Güttinger, *Trends Cell Biol.* **15**, 121 (2005).
8. O. Gadal *et al.*, *Mol. Cell Biol.* **21**, 3405 (2001).
9. T. I. Moy, P. A. Silver, *J. Cell Sci.* **115**, 2985 (2002).
10. A. W. Johnson, E. Lund, J. Dahlberg, *Trends Biochem. Sci.* **27**, 580 (2002).
11. M. T. Bohnsack *et al.*, *EMBO J.* **21**, 6205 (2002).
12. M. Ohno, A. Segref, A. Bachi, M. Wilm, I. W. Mattaj, *Cell* **101**, 187 (2000).
13. J. Huber *et al.*, *EMBO J.* **17**, 4114 (1998).
14. E. Paraskeva *et al.*, *J. Cell Biol.* **145**, 255 (1999).
15. U. Fischer, J. Huber, W. C. Boelens, I. W. Mattaj, R. Lührmann, *Cell* **82**, 475 (1995).
16. W. Wen, J. Meinkoth, R. Tsien, S. Taylor, *Cell* **82**, 463 (1995).
17. A. Strasser, A. Dickmanns, R. Lührmann, R. Ficner, *EMBO J.* **24**, 2235 (2005).
18. D. Wohlwend, A. Strasser, A. Dickmanns, R. Ficner, *J. Mol. Biol.* **374**, 1129 (2007).
19. G. Mitrousis, A. S. Olia, N. Walker-Kopp, G. Cingolani, *J. Biol. Chem.* **283**, 7877 (2008).
20. D. Görlich, P. Henklein, R. Laskey, E. Hartmann, *EMBO J.* **15**, 1810 (1996).
21. Materials and methods are available as supporting material on Science Online.
22. Y. Matsuura, M. Stewart, *Nature* **432**, 872 (2004).
23. C. Petosa *et al.*, *Mol. Cell* **16**, 761 (2004).
24. D. Görlich *et al.*, *J. Cell Biol.* **138**, 65 (1997).
25. M. A. Andrade, C. Petosa, S. I. O'Donoghue, C. W. Müller, P. Bork, *J. Mol. Biol.* **309**, 1 (2001).
26. G. Cingolani, C. Petosa, K. Weis, C. W. Müller, *Nature* **399**, 221 (1999).
27. B. J. Lee *et al.*, *Cell* **126**, 543 (2006).
28. A. Cook *et al.*, *Mol. Cell* **18**, 355 (2005).
29. N. Kudo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9112 (1999).
30. I. R. Vetter, A. Arndt, U. Kutay, D. Görlich, A. Wittinghofer, *Cell* **97**, 635 (1999).
31. Y. M. Chook, G. Blobel, *Nature* **399**, 230 (1999).
32. M. Fornerod *et al.*, *EMBO J.* **16**, 807 (1997).
33. K. Scheffzek, C. Klebe, K. Fritz-Wolf, W. Kabsch, A. Wittinghofer, *Nature* **374**, 378 (1995).
34. M. Stewart, H. M. Kent, A. J. McCoy, *J. Mol. Biol.* **284**, 1517 (1998).
35. We thank the staff of synchrotron beamlines at BESSY (Berlin), EMBL/DESY (Hamburg), and ESRF (Grenoble) for assistance during data collection; D. Deichsel for excellent technical help; and S. Frey and S. Güttler for comments on the manuscript. Funding was from the DFG (SFB523, to R.F.), the MPG (to D.G. and T.G.), the Alfred Krupp-Foundation (to D.G. and T.G.), and the Boehringer Ingelheim Fonds (to T.G.). Coordinates and structure factors have been deposited in the Protein Data Bank with accession code 3GJX.

6.1 Supporting online material

MATERIALS AND METHODS

Protein expression and purification

Full-length mouse CRM1¹⁻¹⁰⁷¹ was expressed at 16 °C in *E. coli* BLR as an N-terminal His-*zz*- [TEV] fusion protein from a pQE80-derived plasmid (QIAGEN, Hilden, Germany). Cells were lysed in 50 mM Tris (pH 7.5), 500 mM NaCl, 2 mM Mg(OAc)₂, 2 mM imidazole, 5 mM DTT. The protein was purified via Ni²⁺-chelate affinity chromatography with elution in lysis buffer containing 200 mM imidazole. After cleavage of the His-*zz*-tag during dialysis to the lysis buffer, tag and the bulk of contaminants were removed via another Ni²⁺-chelate column. The flow-through was further purified by gel filtration on a Superdex 200 column (equilibrated in 50 mM Tris (pH 7.5), 50 mM NaCl, 2 mM Mg(OAc)₂, 5 mM DTT). We truncated the C-terminus of Ran (residues 181-216), because it is disordered in other transport receptor complexes (1, 2), it destabilizes the GTP-bound form of Ran and weakens the interactions with transport receptors (3). The Q69L mutation blocks the GTPase activity (4). Human RanQ69L¹⁻¹⁸⁰ was expressed as an N-terminal His-*zz*-[TEV] fusion at 20°C in *E. coli* BL21 (DE3). Cells were lysed in 50 mM K-Phosphate (pH 7.0), 500 mM NaCl, 5 mM Mg(OAc)₂, 2 mM imidazole, 2 mM DTT. The purification of Ran was analogous to that of CRM1 with the difference that all steps performed after cell lysis were in the presence of 30 μM GTP. The nucleotide state of Ran was validated by extracting the nucleotide in deionized urea, followed by analysis on MonoQ (Amersham Biosciences). Expression and purification of human SPN1 for crystallization was as described (5). Wild-type and mutant SPN1 constructs for binding assays were expressed as N-terminal His-[TEV] or His-*zz*-[TEV] fusions in *E. coli* BLR at 20-25 °C and purified via Ni²⁺-chelate affinity chromatography (lysis buffer: 50 mM Tris (pH 7.5), 200 mM NaCl, 2 mM Mg(OAc)₂, 2 mM imidazole, 2 mM DTT). Where indicated, the His-tag had been cleaved off by TEVprotease and removed as described above. Untagged and His-*zz*-tagged SPN1^{1-360/2-360} were further purified on a Superdex 200 gel filtration column equilibrated to 50 mM Tris (pH 7.5), 200 mM NaCl, 2 mM Mg(OAc)₂, 2 mM DTT. His-*zz*-[TEV]-Hs SPN1^{1-21/2-21} were expressed in *E. coli* BLR at 37 °C, purified under denaturing conditions via the His-tag (lysis in 50 mM Tris (pH 8.0), 6 M Guanidinium hydrochloride, 1 mM EDTA, 2 mM DTT; elution in 50 mM Tris (pH 7.5), 8 M urea, 50 mM NaCl, 1 mM EDTA, 200 mM imidazole, 2 mM DTT) and refolded by dialysis to 50 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, 2 mM DTT. Expression and purification of human importin β was as described (6).

Binding assays

Binding assays were performed by incubating the specified components (see Fig. 1 and fig. S1) for 3 h with 20 μ l IgG-Sepharose 6 FastFlow (Amersham Biosciences) at 50 mM Tris (pH 7.5), 200 mM NaCl (if not denoted differently), 2 mM Mg(OAc)₂, 1 mM DTT, 0.005% (w/v) digitonin in the presence of an ATP/GTP-regenerating system (6). Where indicated, 3 μ M RanQ69L¹⁻¹⁸⁰ (GTP-form) was added. The binding volume was 500 μ l. Beads were washed 3x with 500 μ l of the respective binding buffer; bound material was eluted with 50 mM Tris (pH 7.5), 1.5 M MgCl₂ and precipitated with isopropanol (95% v/v final). Baits immobilized on IgG-Sepharose were then eluted with SDS.

Reconstitution and structure determination of the SPN1·CRM1·RanGTP complex

The complex was prepared by mixing purified CRM1¹⁻¹⁰⁷¹, GTP-RanQ69L¹⁻¹⁸⁰ and SPN1¹⁻³⁶⁰ and further purified by gel filtration on a Superdex 200 column (equilibrated in 50 mM Tris (pH 7.5), 50 mM NaCl, 2 mM Mg(OAc)₂, 5 mM DTT). For crystallization, the protein solution was concentrated to 4 mg·ml⁻¹ and stored on ice for a maximum of 2 weeks. The complex was crystallized by the vapor diffusion method in sitting drop 24-well ChrysChem-plates (Hampton Research, CA, USA). 1 μ l of a reservoir solution containing 11% (w/v) PEG1000 and 100 mM Tris pH 8.05 was mixed with 1 μ l of the prepared protein complex solution. Single crystals with dimensions of 50 μ m \times 50 μ m \times 300 μ m grew within 4 days at 293 K and belonged to the space group *P*2₁ with cell dimensions of *a* = 72.17 Å, *b* = 225.72 Å and *c* = 163.41 Å and angles of α = 90.0°, β = 100.6° and γ = 90.0°. The crystals were soaked in 15.7 % (v/v) propanediol for 10 seconds and flash-frozen.

For structure determination more than 500 crystals had to be tested. The X-ray diffraction data of two crystals, diffracting to a maximum resolution of 2.5 Å, were integrated, scaled, reduced and merged using XDS (7). The structure was solved by means of molecular replacement using PHASER (8) with the crystal structures of GTP-Ran⁷⁻¹⁷⁶ (PDB-ID 1WA5) (9), the m₃G-cap-binding domain of SPN1 (residues 97-300; PDB-ID 1XK5) (10) and CRM1⁷⁰⁷⁻¹⁰²⁷ (PDB ID 1W9C) (11) as search models. The resulting electron density map was used to complete the initial search model. The structure was improved by iterative cycles of refinement using CNS (12) and PHENIX (13), as well as model building in COOT (14). Waters were built manually in COOT. In the final model of SPN1, residues 31-33, 72-91, 162-165 and 288-348 of one complex in the asymmetric unit, and residues 30-32, 74-92, 162-165 and 291-348 of the other one were not defined in the electron density map and thus they were not built. The electron density corresponding to the C-terminal residues of SPN1 could not be interpreted unambiguously. The model containing residues 349-360 yielded the best R-factor and R_{free}-value, and strong electron density peaks correlated with the positions of the sulfur atoms of Cys³⁵⁶ and Met³⁵⁸. Ran comprises the residues 9-179, the bound GTP molecule as well as a coordinated magnesium ion. The polypeptide chain of CRM1 could be traced for residues 12-1055 with exception of the flexible region encompassing residues 67-69. The structure was refined at a resolution of 2.5 Å

to an R_{work} of 24.4 % and an R_{free} -value of 28.1 % (see [Table S1](#)). In the final model, 88 % of the residues are located within the most favored regions of the Ramachandran plot, 11 % in the additionally allowed ones, 1% in the generously allowed regions and none in the disallowed regions. Contact surfaces were calculated with the program AREAIMOL as implemented in the CCP4 suite (version 6.0.2) ([15](#)). Figures were prepared with PyMOL (DeLano, W.L. The PyMOL Molecular Graphics System (2002), DeLanoScientific, USA).

References and Notes to Materials and Methods

1. I. R. Vetter, A. Arndt, U. Kutay, D. Görlich, A. Wittinghofer, *Cell* **97**, 635 (1999).
2. A. Cook, F. Bono, M. Jinek, E. Conti, *Annu Rev Biochem* **76**, 647 (2007).
3. S. A. Richards, K. M. Lounsbury, I. G. Macara, *J Biol Chem* **270**, 14405 (1995).
4. C. Klebe, F. Bischoff, H. Ponstingl, A. Wittinghofer, *Biochemistry* **34**, 639 (1995).
5. A. Strasser *et al.*, *Acta Crystallogr D Biol Crystallogr* **60**, 1628 (2004).
6. S. Jäkel, D. Görlich, *EMBO J* **17**, 4491 (1998).
7. W. Kabsch, *J Appl Crystallogr* **26**, 795 (1993).
8. A. J. McCoy, *Acta Crystallogr D Biol Crystallogr* **63**, 32 (2007).
9. Y. Matsuura, M. Stewart, *Nature* **432**, 872 (2004).
10. A. Strasser, A. Dickmanns, R. Luhrmann, R. Ficner, *EMBO J* **24**, 2235 (2005).
11. C. Petosa *et al.*, *Mol Cell* **16**, 761 (2004).
12. A. T. Brunger, *Nat Protoc* **2**, 2728 (2007).
13. P. D. Adams *et al.*, *Acta Crystallogr D Biol Crystallogr* **58**, 1948 (2002).
14. P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (2004).
15. Collaborative Computational Project, No. 4, *Acta Crystallogr D Biol Crystallogr* **50**, 760 (1994).

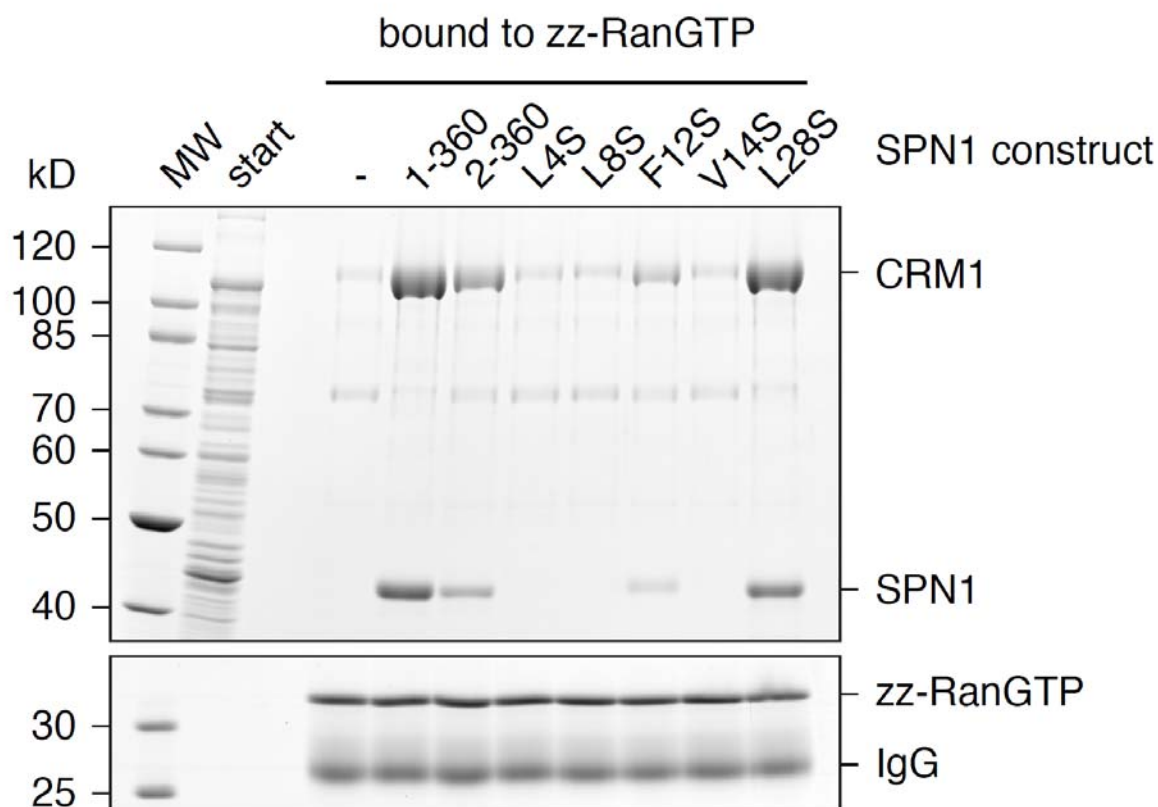


Fig. S1. Met¹, Leu⁴, Leu⁸, Phe¹² and Val¹⁴ of SPN1 are critical for cooperative export complex formation. The experimental setup was essentially identical to that described in Fig. 1B with the difference that here, RanGTP instead of CRM1 was zz-tagged and immobilized on IgG-Sepharose. The beads were incubated at 200 mM NaCl with an *E. coli* extract that contained 1 μ M CRM1 and 1 μ M of wild-type or mutant SPN1 as indicated. RanGTP-bound ligands were eluted with MgCl₂ and analyzed by SDS-PAGE and Coomassie-staining. As the affinity of CRM1 for RanGTP is low in the absence of export cargo, RanGTP recruited CRM1 only very inefficiently without SPN1. However, addition of wild-type SPN1 (1-360) promoted CRM1-binding to RanGTP. Note that the deletion of Met¹ of SPN1 or the change of either Leu⁴, Leu⁸, Phe¹² or Val¹⁴ to Ser abolished or strongly weakened the CRM1·RanGTP interaction. Mutating Leu²⁸ to Ser had no effect. See also Fig. 1B.

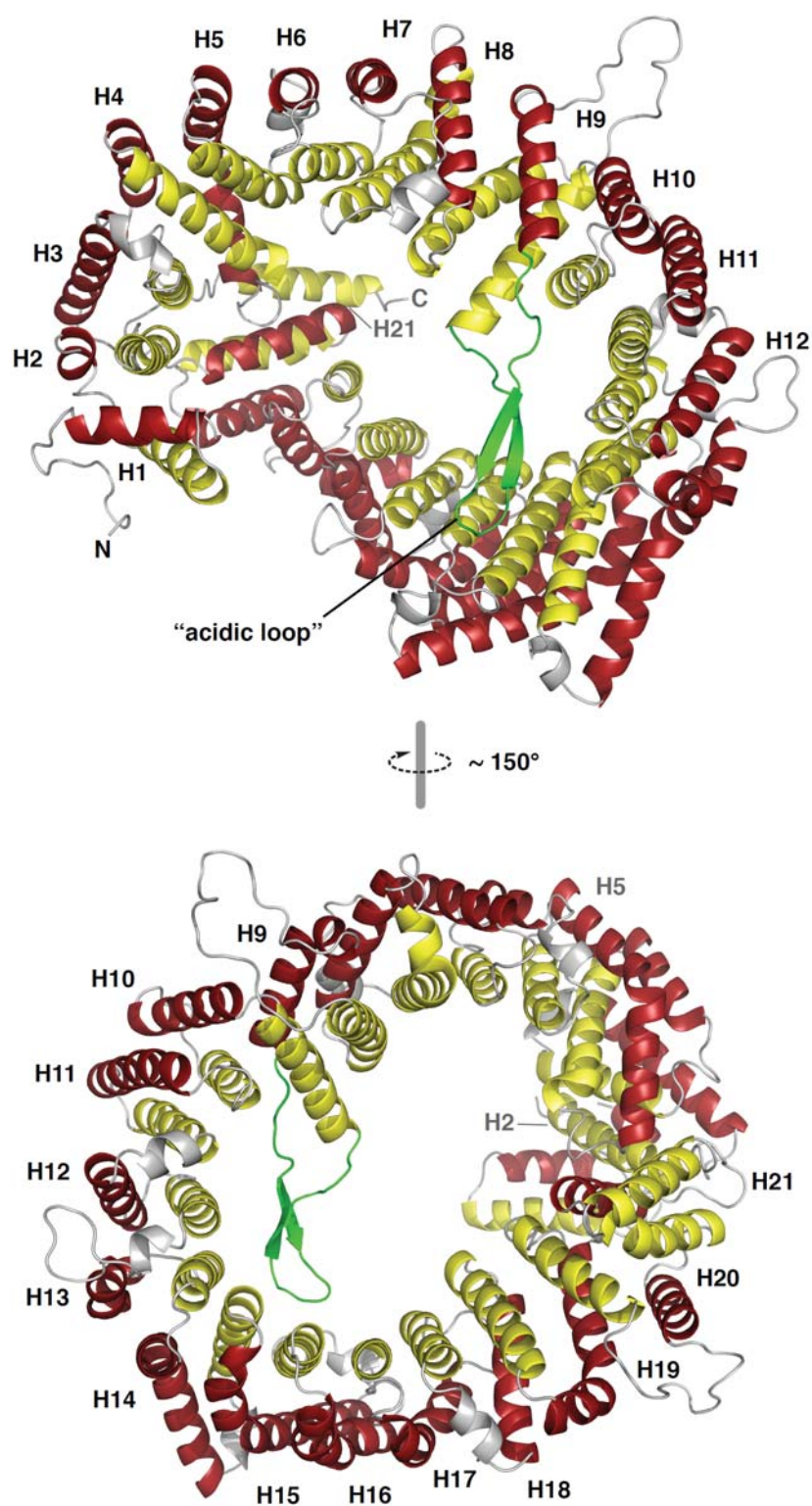


Fig. S2. HEAT repeat organization of CRM1. The panels illustrate the organization of CRM1 from consecutive HEAT repeats, numbered H1-H21. Two views are depicted. A-helices (colored in red) are located on the outside of the torus, while B-helices (yellow) face the inside. However, this topology is inverted for HEAT 21. HEAT 21 also circularizes CRM1 by contacting HEATs 2 and 5. The long intra-repeat "acidic loop" within HEAT 9 is shown in green.

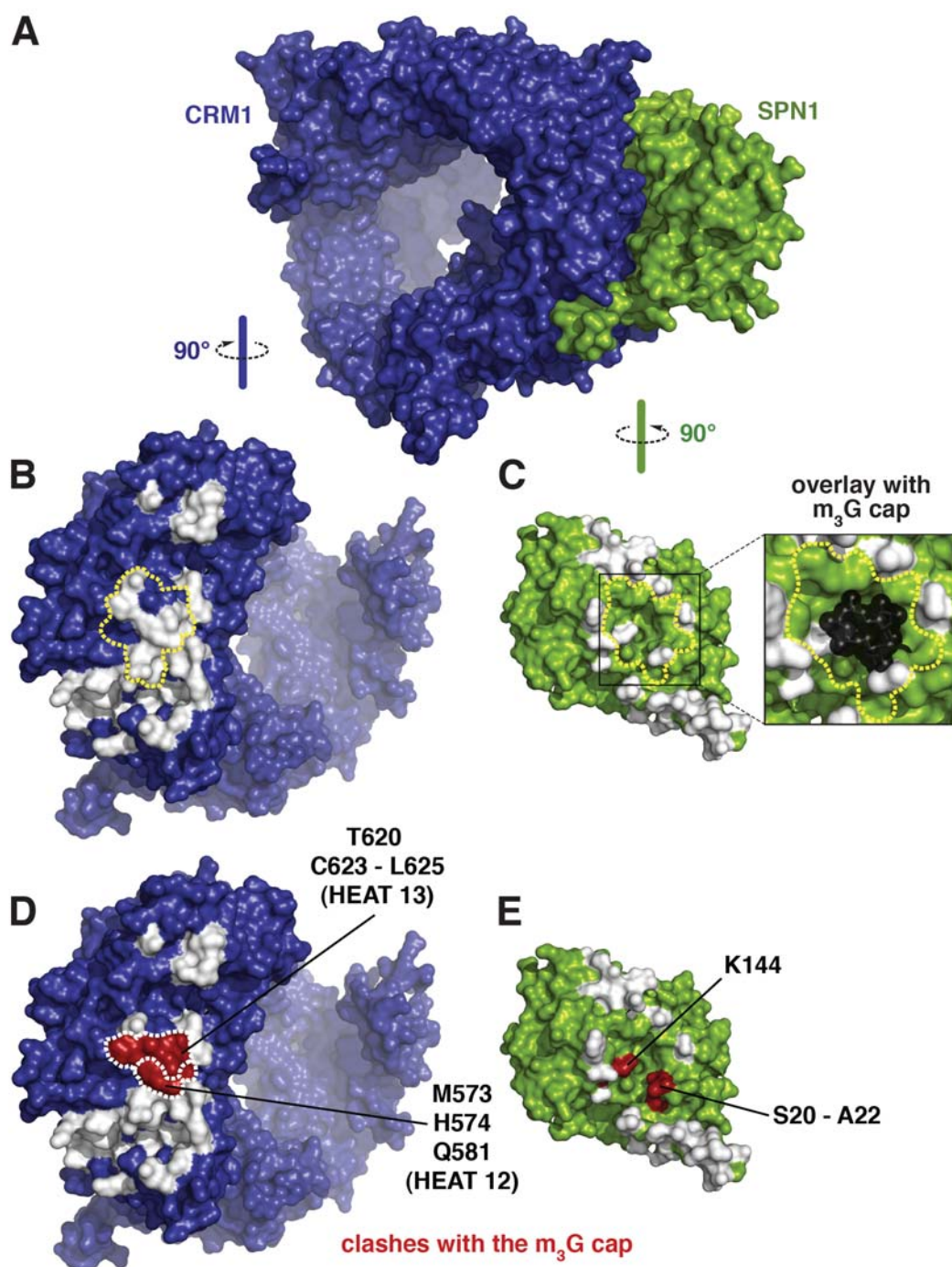


Fig. S3. m₃G cap binding is excluded in the export complex. **(A)** The complex of CRM1 (blue) and SPN1 (green) is shown as a surface representation. In the following panels, it is flipped open to illustrate the contact sites (shown in white). Deviation from exact complementarity is due to coloring interacting residues and not interacting atoms. **(B)** The CRM1 molecule from (A) is shown alone and rotated as indicated. Areas that contact SPN1 are colored in white. The region that covers the m₃G cap-binding site of SPN1 is demarcated by a yellow dashed line. **(C)** The SPN1 molecule from (A) is shown alone and rotated as indicated. Residues contacting CRM1 are shown in white. The m₃G cap-binding site is bordered by a yellow dashed line. The insert shows superposition with an m₃G cap (in black) from the SPN1·m₃GpppG complex (PDB accession code 1XK5). **(D)** The CRM1 molecule is shown as in (B), but those residues are shown in red that would clash with an m₃G cap bound to SPN1. **(E)** The SPN1 molecule is shown as in (C), but residues that would clash with a bound m₃G cap are shown in red.

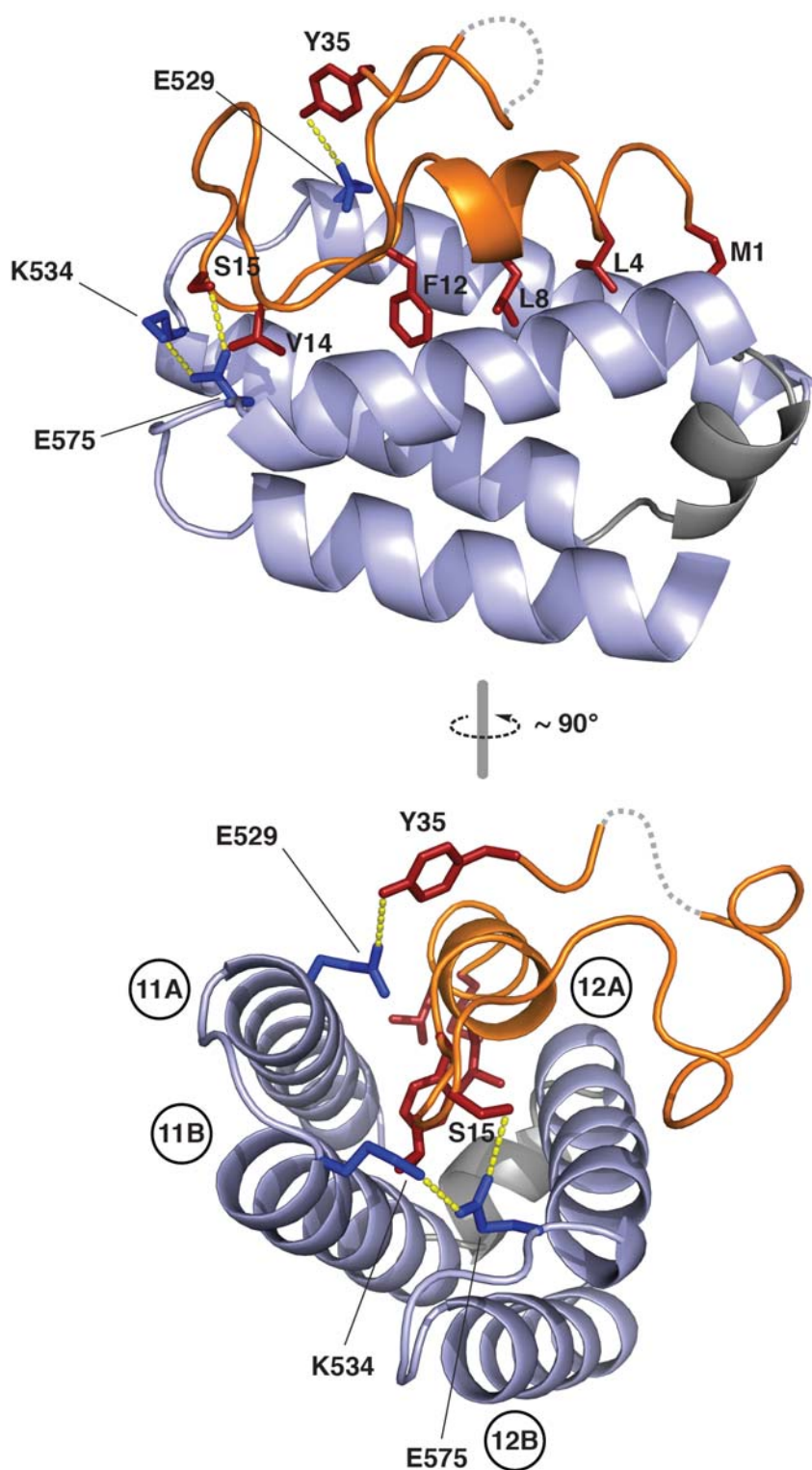


Fig. S4. Molecular details of the CRM1·SPN1¹⁻³⁵ interaction. HEAT repeats 11-12 (CRM1⁵¹⁰⁻⁵⁹⁵) are shown in blue, SPN1¹⁻³⁵ in orange. SPN1 residues Met¹, Leu⁴, Leu⁸, Phe¹² and Val¹⁴ (shown as red sticks) wedge into the hydrophobic cleft that is formed by CRM1 helices 11A and 12A and the helical linker (gray) connecting helices 11B and 12A. Hydrogen bonds (CRM1^{E529}-SPN1^{Y35}; CRM1^{E575}-SPN1^{S15}) and the salt bridge (CRM1^{K534}-CRM1^{E575}) are illustrated as yellow dashed lines. CRM1 residues engaged in polar contacts are shown as blue sticks. SPN1³⁰⁻³² remained unresolved and are shown as a gray dashed line. See main text for further details.

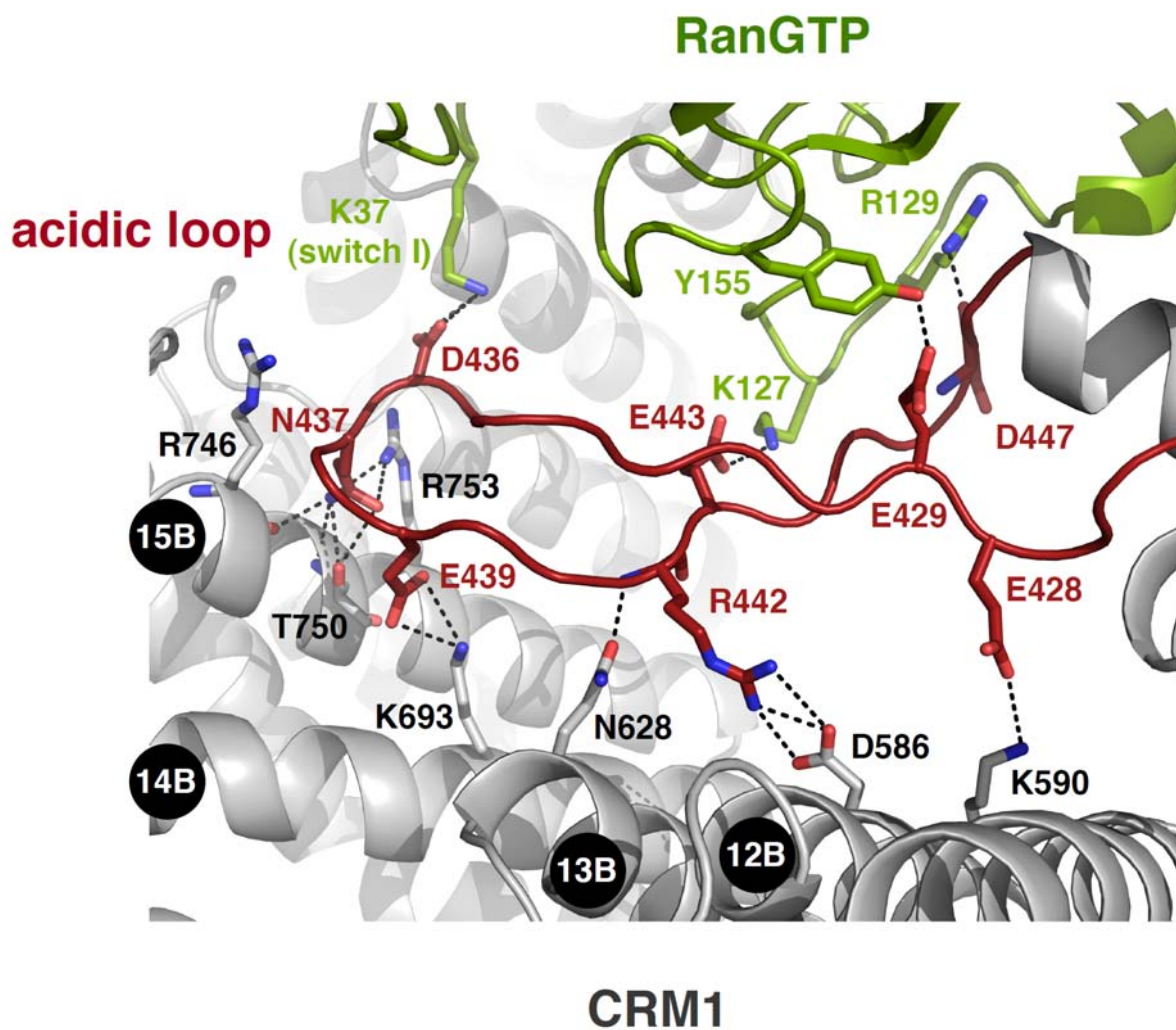


Fig. S5. Detailed view on the interactions of the acidic loop with Ran and HEAT repeats of CRM1. The acidic loop is colored in dark red, other parts of CRM1 in gray, and Ran in green. Atoms involved in polar interactions (black dashed lines) are colored in blue (nitrogen) or in light red (oxygen). Numbers in black circles denote those B-helices of CRM1 HEATs that contact the acidic loop with a distance of less than 3.6 Å. The extensive contacts immediately suggest that the acidic loop helps to couple Ran binding to those conformational changes that activate the cargo-binding site at HEATs 11 and 12. See also Fig. 4.

Table S1: Crystallographic data statistics

Crystal	SPN1·CRM1·RanGTP
Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	72.17, 225.72, 163.41
α , β , γ (°)	90.00, 100.56, 90.00
Wavelength (Å)	0.9
X-ray source	BL14.1, BESSY (Berlin)
Resolution range (Å)	38.84-2.50 (2.60-2.50)
No. of reflections	749341 (64346)
Completeness (%)	97.3 (96.1)
R_{merge}^a (%)	10.8 (44.3)
Average I/ σ	10.8 (2.9)
Redundancy	4.4 (3.4)
Refinement	
Resolution (Å)	38.8-2.5
Complexes per AU	2
No. of atoms	
Protein	24247
Ligand	66
Waters	959
R_{work}^b (%)	24.4 (28.1)
R_{free}^c (%)	28.1 (32.9)
Figure of merit	0.80
Average B factors (Å ²)	
Protein	48.9
Ligand	29
Waters	40
RMS deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.079

Values in parentheses indicate the specific values in the particular highest resolution shell.

^(a) $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I_i(hkl) \rangle| / \sum_{hkl} \sum_i \langle I_i(hkl) \rangle$, where the sum i is over all separate measurements of the unique reflection h, k, l .

^(b) $R_{\text{work}} = \sum_{hkl} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{hkl} |F_{\text{obs}}|$.

^(c) R_{free} as R_{work} , but summed over a 5 % test set of reflections.

Table S2: HEAT repeat helices in CRM1. See [Figure S2](#) for further details.

HEAT repeat	Helix A residues	Helix B residues
1	25 – 35	38 – 52
2	59 – 64	69 – 90
3	96 – 115	124 – 140
4	148 – 159	161 – 181
5	188 – 215	219– 234
6	246 – 254	261 – 273
7	280 – 297	314 – 339
8	344 – 359	363 – 383
9	405 – 423	
Acidic loop	424 – 448	
9		449 – 467
10	469 – 485	491 – 504
11	510 – 530	534 – 550
12	559 – 574	580 – 595
13	610 – 623	627 – 643
14	647 – 674	682 – 702
15	704 – 735	746 – 765
16	769 – 790	798 – 811
17	815 – 834	842 – 858
18	868 – 883	887 – 906
19	908 – 931	939 – 954
20	970 – 985	991 – 1005
21	1008 – 1023	1037 – 1052

Table S3: List of CRM1 residues that contact the RanGTP molecule with a distance of less than 3.6 Å. Corresponding HEAT repeats are numbered. "A": A helix; "B": B helix; "A-B": Loop between A and B helix; "AL": acidic loop within HEAT 9. Contact areas of CRM1 for Ran are numbered as in [Figure 4](#).

Residue of CRM1	CRM1 HEAT	Contact area of CRM1	Distance [Å]	Contacted residue of Ran
Tyr36	1A-1B	1	3.28	Gln82
Tyr77	2B	1	3.54	Asp77
Gln81	2B	1	2.66	Leu75
Gln81	2B	1	2.56	Asp77
Gln81	2B	1	3.27	Gly78
Lys129	3B	1	3.51	Asp77
Glu176	4B	1	3.54	Arg110
Glu176	4B	1	2.58	Arg110
Glu177	4B	1	3.08	Arg110
Glu177	4B	1	2.80	Arg110
Arg231	5B	1	2.99	Lys142
Asp313	7A-7B	2	3.16	Lys167
Asn317	7B	2	3.20	Asn143
Gln320	7B	2	3.59	Arg140
Asn321	7B	2	3.54	Asn143
Glu364	8B	2	2.78	His139
Glu364	8B	2	3.41	Arg140
Glu371	8B	2	3.58	Arg140
Glu429	AL	3	2.53	Tyr155
Asp436	AL	3	3.35	Lys37
Asp436	AL	3	3.26	Lys37
Glu443	AL	3	3.19	Lys127
Asp447	AL	3	2.47	Arg129
Asp449	9B	4	2.87	Asp148
Glu843	17B	4	3.38	Lys37
Asp932	19A-19B	4	2.88	Lys71
Thr933	19A-19B	4	2.62	Glu70
Thr933	19A-19B	4	2.78	Lys71
Thr933	19A-19B	4	3.07	Lys71

Chapter 7 • Crystal structure of the RRM domain of PARN

This manuscript has originally been published in the *Journal of Molecular Biology*.

'Crystal Structure of the RRM Domain of Poly(A)-Specific Ribonuclease Reveals a Novel m⁷G-Cap-Binding Mode'

Thomas Monecke[†], Stephanie Schell[†], Achim Dickmanns and Ralf Ficner[‡]

Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik, Abteilung für Molekulare Strukturbiologie (Göttinger Zentrum für Molekulare Biowissenschaften (GZMB), Ernst-Caspari-Haus), Justus-von-Liebig-Weg 11, D-37077, Göttingen

[†] These authors contributed equally to this work.

[‡] To whom correspondence should be addressed:

E-mail: rficner@uni-goettingen.de

Phone: 0551 – 39 14071

Fax: 0551 – 39 14082

Keywords: PARN; RRM; mRNA; cap; deadenylation

Journal of Molecular Biology, VOL. 382, NO. 4, pp. 827-34, October 17, 2008

© 2008 by Elsevier Ltd.

Received April 17, 2008; Accepted July 23, 2008; published online July 31, 2008

Preface – About the Manuscript

The poly(A)-specific ribonuclease (PARN) catalyzes the deadenylation of the poly(A) tail of mRNAs. It is the major deadenylating enzyme in mammalian cells. While it was known that PARN binds, besides the poly(A) tail, the m⁷G-cap of the target mRNA as well, the structural basis of cap binding has remained completely unknown. The aim of the study and subject of the following publication was the characterization of the cap binding mode of the RRM domain of this deadenylase. The instructive finding was that PARN binds the cap completely different than all other known m⁷G-cap binding proteins. While these proteins usually stack the methylated guanine base between two aromatic or hydrophobic side chains, PARN stacks the m⁷G cap only on one side by a single tryptophan, while the opposing side is not occupied by any protein residue. This novel cap binding mode was confirmed by means of site-directed mutagenesis studies and fluorescence spectroscopy measurements.

The coordinates and structure factors of the crystal structure described in the following publication have been deposited in the protein data bank (PDB) under the PDB ID 3CTR.

Author contributions:

Dr. Achim Dickmanns, Annette Berndt and Prof. Dr. Ralf Ficner initiated the project. Molecular cloning, protein purification and crystallization as well as fluorescence measurements were mainly performed by Stephanie Schell within her diploma thesis under the supervision and guidance of Prof. Dr. Ralf Ficner and me. My further contribution concerned data evaluation, structure determination and refinement as well as structure analysis together with Stephanie Schell.

Abstract

Poly(A)-specific ribonuclease (PARN) is a processive 3'-exoribonuclease involved in the decay of eukaryotic mRNAs. Interestingly, PARN interacts not only with the 3' end of the mRNA but also with its 5' end as PARN contains an RRM domain that specifically binds both the poly(A) tail and the 7-methylguanosine (m⁷G) cap. The interaction of PARN with the 5' cap of mRNAs stimulates the deadenylation activity and enhances the processivity of this reaction. We have determined the crystal structure of the PARN-RRM domain with a bound m⁷G triphosphate nucleotide, revealing a novel binding mode for the m⁷G cap. The structure of the m⁷G binding pocket is located outside of the canonical RNA-binding surface of the RRM domain and differs significantly from that of other m⁷G-cap-binding proteins. The crystal structure also shows a remarkable conformational flexibility of the RRM domain, leading to a perfect exchange of two α -helices with an adjacent protein molecule in the crystal lattice.

Introduction

Two characteristics of eukaryotic mRNAs are the 7-methylguanosine (m⁷G) cap located at the 5' end and the poly(A) tail at its 3' end. Both posttranscriptional modifications play a key role as regulator in mechanisms controlling the fate of mRNA, including its synthesis, maturation, transport, translation, stability, and decay. In particular, both the cap and the poly(A) tail are recognized during the general pathways of eukaryotic mRNA decay.^{1–6} In eukaryotes, most of the mRNAs undergo deadenylation-dependent mRNA degradation, whereby the removal of the 3' poly(A) tail is the first, rate-limiting step. So far, it is poorly understood how and when this event is triggered. Subsequent to deadenylation, mRNAs can be degraded by two independent irreversible mechanisms. Firstly, the 5' cap may be removed by the DCP1–DCP2 complex, which leaves the mRNA susceptible for decay by the XRN1 5'→3' exonuclease.^{7–9} Alternatively, the unprotected 3' end can be attacked in 3'→5' direction by the exosome, and the remaining cap is hydrolyzed by the scavenger decapping enzyme DcpS.^{10,11} To date, several eukaryotic deadenylases have been identified, including the Pan2/3 nuclease,^{12,13} the yeast Ccr4/Pop2/Not complex,^{14–16} and the mammalian poly(A)-specific ribonuclease (PARN).^{17–19} Among all deadenylases, PARN is unique since it interacts with both the cap structure and the poly(A) tail

during deadenylation.^{20–23} Functionally, PARN is a processive poly(A)-specific 3'-exoribonuclease that catalyzes the 3'- to 5'-end deadenylation of singlestranded mRNA with a free 3' hydroxyl group both in the nucleus and in the cytoplasm.^{17,18,24} PARN was shown to be the major deadenylase in mammalian cells and belongs to the DEDD superfamily of exonucleases defined by the four conserved acidic amino acid residues DEDD in the active site.^{20–22,25} These residues coordinate two magnesium ions, which are essential for enzymatic activity²⁶ and protein stability.²⁷ PARN from *Homo sapiens* is a 74-kDa multidomain protein and contains besides the nuclease domain an RRM domain and an R3H domain, which is inserted into the nuclease domain with respect to the primary amino acid sequence.^{18,25,28,29} Analytical ultracentrifugation and gel filtration experiments showed that PARN exists as a homodimer in solution.^{25,30} The crystal structure of the C-terminal truncated human PARN (residues 1–430) comprising the nuclease and the R3H domains as well as mutagenesis studies revealed that the dimerization is mediated by the nuclease domain.^{25,30} Interestingly, the R3H domain, which was shown to be involved in stabilization of the enzyme–substrate complex, protrudes from the nuclease domain and is located on the top of the substrate binding site of the other subunit within the homodimer.²⁵

Addressing the question why the RNA-binding R3H and RRM domains are essential for PARN, a comprehensive comparison of activity and stability of PARN mutants, lacking either one or both of the two RNA-binding domains, was carried out. The experiments demonstrated that the presence of one of these domains is sufficient for substrate binding, but both are required for efficient catalysis.³¹

The 5' cap binding of PARN not only stimulates the deadenylation activity but also enhances the processivity of the deadenylation reaction.^{20–22}

The C-terminal RRM domain of PARN was shown to bind the mRNA 5' cap, but the residues involved in m⁷G cap binding have been only partially identified.²⁹

The comparison of three-dimensional structures of other m⁷G-cap-binding proteins reveals a common binding motif mainly consisting of two aromatic side chains sandwiching the monomethylated guanine base of the m⁷G cap.^{32,33} Hence, it was thought that the m⁷G-cap-binding RRM domain of PARN will also apply this common strategy of m⁷G cap recognition. However, recent biochemical studies of human PARN revealed that only one tryptophan residue (Trp475) is essential for cap binding by the RRM domain.²⁹ Furthermore, the NMR structure of the PARN-RRM from *Mus musculus* lacking a bound m⁷G cap shows that only Trp468 (which corresponds to Trp475 of human PARN) is located in the putative binding pocket of the PARN, whereas the side chain of the closest adjacent aromatic residue (Trp449) is facing antipodal with respect to Trp468 [Protein Data Bank (PDB) code: 1WHV; unpublished data].

These results raised the question whether the binding of the m⁷G cap to the PARN-RRM domain will induce a flip of this Trp449 (which corresponds to Trp456 in human PARN), leading to the canonical aromatic sandwich of the m⁷G base. In order to fully characterize the 5' cap binding pocket, we crystallized the PARN-RRM (residues 445–540) of *H. sapiens* in the presence of m⁷G triphosphate (m⁷GTP) and solved and refined the crystal structure at a resolution of 2.1 Å. The crystal structure

analysis reveals a novel binding mode for the m⁷G cap. The m⁷G binding pocket is located outside of the canonical RNA-binding surface of RRM domains and significantly differs from that of other m⁷G-cap-binding proteins.

Overall Structure

The m⁷G-cap-binding RRM domain of human PARN (residues 445–540) was crystallized in the presence of m⁷GTP. The crystals belong to space group *I*4₁22 and contain one protein molecule per asymmetric unit. The crystallographic phase problem was solved by means of multiwavelength anomalous diffraction (MAD) using selenomethionine (SeMet)-labeled protein, and the crystal structure was refined at a resolution of 2.1 Å (Table 1).

The protein folds into a three-stranded antiparallel β -sheet that is flanked by one α -helix connecting β -strands β 1 and β 2. After the third β -strand, another two α -helices, which contain the C-terminal residues of the crystallized domain, protrude from the α/β -core (Fig. 1a). The protruding α -helix, α 2, packs against the corresponding helix of another, symmetry-related protein molecule in the crystal (Fig. 1b). In addition, the short α 3 helices of the monomers are bound in an almost antiparallel fashion to the pair of α 2 helices. Furthermore, each of the two α 2 helices packs against the β -sheet of yet another symmetry-related molecule, leading to a homotetramer with perfect 222 symmetry. The C-terminal α 3 helix of each monomer forms multiple contacts with the β 1 strand of the symmetry-related protein molecule. Interestingly, a part of the loop connecting α 2 and α 3 helices adopts almost a β -strand conformation and interacts with the β 1 strand of the adjacent monomer, thus enlarging this β -sheet to four strands.

However, gel filtration experiments show that the PARN-RRM domain predominantly exists as a monomer, but about 5% elute at a volume corresponding to a homodimer (data not shown). This suggests that the observed fold of the two protruding helices is most likely a consequence of crystal packing. This is further confirmed by

the fact that the crystal structure significantly differs from the three-dimensional structure determined independently by NMR (PDB code: 1WHV; unpublished data). In the NMR structure, the $\alpha 2$ and $\alpha 3$ helices pack against the

four-stranded β -sheet, forming a compact α/β -sandwich structure resembling the canonical RRM fold. Interestingly, within the crystal, the RRM fold is mimicked by the protruding helices of a symmetry-related molecule (Fig. 2).

Table 1. Data statistics of data sets of PARN-RRM (amino acids 445–540) with bound m^7 GTP

Crystals	Native	SeMet		
		Peak	Inflection	High remote
Wavelength (Å)	1.078	0.9778	0.97861	0.91841
Data collection				
Cell dimensions (Å)	a=b=81.11 c=78.06	a=b= 81.38 c=78.62	a=b=81.43 c=78.68	a=b=81.47 c=78.69
Space group	$I4_122$	$I4_122$	$I4_122$	$I4_122$
Resolution range (Å)	30-2.10 (2.18-2.10)	30-2.50 (2.59-2.50)	30-2.60 (2.69-2.60)	30-2.50 (2.59-2.50)
No. of reflections	7918 (776)	4820 (461)	4310 (421)	4804 (463)
Average I/σ	67.9 (14.4)	57.1 (29.4)	55.2 (33.0)	48.7 (12.0)
Completeness (%)	99.9 (100.0)	99.9 (99.6)	99.9 (99.5)	99.9 (99.8)
Redundancy	13.6 (7.8)	11.8 (11.9)	11.8 (12.0)	11.8 (11.8)
R_{merge} (%)	3.8 (20.2)	3.4 (7.7)	3.2 (7.5)	4.5 (19.0)
No. of Se sites (N)		1	1	1
Refinement				
Resolution limits (Å)	15.00-2.1			
No. of used reflections	7448			
No. of protein atoms	602			
No. of ligand atoms	33			
No. of water atoms	69			
R-factor (%)	21.1			
R_{free} (%)	24.3			
Figure of merit	80.6			
Ramachandran plot (%)				
Most favourable region	91.2			
Additionally allowed regions	8.8			
Generously allowed regions	0.0			
Disallowed regions	0.0			
<i>r.m.s. d. from ideal values</i>				
Bond distances (Å)	0.012			
Angles (°)	1.287			
Average B-value (\AA^2)				
Protein molecule	43.5			
m^7 Gppp	63.0			
H ₂ O	53.4			

Values in parentheses indicate the specific values in the highest resolution shell.

$R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I_i(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i \langle I_i(\text{hkl}) \rangle$, where the sum i is over all separate measurements of the unique

reflection hkl. R_{free} as R-factor, but summed over a 5% test set of reflections.

Cloning, expression, and purification of PARN-RRM: The PARN fragment comprising amino acids 445–540 was amplified from a human cDNA library (MegaMan, Stratagene, USA) and cloned into the expression vector pGEX-6P-1. The glutathione S-transferase (GST)-PARN-RRM fusion protein was expressed in *Escherichia coli* BL21(DE3) in 2YT medium at 16 °C. All purification steps were carried out at 4 °C. Cell lysis was performed using a microfluidizer 110S (Microfluidics, USA) in a buffer containing 400 mM NaCl, 50 mM Tris/HCl, pH 8.0, 2 mM ethylenediaminetetraacetic acid (EDTA), and 2 mM DTT. The lysate was centrifuged at 30,000g, and the supernatant was loaded onto a glutathione-Sepharose column (GE Healthcare, Germany). GST-PARN-RRM was eluted with lysis buffer containing 30 mM reduced glutathione. GST-PARN-RRM was incubated with PreScission protease (GE Healthcare) at 4 °C overnight. A final gel filtration (Superdex 75, GE Healthcare) was performed using a buffer containing 300 mM NaCl, 20 mM Tris/HCl, pH 8.0, and 2 mM DTT. PARN-RRM was concentrated to 8.8 mg ml⁻¹ using a vivaspin concentrator (Sartorius, Germany), and m⁷GTP (Sigma, Germany) was added in sixfold molar excess. SeMet-containing PARN-RRM was expressed according to the protocol described by Reuter *et al.*³⁴ Purification was analogous to that of native PARN-RRM, with the exception that, in all buffers, the DTT concentration was elevated to 5 mM.

Crystallization and data collection: PARN-RRM was crystallized in 2 M Li₂SO₄ and 0.05 M 4-morpholineethanesulfonic acid, pH 5.5, for the native PARN–cap complex. In the case of the SeMet–PARN–cap complex, the reservoir solution additionally contained 0.159 M NaBr and 1% (v/v) β-mercaptoethanol. The crystals were flash-cooled in a 100 K cryostream and annealed once for 5 s before data collection. A complete data set was collected at beamline BW7A of the Deutsches Elektronen-Synchrotron/European Molecular Biology Laboratory (Hamburg) up to a resolution of 2.1 Å. MAD data sets of SeMet–PARN-RRM were collected at beamline 14.2 at BESSY (Berlin).

Data processing, phasing and density modification: Data were processed with HKL2000 (HKL Research, USA), and selenium sites were found and refined by SHELX/HKL2MAP.³⁵ An initial model of PARN-RRM was built by ARP/wARP.³⁶ The model was improved manually using Coot and refined by REFMAC5.^{37–39} The final model consists of residues 445–514 of PARN. The 25 C-terminal residues are not defined in the electron density map. Figures were generated using PyMOL [PyMOL Molecular Graphics System (2002), DeLano Scientific, USA].

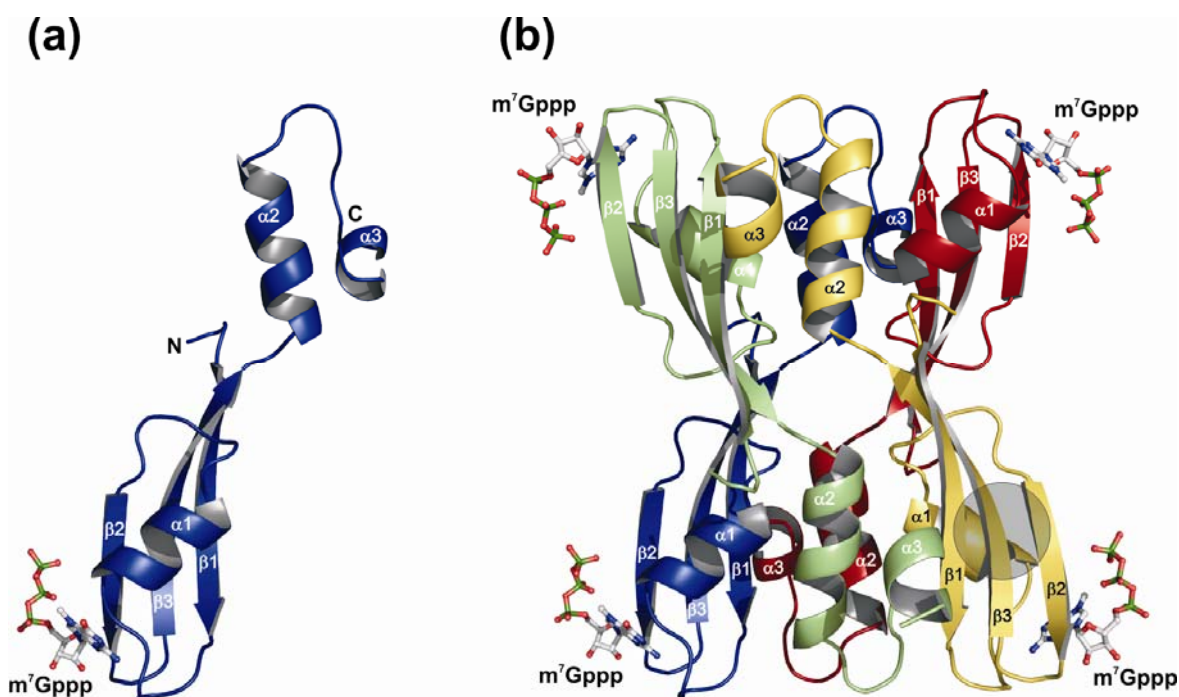


Fig. 1. Structure of the PARN-RRM domain. **(a)** PARN-RRM with bound m^7 GTP. The PARN-RRM (residues 445–514, shown in cartoon mode) reveals an extended shape with the two C-terminal α -helices $\alpha 2$ and $\alpha 3$ protruding from protein core. The m^7 GTP shown in ball-and-stick mode (red, oxygen; green, phosphor; blue, nitrogen; and gray, carbon) is bound on the surface of the N-terminal part of the RRM domain involving the loops connecting $\beta 2$ – $\beta 3$ and $\beta 1$ – $\alpha 1$, respectively. **(b)** Crystallographic tetramer of PARN-RRM. The protruding helices $\alpha 2$ and $\alpha 3$ mediate the assembly of a homotetramer in the crystal. All four m^7 GTPs emanate from the corners of the tetramer. The position of the canonical RRM-RNA binding site is indicated on the gold subunit as gray area.

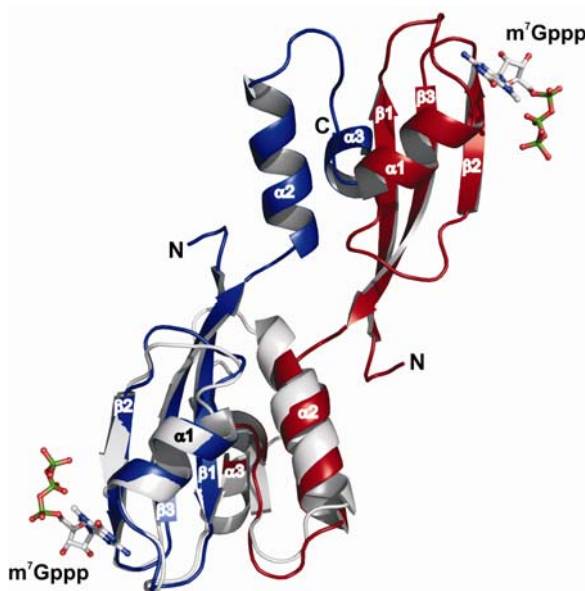


Fig. 2. The crystallographic dimer mimics the NMR structure. The protruding C-terminal $\alpha 2$ and $\alpha 3$ helices of each monomer pack on the N-terminal β -sheet in a way that they mimic the NMR structure (colored gray) of the PARN-RRM (PDB code: 1WHV). The loop between the $\alpha 2$ and $\alpha 3$ helices corresponds to the fourth β -strand of the canonical RRM fold.

Nevertheless, the crystal structure documents a conformational flexibility of the RRM domain, which raises the question of a functional relevance. Biochemical data showing the enhancement of activity and processivity upon m^7 G cap binding indicate a cross talk between the R3H, the nuclease, and the RRM domains of PARN, which might require some conformational flexibility of the PARN domains.³¹

m^7 G cap binding site

The crystal structure analysis reveals additional electron density located on the surface of the RRM domain, which clearly corresponds to a bound m^7 GTP. The beta and gamma phosphates of the m^7 GTP are weakly defined in the electron density map; hence, they appear to be rather

flexible (Fig. 3). The m^7 G cap binding pocket is formed by residues belonging to the $\beta 2$ strand and the two loops connecting $\beta 1$ – $\alpha 1$ and $\beta 2$ – $\beta 3$, respectively. Importantly, the bound m^7 GTP does not form any contacts to crystal symmetry-related molecules. The 7-methyl-guanine stacks on the side chain of Trp475 in a perfect coplanar orientation. The exocyclic NH_2 group of the m^7 G forms hydrogen bonds to the main chain carbonyl oxygens of Lys454, Trp456, and Asp478 (Fig. 3). Another hydrogen bond occurs between the N1 of m^7 G and the carbonyl oxygen of Trp456, and a water molecule mediates hydrogen bonding between the Thr458 side chain and the O6 atom of the m^7 G.

Surprisingly, the m^7 G cap binding pocket of PARN differs significantly to that of other known m^7 G cap binding proteins, namely, the small subunit of the cap-binding complex (CBC20), which binds to the 5' cap of RNAs in the nucleus,^{40,41} the translation initiation factor eIF4E,^{42–46} and the viral methyltransferase VP39.^{47–50} The analysis of the three-dimensional structures of these proteins reveals a common strategy for specific recognition of m^7 G cap involving two coplanar aromatic side chains that sandwich the bound m^7 G base (Fig. 4). The m^7 G base is bound by π – π interactions and, additionally, by a cation– π interaction due to the cationic charge of the m^7 G base. This cation– π interaction was proposed to be responsible for the discrimination of unmethylated and uncharged GTP. This is in line with our findings that GTP binds to the PARN-RRM with an affinity that is at least 100 times lower than that for m^7 GTP (data not shown). It has been proposed that, also, in PARN, two aromatic residues, namely, Trp456 and Trp475, form the canonical aromatic sandwich binding the m^7 G cap. The crystal structure clearly shows that Trp475 provides the expected base-stacking interaction, while the side chain of Trp456 is not

in contact to the bound m^7 GTP. However, the main chain carbonyl oxygen of Trp456 is within perfect hydrogen-bonding distance of the m^7 G N1 atom. Trp456 is part of the loop connecting the $\beta 1$ strand and the $\alpha 1$ helix, which is very

well defined in the electron density map, indicating its rigidity. The side chain of Trp456 is packed between the side chains of Phe452, Pro453, and Asp460, fixing the loop in the observed conformation.

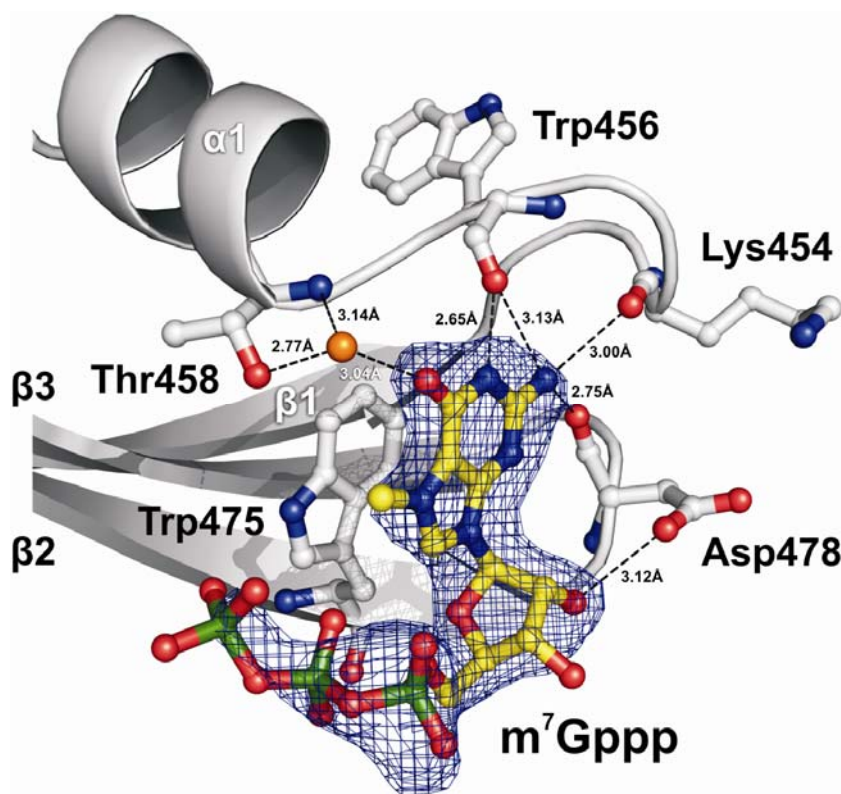


Fig. 3. Detailed view of the cap binding site of the PARN-RRM with bound m^7 GTP. The m^7 GTP cap (shown in ball-and-stick mode) binds on the surface of the N-terminal part of the RRM domain involving the loops connecting $\beta 2$ and $\beta 3$ strands as well as the $\beta 1$ strand and the $\alpha 1$ helix. The $2|F_o| - |F_c|$ electron density (blue) corresponding to the bound m^7 GTP is contoured at 1.1 σ . The interacting residues are depicted in ball-and-stick mode. The guanine moiety stacks on the side chain of Trp475 and is bound by several hydrogen bonds. A water molecule mediating hydrogen bonds between Thr458 and the exocyclic oxygen O6 of the cap guanine is colored orange.

This stable loop conformation is important as the carbonyl oxygens of Trp456 and the adjacent Lys454 form direct hydrogen bonds to the m^7 GTP. Hence, the mutation of Trp456 to Ala is expected to reduce the loop's stability and, therefore, to weaken the hydrogen bonds to the m^7 G. This is consistent with the previously reported K_d values for m^7 GTP that have been determined by means of fluorescence spectroscopy.²⁹ While the mutation of Trp475 in full-length PARN completely abolishes m^7 GTP binding, the mutation Trp456Ala increases the K_d value from 1.6 to 5.6 μ M. Interestingly, the RRM domain by itself shows a reduced K_d value

of 11 μ M, indicating that residues outside the RRM domain contribute to m^7 G cap binding as well.

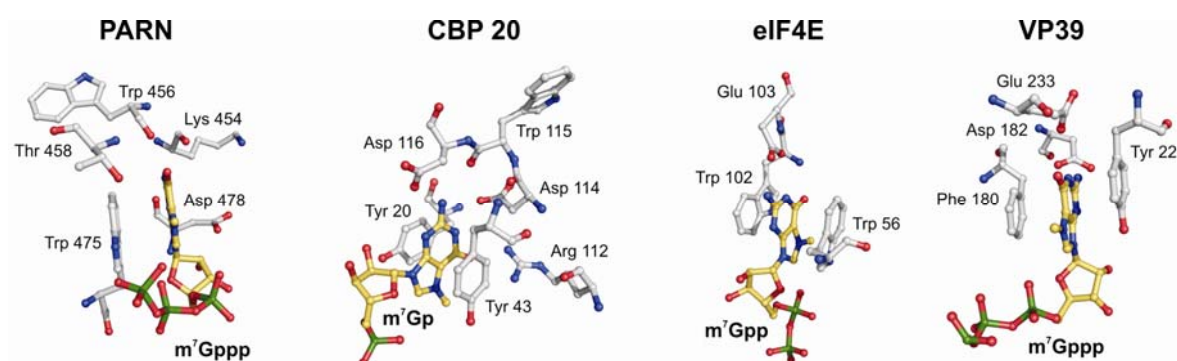
Mutational analysis of the m^7 G cap binding site

In order to characterize the binding of the cap analog m^7 GTP to PARN-RRM (amino acids 445–540) as well as the effect of several point mutations, we performed fluorescence titrations. The results of the titration measurements are summarized in Table 2.

Table 2. Binding constants of m⁷Gppp for human PARN-RRM domain and mutants

$K_d \pm \Delta K_d$ (μ M) m ⁷ Gppp
--

PARN (aa445-540)	6.94 \pm 2.08
PARN (aa445-540, K454A)	20.03 \pm 3.43
PARN (aa445-540, T458A)	30.58 \pm 4.29
PARN (aa445-540, D478A)	20.00 \pm 7.03

**Fig. 4.** Comparison of cap binding pockets. m⁷G cap binding pockets of CBP20, eIF4E, and the viral nucleoside 2'-O-methyltransferase (VP39) in comparison to the m⁷G cap binding pocket of PARN-RRM. The m⁷G cap molecules are shown in ball-and-stick mode (red, oxygen; green, phosphor; blue, nitrogen; and yellow, carbon).

Site-directed mutagenesis: The PARN-RRM mutants K454A, T458A, and D478A were generated from the wild-type clone (pGEX-6P-1 PARN445–540) using the QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene, USA) following the manufacturer's protocol. All mutations were confirmed by DNA sequencing.

Fluorescence spectroscopy: Expression and purification of PARN (amino acids 445–540) and PARN mutants used for fluorescence measurements were analogous to that of PARN-RRM (amino acids 445–540) used for crystallization, with the exception that the final gel filtration buffer contained 100 mM KCl, 20 mM Hepes–KOH, pH 7.0, 10% glycerol, 1.5 mM MgCl₂, 0.2 mM EDTA, and 0.5 mM DTT. To determine the binding affinity of PARN-RRM (amino acids 445–540) and the mutants K454A, T458A and D478A for m⁷GTP, we mixed 1 μ M purified protein with increasing concentrations of m⁷Gppp (0–200 μ M).

Fluorescence measurements were performed at 20 °C using a Fluoromax3[™] spectrofluorimeter (Jobin Yvan) in 0.5 cm \times 1.0 cm section cuvettes (Hellma) with a buffer containing 100 mM KCl, 20 mM Hepes–KOH, pH 7.0, 10% glycerol, 1.5 mM MgCl₂, 0.2 mM EDTA, and 0.5 mM DTT. Fluorescence emission at 315 nm (bandwidth, 10 nm) was recorded during excitation at 295 nm for 60 s (bandwidth, 1 nm) with time constants of 0.5 s. Values of two independent measurements were averaged, and the spectra were normalized with respect to the cap and buffer contributions. Equilibrium dissociation constants were obtained by fitting the solutions of a quadratic function, assuming a 1:1 stoichiometry and

taking into account the fluorescence of the cap and the buffer using SigmaPlot.

For the wild-type PARN-RRM, a K_d value of 6.94 \pm 2.08 μ M was determined. This K_d value differs slightly from the one obtained by Nilsson *et al.* (11.1 \pm 0.2 μ M) using a marginally different PARN fragment (amino acids 445–560).²⁹ In comparison to other m⁷G-cap-binding proteins, the K_d value of the PARN-RRM is about 3 orders of magnitude higher, as the K_d values of eIF4E and CBC have been reported to be 10 and 33 nM, respectively.^{44,51} This significant difference is most likely due to the fact that the positively charged 7-methylguanine stacks between two aromatic side chains of eIF4E and CBC (Fig. 4), while it stacks only one tryptophan side chain of the PARN-RRM. The crystal structure of the PARN-RRM–m⁷GTP complex demonstrates the contribution of other residues, namely, Lys454, Thr458, and Asp478, in m⁷G cap binding. Three PARN mutants (Lys454Ala, Thr458Ala, and Asp478Ala) were generated and their m⁷GTP-binding affinities were determined by fluorescence titrations to further characterize the impact of these residues in m⁷G cap binding (Table 2). All mutants exhibit increased K_d values for m⁷GTP, which

are at least twofold higher than that for the wild-type PARN-RRM. This confirms that Lys454, Thr458, and Asp478 also contribute significantly to the binding enthalpy, but Trp475 is the most important residue for m⁷G cap binding.

Interestingly, full-length PARN was reported to bind the m⁷GTP with an almost sevenfold higher affinity (1.59±0.11 μM) than the PARN-RRM (amino acids 445–560).²⁹ These data as well as the crystal structure of the PARN-RRM–m⁷GTP complex suggest that residues belonging to the R3H domain and/or the nuclease domain complete the cap binding pocket.

PDB accession code

The coordinates of the PARN-RRM–m⁷GTP complex and the structure factors have been deposited in the Research Collaboratory for Structural Bioinformatics PDB and are available under the accession code 3CTR.

Acknowledgements

We are grateful to Manfred Weiß for support and help during data collection at Deutsches Elektronen-Synchrotron/European Molecular Biology Laboratory beamline BW7A and to Uwe Müller for help and support during MAD data collection at BESSY beamline 14.2. We thank our colleagues Annette Berndt for technical assistance and Daniel Wohlwend for crystallographic support and discussions. This work was supported by the Deutsche Forschungsgemeinschaft (SFB523 and FOR855).

References

1. Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr. Opin. Cell Biol.* **17**, 242–250.
2. Garneau, N. L., Wilusz, J. & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126.
3. Meyer, S., Temme, C. & Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 197–216.
4. Parker, R. & Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* **11**, 121–127.
5. Wilusz, C. J., Wormington, M. & Peltz, S. W. (2001). The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* **2**, 237–246.
6. Sachs, A. & Wahle, E. (1993). Poly(A) tail metabolism and function in eucaryotes. *J. Biol. Chem.* **268**, 22955–22958.
7. van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E. & Seraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* **21**, 6915–6924.
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. Hsu, C. L. & Stevens, A. (1993). Yeast cells lacking 5'→3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* **13**, 4826–4835.
10. Chen, N., Walsh, M. A., Liu, Y., Parker, R. & Song, H. (2005). Crystal structures of human DcpS in ligandfree and m⁷GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J. Mol. Biol.* **347**, 707–718.
11. van Dijk, E., Le Hir, H. & Seraphin, B. (2003). DcpS can act in the 5'–3' mRNA decay pathway in addition to the 3'–5' pathway. *Proc. Natl Acad. Sci. USA*, **100**, 12081–12086.
12. Boeck, R., Tarun, S., Jr, Rieger, M., Deardorff, J. A., Muller-Auer, S. & Sachs, A. B. (1996). The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.* **271**, 432–438.
13. Uchida, N., Hoshino, S., Imataka, H., Sonenberg, N. & Katada, T. (2002). A novel role of the mammalian GSPT/eRF3 associating with poly(A)-binding protein in Cap/Poly(A)-dependent translation. *J. Biol. Chem.* **277**, 50286–50292.
14. Daugeron, M. C., Mauxion, F. & Seraphin, B. (2001). The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* **29**, 2448–2455.
15. Tucker, M., Staples, R. R., Valencia-Sanchez, M. A., Muhlrads, D. & Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *EMBO J.* **21**, 1427–1436.

16. Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L. & Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell*, **104**, 377–386.
17. Astrom, J., Astrom, A. & Virtanen, A. (1992). Properties of a HeLa cell 3' exonuclease specific for degrading poly(A) tails of mammalian mRNA. *J. Biol. Chem.* **267**, 18154–18159.
18. Korner, C. G. & Wahle, E. (1997). Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *J. Biol. Chem.* **272**, 10448–10456.
19. Korner, C. G., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E. & Wahle, E. (1998). The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes. *EMBO J.* **17**, 5427–5437.
20. Dehlin, E., Wormington, M., Korner, C. G. & Wahle, E. (2000). Cap-dependent deadenylation of mRNA. *EMBO J.* **19**, 1079–1086.
21. Martinez, J., Ren, Y. G., Nilsson, P., Ehrenberg, M. & Virtanen, A. (2001). The mRNA cap structure stimulates rate of poly(A) removal and amplifies processivity of degradation. *J. Biol. Chem.* **276**, 27923–27929.
22. Gao, M., Wilusz, C. J., Peltz, S. W. & Wilusz, J. (2001). A novel mRNA-decapping activity in HeLa cytoplasmic extracts is regulated by AU-rich elements. *EMBO J.* **20**, 1134–1143.
23. Opyrchal, M., Anderson, J. R., Sokoloski, K. J., Wilusz, C. J. & Wilusz, J. (2005). A cell-free mRNA stability assay reveals conservation of the enzymes and mechanisms of mRNA decay between mosquito and mammalian cell lines. *Insect Biochem. Mol. Biol.* **35**, 1321–1334.
24. Copeland, P. R. & Wormington, M. (2001). The mechanism and regulation of deadenylation: identification and characterization of *Xenopus* PARN. *RNA*, **7**, 875–886.
25. Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E. & Song, H. (2005). Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.* **24**, 4082–4093.
26. Ren, Y. G., Kirsebom, L. A. & Virtanen, A. (2004). Coordination of divalent metal ions in the active site of poly(A)-specific ribonuclease. *J. Biol. Chem.* **279**, 48702–48706.
27. Liu, W. F., Zhang, A., Cheng, Y., Zhou, H. M. & Yan, Y. B. (2007). Effect of magnesium ions on the thermal stability of human poly(A)-specific ribonuclease. *FEBS Lett.* **581**, 1047–1052.
28. Moser, M. J., Holley, W. R., Chatterjee, A. & Mian, I. S. (1997). The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res.* **25**, 5110–5118.
29. Nilsson, P., Henriksson, N., Niedzwiecka, A., Balatsos, N. A., Kokkoris, K., Eriksson, J. & Virtanen, A. (2007). A multifunctional RNA recognition motif in poly(A)-specific ribonuclease with cap and poly(A) binding properties. *J. Biol. Chem.* **282**, 32902–32911.
30. Dehlin, E. (2004). Cap dependence of the poly(A)-specific ribonuclease PARN. Dissertation Universität Halle-Wittenberg.
31. Liu, W. F., Zhang, A., He, G. J. & Yan, Y. B. (2007). The R3H domain stabilizes poly(A)-specific ribonuclease by stabilizing the RRM domain. *Biochem. Biophys. Res. Commun.* **360**, 846–851.
32. Hsu, P. C., Hodel, M. R., Thomas, J. W., Taylor, L. J., Hagedorn, C. H. & Hodel, A. E. (2000). Structural requirements for the specific recognition of an m⁷G mRNA cap. *Biochemistry*, **39**, 13730–13736.
33. Quijcho, F. A., Hu, G. & Gershon, P. D. (2000). Structural basis of mRNA cap recognition by proteins. *Curr. Opin. Struct. Biol.* **10**, 78–86.
34. Reuter, K., Nottrott, S., Fabrizio, P., Luhrmann, R. & Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* **294**, 515–525.
35. Pape, T. & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844.
36. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiplerefinement dummy atomic models. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 448–455.
37. Collaborative Computational Project, No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
38. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.

39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
40. Calero, G., Wilson, K. F., Ly, T., Rios-Steiner, J. L., Clardy, J. C. & Cerione, R. A. (2002). Structural basis of m⁷GpppG binding to the nuclear cap-binding protein complex. *Nat. Struct. Biol.* **9**, 912–917.
41. Mazza, C., Segref, A., Mattaj, I.W. & Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* **21**, 5548–5557.
42. Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1997). Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell*, **89**, 951–961.
43. Matsuo, H., Li, H., McGuire, A. M., Fletcher, C. M., Gingras, A. C., Sonenberg, N. & Wagner, G. (1997). Structure of translation factor eIF4E bound to m⁷GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.* **4**, 717–724.
44. Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszyńska, A., Dadlez, M. *et al.* (2002). Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4EBP1 proteins. *J. Mol. Biol.* **319**, 615–635.
45. Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S. *et al.* (2002). Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region. *Biochem. J.* **362**, 539–544.
46. Tomoo, K., Shen, X., Okabe, K., Nozoe, Y., Fukuhara, S., Morino, S. *et al.* (2003). Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations. *J. Mol. Biol.* **328**, 365–383.
47. Hodel, A. E., Gershon, P. D. & Quijcho, F. A. (1998). Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol. Cell*, **1**, 443–447.
48. Hodel, A. E., Gershon, P. D., Shi, X., Wang, S. M. & Quijcho, F. A. (1997). Specific protein recognition of an mRNA cap through its alkylated base. *Nat. Struct. Biol.* **4**, 350–354.
49. Hu, G., Gershon, P. D., Hodel, A. E. & Quijcho, F. A. (1999). mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. USA*, **96**, 7149–7154.
50. Hu, G., Oguro, A., Li, C., Gershon, P. D. & Quijcho, F. A. (2002). The “cap-binding slot” of an mRNA capbinding protein: quantitative effects of aromatic side chain choice in the double-stacking sandwich with cap. *Biochemistry*, **41**, 7677–7687.
51. Worch, R., Niedzwiecka, A., Stepinski, J., Mazza, C., Jankowska-Anyszka, M., Darzynkiewicz, E. *et al.* (2005). Specificity of recognition of mRNA 5' cap by human nuclear cap-binding complex. *RNA*, **11**, 1355–1363.

Chapter 8 • Additional Results and Discussion

The aim of the present work was the structural and biochemical characterization of a protein and a protein complex involved in the biogenesis of spliceosomal UsnRNPs and of the RNA recognition motif (RRM) of poly(A)-specific ribonuclease PARN. The special focus within the first subject was the dimethyltransferase TGS1 (Trimethylguanosine Synthase 1) ([Chapter 4](#) and [5](#)) as well as the nuclear export complex CRM1·SPN1·RanGTP ([Chapter 6](#)). TGS1 catalyzes the double methylation of the m⁷G-cap of spliceosomal snRNAs. The hypermethylated cap then serves as signal that the cytoplasmic assembly process is finished and the generated m₃G-cap is recognized by the import adapter snurportin 1 (SPN1). After nuclear import of the snRNP-SPN1 complex using the import receptor importin β and subsequent release of the snRNP in the nucleus, the import adapter has to be recycled back to the cytoplasm. The recycling is mediated by the nuclear export receptor CRM1. The investigated nuclear export complex consists of the export receptor, which, in its RanGTP bound form, is able to bind the export cargo SPN1 and to pass the nuclear pore. Apart from these two closely related subjects, the third more distant focus concerned the characterization of the m⁷G-cap binding mode of the RRM domain of poly(A)-specific ribonuclease (PARN) ([Chapter 7](#)). The following chapter is focused on additional and supporting results of the three projects as well as their discussion.

8.1 The Trimethylguanosine Synthase 1

The dimethyltransferase TGS1 is biochemically well characterized in numerous organisms including *Homo sapiens*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Giardia lamblia*, *Trypanosoma brucei* and *Drosophila melanogaster* (Colau *et al.*, 2004; Enunlu *et al.*, 2003; Franke *et al.*, 2008; Girard *et al.*, 2008; Hausmann *et al.*, 2007; Hausmann and Shuman, 2005; Hausmann *et al.*, 2008; Komonyi *et al.*, 2005; Misra *et al.*, 2002; Mouaikel *et al.*, 2003a; Mouaikel *et al.*, 2003b; Mouaikel *et al.*, 2002; Plessel *et al.*, 1994; Ruan *et al.*, 2007; Zhu *et al.*, 2001). Despite its solid functional description, the three dimensional structure of that enzyme, allowing a sequence-structure-function relationship analysis has been unavailable so far.

In order to reveal the structural organization of the predicted and conserved C-terminal methyltransferase domain of *H. sapiens* TGS1 (hTGS1) as well as the catalytic mechanism of dimethyltransfer, this domain was purified and crystallized. Initially, a C-terminal fragment comprising the amino acids 653-853 (hTGS1₆₅₃₋₈₅₃) has been crystallized in the presence of the cap dinucleotide m⁷GpppA (Chapter 4; Figure 8-1 A).

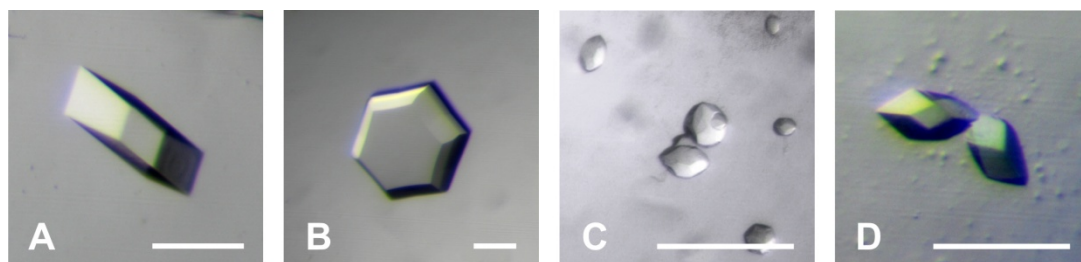


Figure 8-1: Crystals of hTGS1 fragments of various lengths bound to different substrates. (A) Tetragonal crystals of inactive hTGS1₆₅₃₋₈₅₃ bound to the cap dinucleotide m⁷GpppA (Chapter 4). (B) Trigonal crystals of hTGS1₆₅₃₋₈₅₃ in complex with the methyl group donor *S*-adenosyl-L-methionine (Chapter 4). (C) Hexagonal crystals of inactive hTGS1₆₆₄₋₈₅₃ crystallized together with both, m⁷GTP and *S*-adenosyl-L-homocysteine. (D) Rhombohedral crystals of the active form of hTGS1₆₁₈₋₈₅₃ bound to m⁷GTP and *S*-adenosyl-L-homocysteine (Chapter 5). The scale bar at the bottom of each panel corresponds to a length of 100 μm.

hTGS1₆₅₃₋₈₅₃ corresponds to the residues 58-266 of the *S. cerevisiae* TGS1 (yTGS1), which in turn were predicted to represent the catalytically active core of the enzyme mediating the cap hypermethylation in yeast (Mouaikel *et al.*, 2003a). The structure was solved by means of SeMet-MAD (selenomethionine multiwavelength anomalous diffraction) to a resolution of 2.2 Å. However, the crystal structure revealed that this fragment lacks an intact m⁷G-cap binding site, as the predicted pocket was not occupied by the m⁷guanine (Chapter 4). Instead, the second base, an adenine, of the cap dinucleotide was bound in the highly conserved *S*-adenosyl-L-methionine (AdoMet) binding site of the protein. This conformation strongly suggested that the hTGS1₆₅₃₋₈₅₃ is not able to bind the m⁷G-cap in the correct way and if so, it should be catalytically inactive. To prove this hypothesis an HPLC-based activity assay was developed and the purified hTGS1₆₅₃₋₈₅₃ was tested in this methyltransferase assay. It turned out that this fragment indeed lacks catalytic activity, as it was not able to convert m⁷GpppA and AdoMet to m^{2,2,7}GpppA and *S*-adenosyl-L-homocysteine (AdoHcy), respectively. While the cap binding site of hTGS1₆₅₃₋₈₅₃ seems to be incomplete, the structurally conserved AdoMet binding site is functional as indicated by the binding of the adenosine moiety of m⁷GpppA. To confirm this hypothesis the purified protein was mixed with AdoMet in a six-fold molar excess and crystallized (Figure 8-1 B). The crystals belonged to a trigonal spacegroup (*P*3) with the cell constants *a*=*b*=135.5 Å and *c*=165.9 Å and angles of $\alpha=\beta=90^\circ$

and $\gamma=120^\circ$. They diffracted to a maximum resolution of 3.3 Å and data evaluation using HKL2000 resulted in an R_{merge} of 8.9% and an overall completeness of 97%. The structure was solved by means of molecular replacement using the previously solved SeMet-hTGS1₆₅₃₋₈₅₃ structure in PHASER (McCoy, 2007) as starting model. However, during structure refinement with REFMAC5 (Murshudov *et al.*, 1997), SHELXL (Sheldrick, 2008) and CNS (Brunger, 2007) the R_{work} and R_{free} -values converged to 30% and 35%, respectively. After careful data reexamination including analysis of the cumulative intensity distribution and further tests using the program PHENIX.XTRIAGE (Adams *et al.*, 2002) it turned out that these crystals were almost perfectly twinned, which typically results in high R-factors during structure refinement. Nevertheless, additional electron density in the predicted AdoMet binding site clearly corresponds to the cocrystallized methyl group donor AdoMet (not shown). This situation further confirms the hypothesis of an intact AdoMet binding site, while the cap binding site is incomplete leading to a catalytically inactive enzyme.

In contrast to hTGS1₆₅₃₋₈₅₃ a fragment containing 17 additional residues N-terminally (hTGS1₆₃₆₋₈₅₃) exhibits catalytic activity as shown by the applied activity test (Chapter 4). In order to determine the shortest possible active hTGS1 fragment and thus the exact N-terminal residue required, several additional TGS1 fragments were purified including hTGS1 aa644-853, aa646-853, aa647-853 and aa648-853. All purified fragments were tested in the HPLC-based activity test and sections of the chromatograms are shown in Figure 8-2.

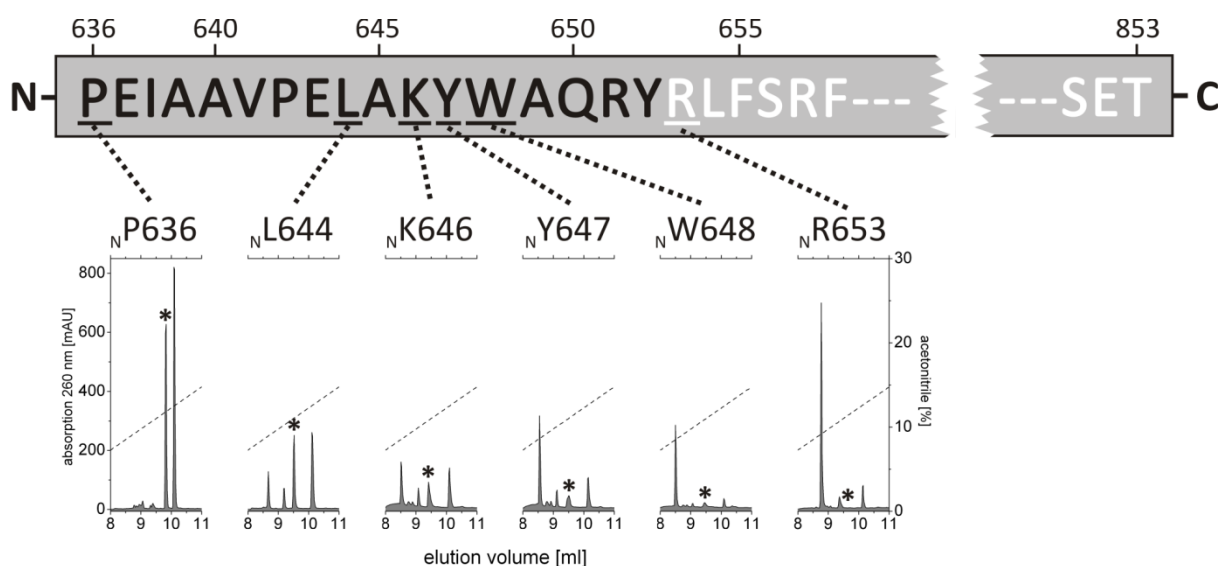


Figure 8-2: Dissection of the methyltransferase-domain and importance of N-terminal residues for catalytic activity. On the top the amino acid sequence of the N-terminus is shown, while the corresponding activity test chromatograms are shown below. The reaction product m^{2,2,7}(₃)GpppA is marked by an asterisk. The crystallized MTase domain (aa653-853; N₆₅₃) is catalytically inactive. Gradual N-terminal elongation of the MTase domain leads to catalytic activity.

The upper panel shows the amino acid sequence of the N-terminal part of hTGS1 fragments. Below, magnifications of the corresponding chromatograms are shown indicating the peak for the reaction product $m^{2,2,7}\text{GpppA}$, which is marked by an asterisk. As protein and substrate concentrations as well as the incubation time were identical, a higher $m^{2,2,7}\text{GpppA}$ peak indicates a higher activity of the tested fragment. Interestingly, the activity is not dependent on a single key residue, but increasing continuously with each additional residue added to the N-terminus. The addition of further amino acid residues to the N-terminus and thus elongating the fragment beyond Pro636 (Figure 8-2) did not lead to a significant increase of activity (not shown). The results indicate that the requirement for these residues might be more structural than functional, which means that the supplemental N-terminal amino acids may stabilize an active conformation of this part or contribute to a functional $m^7\text{G}$ binding site rather than providing a key residue for catalysis (Chapter 5).

Since the crystallization of the shortest fully active hTGS1₆₃₆₋₈₅₃ did not succeed, neither in the presence nor in the absence of substrates or substrate analogs, several further elongated and proven active fragments were tested in crystallization. The crystallization of the active fragment hTGS1₆₁₈₋₈₅₃ bound to the minimal substrate $m^7\text{GTP}$ as well as to the substrate analog AdoHcy led to rhombohedral crystals, which diffracted to a maximum resolution of 2.0 Å (Figure 8-1 D). The structure was solved by means of molecular replacement using the previously solved structure of the inactive SeMet-hTGS1₆₅₃₋₈₅₃. The crystal structure analysis revealed that besides the conserved methyltransferase fold, a small N-terminal, α -helical extension (NTE) is required for enzymatic activity. This small domain is stabilized by a hydrophobic core, destruction of which leads to an inactive conformation, similar to that observed in the crystal structure of hTGS1₆₅₃₋₈₅₃. This instructive finding together with the facts that the activity is not dependent on a single N-terminal residue and that the AdoMet binding site in the inactive methyltransferase fragment is functional, finally strongly supports the hypothesis mentioned above.

The 5'-cap hypermethylation, as a process which changes the properties of RNA molecules dramatically, has to be strictly controlled, since other RNA types (e.g. mRNAs) carry an $m^7\text{G}$ cap at their 5' ends as well, which has to be protected from this modification. It has been shown in detail that both human and yeast TGS1 interact with Sm-proteins common to all snRNPs as well as with proteins, which are specific for particular RNPs (Mouaikel *et al.*, 2003a; Mouaikel *et al.*, 2003b; Mouaikel *et al.*, 2002; Plessel *et al.*, 1994). In more detail, the results of several binding studies suggested that TGS1 is able to bind to the C-terminal tails of

Sm-proteins B and D1, which are known to bind to the snRNAs during their maturation. Besides these interactions the snoRNP specific proteins Cbf5 and Nop58 were shown to interact with TGS1. Hence, it is thought that binding to these proteins may be a control mechanism and thus the interaction could enhance the catalytic activity of the enzyme or, *in vivo*, even enable it (Raker *et al.*, 1996). In order to investigate these interactions structurally and to reveal their possible impact in activation of TGS1 the future work on this enzyme will mainly focus on its cocrystallization with interacting proteins. These are for example the C-terminal tails of Sm-proteins B and D1, the survival of motor neuron protein (SMNp) or the snoRNP-specific proteins Cbf5 and Nop58.

Prior to the knowledge of the crystal structure of the active hTGS1 methyltransferase domain a reaction mechanism was proposed on the basis of structures of other methyltransferases and mutagenesis studies (Hausmann *et al.*, 2007; Hausmann *et al.*, 2008; Mouaikel *et al.*, 2003a). This mechanism requires the positioning of the guanine N2-atom by two main chain carbonyls (Ser763 and Pro764) and additionally stacking of the m⁷guanine between two aromatic or hydrophobic amino acids. Moreover, it was suggested that the side chain of Ser763 might act as a general base, which abstracts a proton from the guanine N2 in order to enhance its nucleophilicity to attack the reactive methyl group of AdoMet. However, details of the mechanism and of activation of this serine residue or the final proton acceptor have been completely unknown. Using the structural information of the active methyltransferase, the reaction mechanism for dimethyltransfer could in general be confirmed with some modifications (Chapter 5 and Figure 8-3). The role of hTGS1 Ser763 is controversially discussed, since mutations of this residue in *D. melanogaster*, *S. cerevisiae* and human TGS1 led to different effects. Nevertheless, the Ser763 hydroxyl is a perfect candidate for the role as primary proton carrier as observed in the crystal structure and by mutational analyses (Chapter 5). In the crystal structure, the distance between the Ser763-OH and the guanine N2 averages 4.1 Å representing an interspace, which is too big for a direct proton transfer. Instead, the hydroxyl group is hydrogenbonded to a water molecule, which is positioned by the Asp696 carboxyl and the Phe761 main chain carbonyl group. This could allow a transfer from N2 *via* Ser763 and the water molecule finally to the side chain of Asp696. To verify this hypothesis the single amino acid mutants Asp696Ala and Asp696Asn were generated in order to show the importance of the side chain carboxyl group. However, during protein purification of both mutants the protein precipitated immediately subsequent to elution from the GSH-sepharose column. This effect was highly reproducible as the mutants were sequenced and

checked twice and the purification was repeated three times. This effect was observed for both, the aspartate-alanine and the more conservative aspartate-asparagine mutation, indicating the requirement for the long, negatively charged side chain in that position. In contrast to the results mentioned above, another group reported the successful purification and characterization of the mutant Asp696Ala (Hausmann *et al.*, 2008). In this study it reduced the activity of the fragment hTGS1₆₃₁₋₈₅₃ to below 1%, while the protein was stable. A possible reason for this difference could be that Hausmann *et al.* used a different, slightly shorter fragment of hTGS1. Nevertheless, it remains unclear why hTGS1₆₃₁₋₈₅₃ is stable in solution, while the longer hTGS1₆₁₈₋₈₅₃ precipitates during purification, when a structurally and functionally important residue is mutated to alanine.

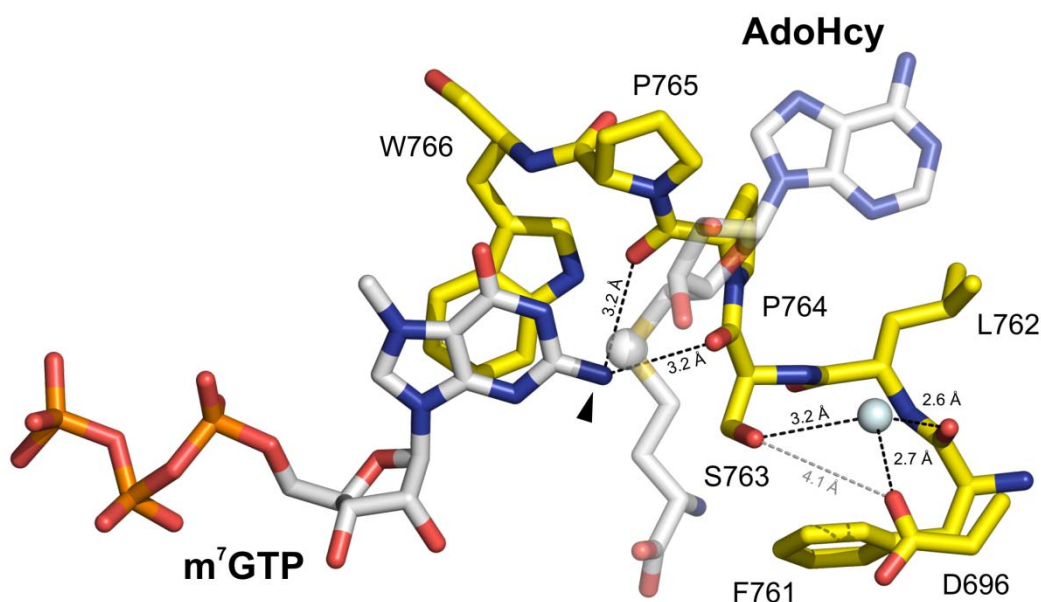


Figure 8-3: Proposed catalytic mechanism of methyltransfer by hTGS1 by an S_N2 substitution. The methylation target N2 (arrow) is hydrogen bonded by main chain atoms of Ser763 and Pro764. These interactions in combination with the net positive charge of the methylated purine ring and the activation by Ser763 may enhance the nucleophilicity of N2, which in turn attacks the AdoMet methyl group. The released proton is transferred to the coordinated water molecule (light blue) and further to Asp696. A direct proton transfer from Ser763 to Asp696 seems to be unlikely due to the big distance of 4.1 Å (gray dashed line). The AdoHcy molecule is shown transparent for clarity reasons and the sulfur atom is depicted as white sphere. Important hydrogen bonds are drawn by dashed lines and distances are indicated.

Although, the structural and biochemical arguments for the proposed model are very conclusive, the possibility that the dimethylation and especially the second methyl transfer reaction applies to a different mechanism cannot be fully excluded. Hence, further mutagenesis experiments as well as structure analyses with substrate analogs like a cap dinucleotide (e.g. m^7GpppG) and reaction intermediates (e.g. $m^{2,7}GTP$) have to be performed in order to verify the proposed catalytic mechanism.

8.2 The nuclear export complex CRM1·SPN1·RanGTP

The nuclear export receptor exportin 1 (Xpo1) or CRM1 (chromosome region maintenance 1) is probably the most versatile exportin in human cells as it exports hundreds of proteins which carry a leucine rich nuclear export signal (NES) (Fornerod *et al.*, 1997; Gadgil *et al.*, 2001; Hutten and Kehlenbach, 2007; Johnson *et al.*, 2002; Kutay and Guttinger, 2005; Moy and Silver, 2002; Stade *et al.*, 1997). CRM1 mediates the export of distinct ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and viral RNAs (HIV mRNA) *via* diverse adapter molecules. Besides adapter-proteins bound to RNA it also mediates the export of proteins *per se* and the prototypical example for this scenario is the snRNP import-adapter snurportin 1 (SPN1) (Ospina *et al.*, 2005; Paraskeva *et al.*, 1999). The competence of CRM1 to bind the export cargoes with high affinity is dependent on binding to the molecular switch Ran in its GTP bound form. Despite its sound biochemical and cell biological characterization the structural information of this nuclear export receptor was limited to its C-terminal third composed of the six HEAT repeats 14-19 and comprising the residues 707-1027 (Petosa *et al.*, 2004). However, the cargo-binding region including the part, which is responsible for the recognition of the nuclear export signal (NES), and the interaction site for the molecular switch RanGTP, located within the N-terminal two-thirds of the protein, is missing. In order to characterize the structural organization of the nuclear export complex composed of CRM1, SPN1 and RanGTP the complex was crystallized (Figure 8-4) and its structure was solved to a resolution of 2.5 Å. The crystal structure shows the molecular basis for cargo/NES recognition and cooperativity of cargo-RanGTP binding in the investigated complex (Chapter 6).

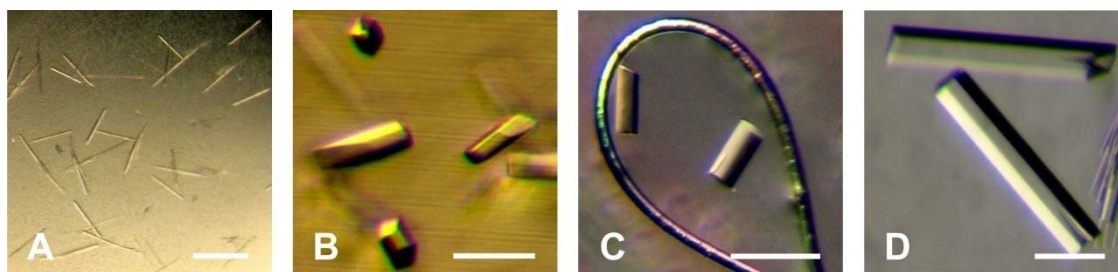


Figure 8-4: Crystals of the export complex CRM1·SPN1·RanGTP. (A) Initial, needle shaped crystals of the export complex in a PEG 4000 condition. (B) Optimized crystals in a condition containing 13% (w/v) PEG 1000 and HEPES pH 8.0. (C) Crystals of (B) mounted in a cryo loop. (D) Crystals of the export complex in 11% (w/v) PEG 1000 and 0.1 M Tris pH 8.05. These crystals typically diffracted to a resolution of 10-4 Å, depending on the cryo protectant used. Two out of more than 500 tested crystals diffracted X-rays even to a maximum resolution of 2.5 Å and they were used for structure determination. The scale bar at the bottom of each panel corresponds to a length of 50 μm.

Quite unusually, protein crystals grown out of a fresh protein preparation stored at 4 °C always led to thin needles only (Figure 8-4 A), while a protein solution, which was frozen once in liquid nitrogen and subsequently thawed, yielded big crystals with dimensions up to 50 μm \times 50 μm \times 300 μm (Figure 8-4 D). One possible reason for this effect could be that aggregated or otherwise inhomogeneous protein precipitates during this procedure and is removed by the subsequent centrifugation step. This may allow the growth of crystals with a higher degree of order and therefore of a bigger size. Finally, only two out of more than 500 tested crystals diffracted to a resolution higher than 2.6 Å (typical resolution 10-4 Å) and these crystals were used for structure determination.

Notably, even the most sophisticated HEAT repeat predictions, failed to forecast the number and exact position of the individual repeats (Petosa *et al.*, 2004). Based on structural, biochemical and *in silico* analyses, Petosa *et al.* supposed that CRM1 consists of 19 consecutive HEAT repeats, whereas in between the last two repeats an intervening linker helix changes the directionality of helix A and B of HEAT 19. The structure analysis of the export complex presented here revealed that CRM1 is composed out of 21 HEAT repeats and adopts an overall superhelical, toroid-like structure (Chapter 6). The directionality of the accordant A- and B-helices is consistent throughout the whole molecule, leading to an inversion of the positions of helix A and B in HEAT repeat 21. In contrast to all other repeats, the A-helix of HEAT 21 faces to the inner side of the CRM1 torus, while helix B faces to the outside (Chapter 6; Figure S2).

Interestingly, the export signature of SPN1 is tripartite, as it binds *via* the N-terminal amphipathic helix, the cap binding domain (CBD) and some residues of the C-terminal part of SPN1 are also found to interact with the exportin. RanGTP is enwrapped by the N-terminal part of the exportin, while the cargo SPN1 binds on the outer surface of CRM1 and makes no direct contacts to RanGTP. The so-called “acidic loop” of CRM1 in between the A- and B-helix of HEAT 9 fixes the Ran molecule like a seat belt on the N-terminal part of CRM1. It touches both switch regions of Ran, binds to them and therefore scans for the GTP-form of this molecular switch. Interestingly, the common molecular switch RanGTP is bound differently by the individual transport receptors. While the regions of Ran, which are contacted by the transport receptor are mostly similar, the detailed interactions between both proteins differ significantly (Figure 8-5). Thus differences in Ran and cargo binding which are observed in the different transport complexes may at least in part rely on the different interaction sites of the transport receptors with RanGTP.

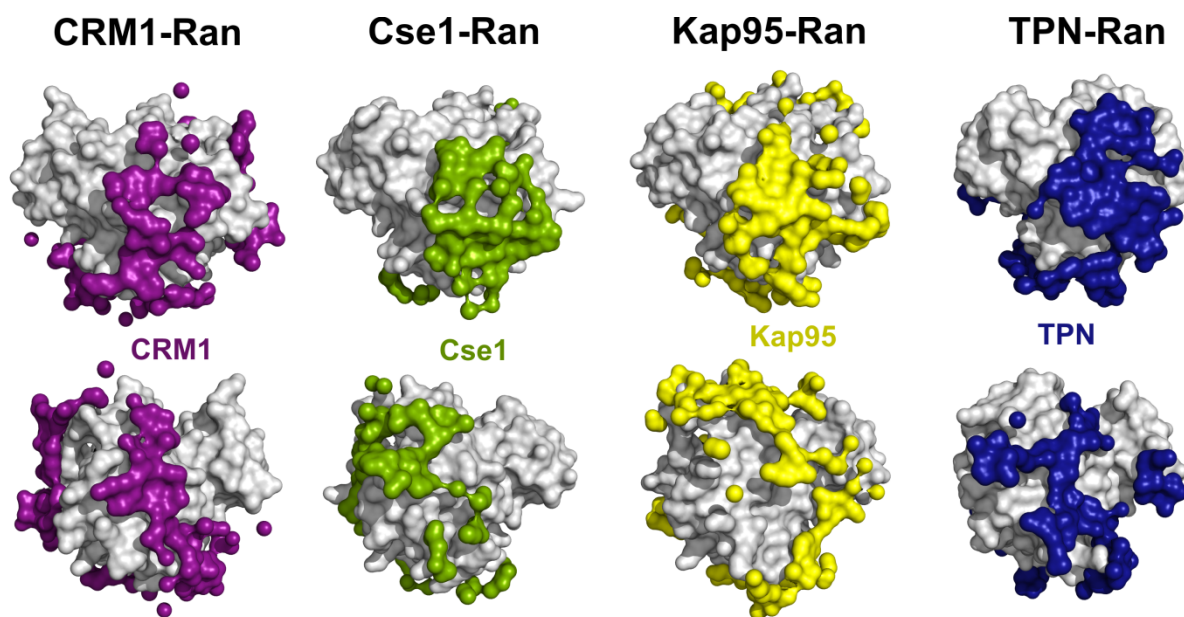


Figure 8-5: Comparison of RanGTP binding of different structurally characterized nuclear transport receptors. Ran is depicted in light gray and the area occupied by the particular transport receptor is indicated by the colored surface (purple=CRM1, PDB ID 3GJX; green=Cse1, (Matsuura and Stewart, 2004) PDB ID 1WA5; yellow=karyopherin 95, (Lee *et al.*, 2005) PDB ID 2BKU; blue=transportin, (Chook and Blobel, 1999) PDB ID 1QBK). The C-terminal residues 179-197 of TPN were removed to increase clarity. The colored surfaces represent transport receptor atoms, which occupy the RanGTP surface (AREAIMOL; radius of probe solvent=1.4Å).

The crystal structure of full length CRM1 bound to the cargo SPN1 was determined recently by another group as well (Dong *et al.*, 2009). Importantly, this incomplete assembly lacks the molecular switch RanGTP, shown to be essential for high affinity cargo binding as well as export competence by CRM1 (Chapter 6; Figure 1). Hence, it is possible that this dimeric assembly shows a certain state, which may be important during the disassembly of the export complex after hydrolysis of RanGTP to its GDP bound form in the cytoplasm. Interestingly, such a dimeric complex bound to nucleoporin 358 and independent of RanGTP is supposed to exist during the late steps of nuclear export (Hutten and Kehlenbach, 2007). Besides the fact that RanGTP is lacking in the crystal structure, the overall structures of CRM1 and Snurportin1 in both complexes are very similar as they superpose with a root mean square deviation of 2.0 Å with respect to all common C $_{\alpha}$ -atoms (Figure 8-6). In the structure of Dong *et al.* the complete first HEAT repeat as well as several loops including the important acidic loop are missing in the final model, owing to missing electron density. These portions are clearly defined in the CRM1·SPN1·RanGTP export complex and the reason for that is most likely the presence of RanGTP which binds to these regions and hence stabilizes them in a certain conformation. The second major difference concerns the C-terminal HEAT repeat of

CRM1. In our export complex structure, HEAT 21 is composed of two antiparallel helices, whose position with respect to the outer and inner side of the CRM1 torus is inverted compared to all other preceding HEAT repeats. By contrast, in the crystal structure lacking RanGTP the last α -helix protrudes through the entire central hole of the toroid touching helix 9B and the loop between helices 10B-11A on the other side of the molecule. This α -helix comprises the residues Thr1030-Val1056 and corresponds to the B-helix of HEAT 21 in our structure. Remarkably, this interaction site is located next to the acidic loop fixing RanGTP and contributing to the cooperative binding of export complex components. This fact and given that RanGTP is not present in the complex, the observed structural flexibility and difference may have a certain relevance in the physiological release of the molecular switch and possibly later in the release of the cargo as well. Interestingly, the C-terminal α -helix in the structure of Dong *et al.* would clash with the bound RanGTP molecule, when superposed as in Figure 8-6. However, whether the structural difference and flexibility of the C-terminal CRM1-helix has indeed a physiological relevance remains to be seen.

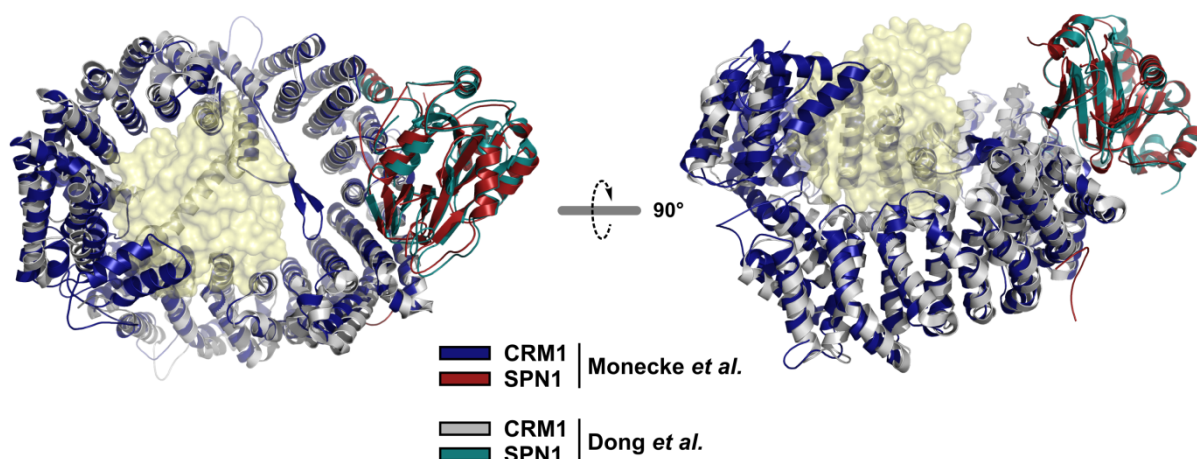


Figure 8-6: Superposition of the export complexes by Monecke *et al.* and Dong *et al.* (PDB ID 3GJX and PDB ID 3GB8, respectively). The export complex by Monecke *et al.* consists of CRM1 (blue), SPN1 (red) and RanGTP (yellow surface). By contrast, the dimeric complex by Dong *et al.* lacks RanGTP. SPN1 does not differ significantly in both complexes, while the differences in CRM1 are restricted to the N- and C-terminal parts as well as several loops. RanGTP is shown as transparent surface to allow the view on the C-terminal helix of CRM1, which in the structure of Dong *et al.*, crosses the entire central hole of the toroid.

During crystallization of the nuclear export complex a second crystal form of the designated complex was observed at 4 °C in a solution containing 10% (w/v) PEG 8000, Tris pH 7.5 and 0.2 M Mg-acetate. The crystals belonged to the orthorhombic space group $P2_12_12$ with the cell dimensions $a=64.9$ Å, $b=117.5$ Å and $c=43.9$ Å (Table 8-1). In fact, this unit cell was too small to harbor the whole export complex, which has a molecular weight of more than 180

kDa. Interestingly, a molecular replacement with the program PHASER (McCoy, 2007) using the crystallized C-terminal third of human CRM1 (PDB ID 1W9C; residues 707-1027) as search model led to a low initial R-factor of 45%. After multiple rounds of refinement to a resolution of 2.25 Å with REFMAC5 (Murshudov *et al.*, 1997) and model building in COOT (Emsley and Cowtan, 2004), the $R_{\text{work}}/R_{\text{free}}$ factors decreased to finally 23.3% and 27.2%, respectively. In the final model the residues 707-1025 of CRM1 are defined in the electron density map with one molecule present in the asymmetric unit, representing a CRM1 fragment which is nearly identical to the initial search model (Petosa *et al.*, 2004).

Table 8-1: Data statistics of the designated export complex at 4 °C.

Crystal	Crm1/RanGTP/SPN1
Data collection	
Space group	$P2_12_12$
Cell dimensions	
a, b, c (Å)	64.9, 117.5, 43.9
α , β , γ (°)	90, 90, 90
Wavelength (Å)	0.873
X-ray source	Microfocus ID23-2, ESRF (Grenoble)
Resolution range (Å)	25.00-2.25 (2.33-2.25)
No. of reflections	15293
Completeness (%)	98.0 (99.9)
R_{merge}^a (%)	6.0 (36.3)
Average I/ σ	17.2 (3.3)
Redundancy	3.3 (3.3)
Mosaicity (°)	0.35
Refinement	
Resolution (Å)	25.00-2.25
Molecules per AU	1
R_{work}^b (%)	23.3
R_{free}^c (%)	27.2
Figure of merit	0.79
RMS deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.632

Values in parentheses indicate the specific values in the particular highest resolution shell.

Both structures superpose well with a root mean square deviation of 0.6 Å and with respect to all common C_{α} -atoms (Figure 8-7). For clarity reasons the HEAT repeat assignment for the C-terminal part of CRM1 in Figure 8-7 was adopted from Petosa *et al.*, although it is obvious that numbers and positions of the individual HEATs are different in the full length CRM1 structure. How this fragment was generated in the crystallization drop remains unclear, since the resulting purified export complex was proven to consist of all three complex components

and was used for crystallization. In fact, filamentous fungi were observed in that particular condition, while crystals did not appear in similar conditions without these fungal filaments. Therefore, possibly a secreted protease of the fungi led to cleavage of an accessible loop of full length CRM1 and thus to the disassembly of the export complex allowing the growth of crystals containing exclusively the C-terminal fragment of the exportin.

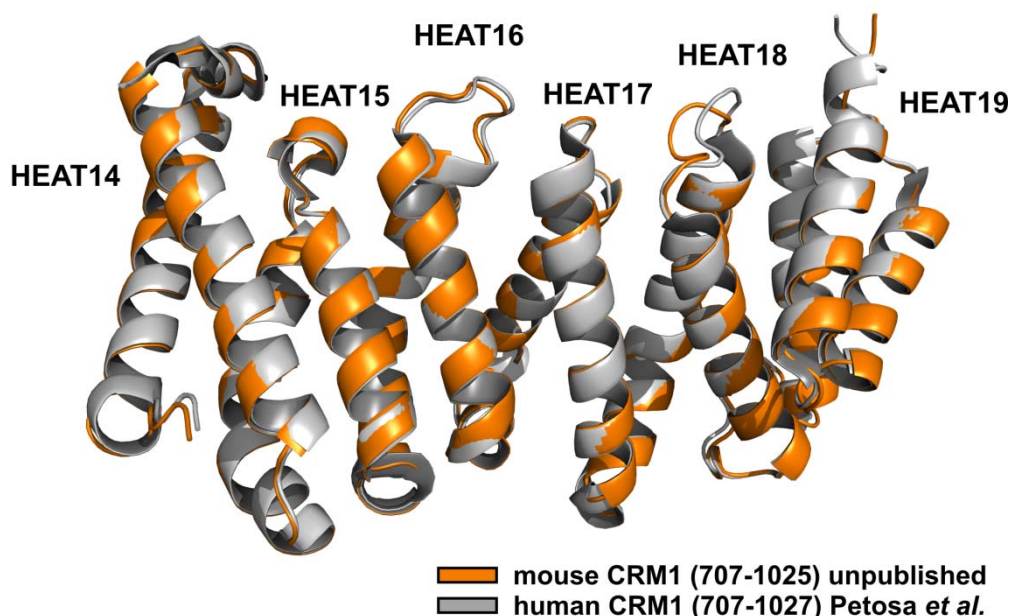


Figure 8-7: Superposition of the structures of human CRM1 (aa707-1027; (Petosa *et al.*, 2004) PDB ID 1W9C; light gray) and HEAT repeats 14-19 (aa707-1025; orange) of mouse CRM1 (unpublished data). The three dimensional structures are nearly identical and the depicted CRM1 parts of *H. sapiens* and *M. musculus* differ in 5 amino acids only at the sequence level. HEAT repeats were assigned according to Petosa *et al.* although this differs significantly from the true HEAT repeat assignment which has been made on the basis of the full length CRM1 structure.

In summary, the three dimensional crystal structure of the nuclear export complex CRM1·SPN1·RanGTP presented here and the additional biochemical data provide significant structural insights into the assembly and disassembly of nuclear export complexes. These results will be a basis for continuative experiments further specifying the details of nuclear transport in the future.

8.3 The poly(A)-specific ribonuclease

The poly(A)-specific ribonuclease (PARN) is a homodimeric and processive 3' exoribonuclease and the major deadenylase in mammalian cells (Astrom *et al.*, 1992; Copeland and Wormington, 2001; Dehlin *et al.*, 2000; Gao *et al.*, 2001; Korner and Wahle, 1997; Martinez *et al.*, 2001; Wu *et al.*, 2005). It is unique compared to other deadenylases, since it binds both the 3'-poly(A) tail as well as the 5'-m⁷G-cap of the target mRNA. In order

to characterize the unknown binding mode of the PARN RNA-recognition-motif (RRM) to the m⁷G-cap, the human PARN-RRM was purified, crystallized (Figure 8-8) and its structure was solved by means of selenomethionine-MAD (SeMet-MAD) to a resolution of 2.1 Å.

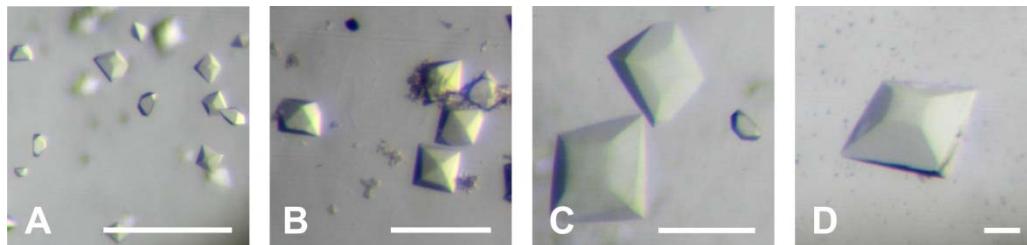


Figure 8-8: Crystals of the poly(A)-specific ribonuclease RRM domain bound to the cap analog m⁷GTP. Tetragonal crystals of PARN-RRM aa445-540-m⁷Gppp-complex (with six-fold molar excess of cap analog) grown in (A) 2 M Li₂SO₄, 10 mM MgCl₂, 50 mM MES pH 5.6 and a protein concentration of 8.8 mg/ml, (B) the same buffer but with a protein concentration of 7.4 mg/ml and (C) 6.5 mg/ml. (D) Tetragonal crystals of selenomethionine (SeMet) PARN-RRM-m⁷Gppp (7 mg/ml) in 2 M Li₂SO₄ and 50 mM MES pH 5.5, 5 mM β-mercaptoethanol and 175 mM NaBr. The structure finally was solved by means of SeMet-MAD and refined against a higher resolution native dataset. The scale bar at the bottom of each panel corresponds to a length of 50 μm.

Interestingly, the PARN-RRM domain binds the m⁷G-cap using a novel modality as the positively charged 7-methylguanosine is stacked by only one tryptophan (Trp475 in human PARN and Trp468 in the mouse ortholog). Contrary to that, in all other structurally characterized m⁷G-cap binding proteins, the cap is bound between two (aromatic or at least hydrophobic) protein residues, generating a tight stack by the contribution of π - π and cation- π interactions. The structure of the PARN-RRM or PARN in its full length was solved independently three times (Chapter 7) (Nagata *et al.*, 2008; Wu *et al.*, 2009). Figure 8-9 shows the superposition of all three available structures. The crystal structure of human PARN in its full length reveals that the RRM domain is located on the top of the nuclease domain with respect to Figure 8-9 (Chapter 3; Figure 3-12). The nuclease domain, which is interrupted by the R3H domain (aa138-245) is responsible for the nucleolytic activity of the enzyme. The latter, however, is not visible in the crystal structure owing to missing electron density for the whole domain. Nevertheless, by means of superposition with an R3H domain containing PARN structure, it was shown that the RRM domain of one monomer and the R3H domain of another may form a circular structure that encloses the active site. The RRM fold present in all three structures, superposes well (left panel in Figure 8-9), while the C-terminal 2 helices in the structure of human PARN-RRM (yellow) protrude from the core of the domain.

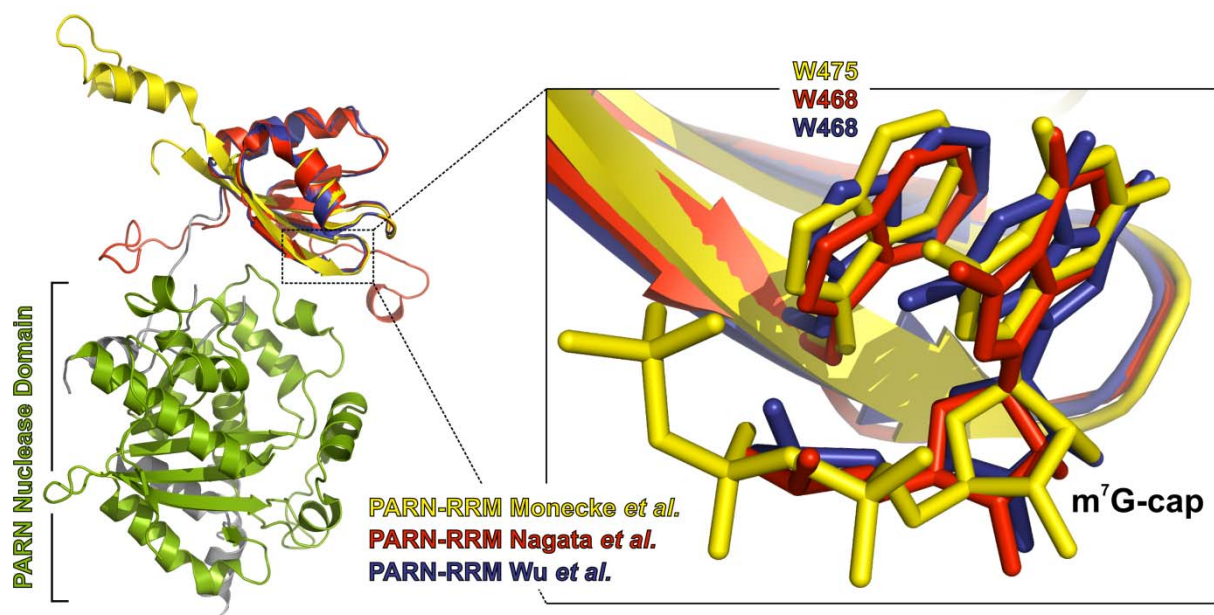


Figure 8-9: Superposition of PARN-RRM domains. The human PARN-RRM (Monecke *et al.*; PDB ID 3CTR) is depicted in yellow whereas the RRM domain of mouse PARN ((Nagata *et al.*, 2008) PDB ID 2ROK) is shown in red. The crystal structure of full length human PARN ((Wu *et al.*, 2009) PDB ID 3D45) is divided into the RRM domain (blue) and the nuclease domain (green). The R3H domain, which is inserted in the nuclease domain with respect to the primary amino acid sequence, is missing in the crystal structure due to flexibility and missing electron density. The RRM is located on the top of the nuclease domain and harbors Trp475 (human PARN) or Trp468 (mouse PARN), representing the only stacking residue for the positively charged 7-methylguanosine-cap (m⁷G). The three RRM domains superpose well, while two yellow helices at the human RRM C-terminus, protrude from the core. The corresponding two α -helices of an adjacent molecule in the asymmetric unit capture their position, thus mimicking the canonical RRM fold perfectly.

In their place, the two corresponding helices of a symmetry related molecule fill the gap and mimic the RRM fold perfectly (Chapter 7; Figure 2). Such an exchange of secondary structure elements raise the question of a physiological relevance since PARN is known to exist and act as a homodimer (Wu *et al.*, 2005). However, the fact that such an exchange is not observed in the two remaining PARN-cap complexes indicates that this interchange is rather a crystallization artifact than an activity-relevant dimerization. Moreover, the RRM domains in the full length PARN structure are located far away from each other and attached to the nuclease domain of the same molecule as well as to the R3H domain of another molecule (see above). Despite these discussed differences between the structures, the binding of the m⁷G-cap by Trp475 and 468 of the PARN-RRMs applies to the same mode (right panel in Figure 8-9). The only observed difference concerns the conformation of the 7-methylguanosine with respect to the stacking tryptophan in one monomer of the full length PARN structure, representing the more closed conformation (Wu *et al.*, 2009). The m⁷guanine in both PARN-RRM domain structures as well as in the closed conformation of full length PARN is in the *anti*-conformation. In contrast, this methylated nucleobase is rotated by 180° around the C1'-

N9 bond (representing the χ angle) in the monomer with the open conformation of the full length PARN structure and therefore it adopts the *syn*-conformation (Figure 8-9 in blue).

In order to confirm the dominant roles of PARN amino acid residues implicated in cap binding and identified by means of X-ray crystallography, dissociation constants for PARN-cap complexes were measured and calculated. A comparison of the binding affinities of individual human PARN fragments to different m⁷G-cap variations is given in Table 8-2.

Table 8-2: Comparison of dissociation constants of selected human PARN fragments and mutants for the cap variants m⁷GTP and m⁷GpppG.

human PARN fragment	K _d ± Δ K _d [μM]		source
	m ⁷ GTP	m ⁷ GpppG	
PARN	1.59 ± 0.11	0.90 ± 0.02	Nilsson <i>et al.</i>
PARN (W475A)	>1000	>1000	
PARN 443-560	11.1 ± 0.2	6.96 ± 0.14	
PARN 443-560 (W456A, W475A)	>1000	>1000	
PARN 445-540	6.94 ± 2.08	—	Monecke <i>et al.</i>
PARN 445-540 (K454A)	20.03 ± 3.43	—	
PARN 445-540 (T458A)	30.58 ± 4.29	—	
PARN 445-540 (D478A)	20.00 ± 7.03	—	

The analysis of these values indeed confirms, that Trp475 is the major contributor to cap binding. It is the only stacking residue for the m⁷G-cap and its mutation leads to a complete loss of binding capability. Interestingly, the binding of the slightly longer PARN fragment comprising the residues 443-560 to the cap analog (11.1 μM) is about two times weaker than the one for the shorter PARN 445-540 (6.94 μM) (Chapter 7) (Nilsson *et al.*, 2007). Differences in the dissociation constants of the full length PARN and the sole PARN-RRMs suggested, that additional residues, which are not part of the RNA recognition motif, contribute to cap binding. In fact, the crystal structure of full length PARN in complex with m⁷GpppG revealed that both, the RRM and the nuclease domain contribute to cap binding significantly (Wu *et al.*, 2009). At least in one of the two PARN molecules, representing the more closed conformation, various residues of the nuclease domain interact with the second guanine nucleotide of the cap analog m⁷GpppG. In detail, Ile34, Leu57, Leu283, Leu284 and Met418 from the nuclease domain form a hydrophobic pocket for the guanine and the atom OD1 of Asp28 interacts indirectly with its ribose *via* a water molecule (Wu *et al.*, 2009). In addition, there are further interactions of the second nucleoside as well as of the phosphates involving the residues Asn281, Ser335, Lys319, His280 and Leu336. In summary, all these interactions may contribute to the seven-fold higher binding affinity of full length PARN

(0.90 μM) to m^7GpppG . However, this observation does not explain the difference in the affinity of the individual PARN constructs for the m^7GTP mononucleotide, which lacks the second guanine (1.59 μM with full length PARN and 11.1 μM with PARN-RRM).

Chapter 9 • References

- Adachi, Y. and Yanagida, M. (1989) Higher order chromosome structure is affected by cold-sensitive mutations in a *Schizosaccharomyces pombe* gene *crm1+* which encodes a 115-kD protein preferentially localized in the nucleus and its periphery. *J. Cell Biol.*, **108**, 1195-1207.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948-1954.
- Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W. and Bork, P. (2001) Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.*, **309**, 1-18.
- Ares, M., Jr. and Weiser, B. (1995) Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog. Nucleic Acid Res. Mol. Biol.*, **50**, 131-159.
- Arnold, M., Nath, A., Hauber, J. and Kehlenbach, R.H. (2006a) Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *J. Biol. Chem.*, **281**, 20883-20890.
- Arnold, M., Nath, A., Wohlwend, D. and Kehlenbach, R.H. (2006b) Transportin is a major nuclear import receptor for c-Fos: a novel mode of cargo interaction. *J. Biol. Chem.*, **281**, 5492-5499.
- Arts, G.J., Fornerod, M. and Mattaj, I.W. (1998) Identification of a nuclear export receptor for tRNA. *Curr. Biol.*, **8**, 305-314.
- Astrom, J., Astrom, A. and Virtanen, A. (1992) Properties of a HeLa cell 3' exonuclease specific for degrading poly(A) tails of mammalian mRNA. *J. Biol. Chem.*, **267**, 18154-18159.
- Bauerle, M., Doenecke, D. and Albig, W. (2002) The requirement of H1 histones for a heterodimeric nuclear import receptor. *J. Biol. Chem.*, **277**, 32480-32489.
- Bayliss, R., Littlewood, T. and Stewart, M. (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. *Cell*, **102**, 99-108.
- Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W. and Medalia, O. (2004) Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science*, **306**, 1387-1390.
- Beck, M., Lucic, V., Forster, F., Baumeister, W. and Medalia, O. (2007) Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature*, **449**, 611-615.
- Bischoff, F.R., Krebber, H., Smirnova, E., Dong, W. and Ponstingl, H. (1995) Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. *EMBO J.*, **14**, 705-715.
- Bischoff, F.R. and Ponstingl, H. (1991) Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. *Nature*, **354**, 80-82.

- Bohnsack, M.T., Regener, K., Schwappach, B., Saffrich, R., Paraskeva, E., Hartmann, E. and Gorlich, D. (2002) Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *EMBO J.*, **21**, 6205-6215.
- Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117-127.
- Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.*, **36**, 333-360.
- Brunger, A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.*, **2**, 2728-2733.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999) Splicing of Precursors to mRNAs by the Spliceosomes. In *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 525-560.
- Calero, G., Wilson, K.F., Ly, T., Rios-Steiner, J.L., Clardy, J.C. and Cerione, R.A. (2002) Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat. Struct. Biol.*, **9**, 912-917.
- Chari, A., Golas, M.M., Klingenhager, M., Neuenkirchen, N., Sander, B., Englbrecht, C., Sickmann, A., Stark, H. and Fischer, U. (2008) An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell*, **135**, 497-509.
- Chari, A., Paknia, E. and Fischer, U. (2009) The role of RNP biogenesis in spinal muscular atrophy. *Curr. Opin. Cell Biol.*, **21**, Epub ahead of print.
- Chi, N.C., Adam, E.J. and Adam, S.A. (1995) Sequence and characterization of cytoplasmic nuclear protein import factor p97. *J. Cell Biol.*, **130**, 265-274.
- Chook, Y.M. and Blobel, G. (1999) Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature*, **399**, 230-237.
- Chook, Y.M. and Blobel, G. (2001) Karyopherins and nuclear import. *Curr. Opin. Struct. Biol.*, **11**, 703-715.
- Chook, Y.M., Cingolani, G., Conti, E., Stewart, M., Vetter, I. and Wittinghofer, A. (1999) Pictures in cell biology. Structures of nuclear-transport components. *Trends Cell Biol.*, **9**, 310-311.
- Cingolani, G., Petosa, C., Weis, K. and Muller, C.W. (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature*, **399**, 221-229.
- Colau, G., Thiry, M., Leduc, V., Bordonne, R. and Lafontaine, D.L. (2004) The small nucleolar RNA cap trimethyltransferase is required for ribosome synthesis and intact nucleolar morphology. *Mol. Cell. Biol.*, **24**, 7976-7986.
- Coller, J. and Parker, R. (2004) Eukaryotic mRNA decapping. *Annu. Rev. Biochem.*, **73**, 861-890.
- Conti, E. and Izaurralde, E. (2001) Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.*, **13**, 310-319.
- Cook, A., Bono, F., Jinek, M. and Conti, E. (2007) Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.*, **76**, 647-671.

- Copeland, P.R. and Wormington, M. (2001) The mechanism and regulation of deadenylation: identification and characterization of Xenopus PARN. *RNA*, **7**, 875-886.
- Coppola, J.A., Field, A.S. and Luse, D.S. (1983) Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc. Natl. Acad. Sci. U S A*, **80**, 1251-1255.
- Cougot, N., van Dijk, E., Babajko, S. and Seraphin, B. (2004) 'Cap-tabolism'. *Trends Biochem. Sci.*, **29**, 436-444.
- Damelin, M., Silver, P.A. and Corbett, A.H. (2002) Nuclear protein transport. *Methods Enzymol.*, **351**, 587-607.
- Dehlin, E., Wormington, M., Korner, C.G. and Wahle, E. (2000) Cap-dependent deadenylation of mRNA. *EMBO J.*, **19**, 1079-1086.
- Dickmanns, A., and Ficner, R. (2005) Role of the 5'-cap in the biogenesis of spliceosomal snRNPs. *Top. Curr. Genet.*, **12**, 179-204.
- Dong, X., Biswas, A., Suel, K.E., Jackson, L.K., Martinez, R., Gu, H. and Chook, Y.M. (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature*, **458**, 1136-1141.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126-2132.
- Enunlu, I., Papai, G., Cserpan, I., Udvardy, A., Jeang, K.T. and Boros, I. (2003) Different isoforms of PRIP-interacting protein with methyltransferase domain/trimethylguanosine synthase localizes to the cytoplasm and nucleus. *Biochem. Biophys. Res. Commun.*, **309**, 44-51.
- Fassati, A., Gorlich, D., Harrison, I., Zaytseva, L. and Mingot, J.M. (2003) Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. *EMBO J.*, **22**, 3675-3685.
- Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. and Luhrmann, R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*, **82**, 475-483.
- Fornerod, M. and Ohno, M. (2002) Exportin-mediated nuclear export of proteins and ribonucleoproteins. *Results Probl. Cell Differ.*, **35**, 67-91.
- Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I.W. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90**, 1051-1060.
- Franke, J., Gehlen, J. and Ehrenhofer-Murray, A.E. (2008) Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J. Cell Sci.*, **121**, 3553-3560.
- Freedman, N.D. and Yamamoto, K.R. (2004) Importin 7 and importin alpha/importin beta are nuclear import receptors for the glucocorticoid receptor. *Mol. Biol. Cell*, **15**, 2276-2286.
- Fried, H. and Kutay, U. (2003) Nucleocytoplasmic transport: taking an inventory. *Cell. Mol. Life Sci.*, **60**, 1659-1688.
- Fu, X., Choi, Y.K., Qu, D., Yu, Y., Cheung, N.S. and Qi, R.Z. (2006) Identification of nuclear import mechanisms for the neuronal Cdk5 activator. *J. Biol. Chem.*, **281**, 39014-39021.

- Fukuda, M., Gotoh, I., Adachi, M., Gotoh, Y. and Nishida, E. (1997) A novel regulatory mechanism in the mitogen-activated protein (MAP) kinase cascade. Role of nuclear export signal of MAP kinase kinase. *J. Biol. Chem.*, **272**, 32642-32648.
- Gadal, O., Strauss, D., Kessel, J., Trumpower, B., Tollervey, D. and Hurt, E. (2001) Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell. Biol.*, **21**, 3405-3415.
- Gao, M., Fritz, D.T., Ford, L.P. and Wilusz, J. (2000) Interaction between a poly(A)-specific ribonuclease and the 5' cap influences mRNA deadenylation rates in vitro. *Mol. Cell*, **5**, 479-488.
- Gao, M., Wilusz, C.J., Peltz, S.W. and Wilusz, J. (2001) A novel mRNA-decapping activity in HeLa cytoplasmic extracts is regulated by AU-rich elements. *EMBO J.*, **20**, 1134-1143.
- Garneau, N.L., Wilusz, J. and Wilusz, C.J. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell. Biol.*, **8**, 113-126.
- Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., Bertrand, E. and Bordonne, R. (2008) Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J. Biol. Chem.*, **283**, 2060-2069.
- Gontan, C., Guttler, T., Engelen, E., Demmers, J., Fornerod, M., Grosveld, F.G., Tibboel, D., Gorlich, D., Poot, R.A. and Rottier, R.J. (2009) Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. *J. Cell Biol.*, **185**, 27-34.
- Gorlich, D. and Kutay, U. (1999) Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.*, **15**, 607-660.
- Gorlich, D., Pante, N., Kutay, U., Aebi, U. and Bischoff, F.R. (1996) Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.*, **15**, 5584-5594.
- Harel, A. and Forbes, D.J. (2004) Importin beta: conducting a much larger cellular symphony. *Mol. Cell*, **16**, 319-330.
- Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M. and Luhrmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U S A*, **99**, 16719-16724.
- Hastings, M.L. and Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.*, **13**, 302-309.
- Hausmann, S., Ramirez, A., Schneider, S., Schwer, B. and Shuman, S. (2007) Biochemical and genetic analysis of RNA cap guanine-N2 methyltransferases from *Giardia lamblia* and *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **35**, 1411-1420.
- Hausmann, S. and Shuman, S. (2005) Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J. Biol. Chem.*, **280**, 4021-4024.
- Hausmann, S., Zheng, S., Costanzo, M., Brost, R.L., Garcin, D., Boone, C., Shuman, S. and Schwer, B. (2008) Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of 2,2,7-trimethylguanosine caps, small nuclear ribonucleoprotein components, pre-mRNA splicing factors, and RNA decay pathways. *J. Biol. Chem.*, **283**, 31706-31718.

- Hernandez, N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.*, **276**, 26733-26736.
- Hopper, A.K., Traglia, H.M. and Dunst, R.W. (1990) The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus. *J. Cell Biol.*, **111**, 309-321.
- Hsu, P.C., Hodel, M.R., Thomas, J.W., Taylor, L.J., Hagedorn, C.H. and Hodel, A.E. (2000) Structural requirements for the specific recognition of an m7G mRNA cap. *Biochemistry*, **39**, 13730-13736.
- 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.
- Huber, J., Dickmanns, A. and Luhrmann, R. (2002) The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. *J. Cell Biol.*, **156**, 467-479.
- Hutten, S. and Kehlenbach, R.H. (2007) CRM1-mediated nuclear export: to the pore and beyond. *Trends Cell Biol.*, **17**, 193-201.
- Isken, O. and Maquat, L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, **21**, 1833-1856.
- Jakel, S., Albig, W., Kutay, U., Bischoff, F.R., Schwamborn, K., Doenecke, D. and Gorlich, D. (1999) The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J.*, **18**, 2411-2423.
- Jakel, S. and Gorlich, D. (1998) Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.*, **17**, 4491-4502.
- Johnson, A.W., Lund, E. and Dahlberg, J. (2002) Nuclear export of ribosomal subunits. *Trends Biochem. Sci.*, **27**, 580-585.
- Jullien, D., Gorlich, D., Laemmli, U.K. and Adachi, Y. (1999) Nuclear import of RPA in *Xenopus* egg extracts requires a novel protein XRIPalpha but not importin alpha. *EMBO J.*, **18**, 4348-4358.
- Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell*, **12**, 5-14.
- Kahle, J., Baake, M., Doenecke, D. and Albig, W. (2005) Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. *Mol. Cell. Biol.*, **25**, 5339-5354.
- Kalderon, D., Richardson, W.D., Markham, A.F. and Smith, A.E. (1984) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature*, **311**, 33-38.
- Kambach, C., Walke, S. and Nagai, K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.*, **9**, 222-230.
- Kataoka, N., Bachorik, J.L. and Dreyfuss, G. (1999) Transportin-SR, a nuclear import receptor for SR proteins. *J. Cell Biol.*, **145**, 1145-1152.

- Kim, S.H. and Lin, R.J. (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.*, **16**, 6810-6819.
- Komonyi, O., Papai, G., Enunlu, I., Muratoglu, S., Pankotai, T., Kopitova, D., Maroy, P., Udvardy, A. and Boros, I. (2005) DTL, the Drosophila homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. *J. Biol. Chem.*, **280**, 12397-12404.
- Korner, C.G. and Wahle, E. (1997) Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *J. Biol. Chem.*, **272**, 10448-10456.
- Kuersten, S., Ohno, M. and Mattaj, I.W. (2001) Nucleocytoplasmic transport: Ran, beta and beyond. *Trends Cell Biol.*, **11**, 497-503.
- Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R. and Gorlich, D. (1997) Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell*, **90**, 1061-1071.
- Kutay, U. and Guttinger, S. (2005) Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biol.*, **15**, 121-124.
- Kutay, U., Lipowsky, G., Izaurralde, E., Bischoff, F.R., Schwarzmaier, P., Hartmann, E. and Gorlich, D. (1998) Identification of a tRNA-specific nuclear export receptor. *Mol. Cell*, **1**, 359-369.
- la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K. and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.*, **17**, 527-536.
- Lai, M.C., Lin, R.I., Huang, S.Y., Tsai, C.W. and Tarn, W.Y. (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J. Biol. Chem.*, **275**, 7950-7957.
- Lanford, R.E. and Butel, J.S. (1984) Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell*, **37**, 801-813.
- Lanford, R.E., Kanda, P. and Kennedy, R.C. (1986) Induction of nuclear transport with a synthetic peptide homologous to the SV40 T antigen transport signal. *Cell*, **46**, 575-582.
- Lee, S.J., Matsuura, Y., Liu, S.M. and Stewart, M. (2005) Structural basis for nuclear import complex dissociation by RanGTP. *Nature*, **435**, 693-696.
- Lei, E.P. and Silver, P.A. (2002) Protein and RNA export from the nucleus. *Dev. Cell*, **2**, 261-272.
- Lejeune, F., Li, X. and Maquat, L.E. (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylation, and exonucleolytic activities. *Mol. Cell*, **12**, 675-687.
- Lesser, C.F. and Guthrie, C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982-1988.
- Lipowsky, G., Bischoff, F.R., Schwarzmaier, P., Kraft, R., Kostka, S., Hartmann, E., Kutay, U. and Gorlich, D. (2000) Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. *EMBO J.*, **19**, 4362-4371.
- Macara, I.G. (2001) Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.*, **65**, 570-594.

- Maco, B., Fahrenkrog, B., Huang, N.P. and Aebi, U. (2006) Nuclear pore complex structure and plasticity revealed by electron and atomic force microscopy. *Methods Mol. Biol.*, **322**, 273-288.
- Madhani, H.D. and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803-817.
- Madhani, H.D. and Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1-26.
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M. and Luhrmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205-2208.
- Malca, H., Shomron, N. and Ast, G. (2003) The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol. Cell. Biol.*, **23**, 3442-3455.
- Maquat, L.E. and Carmichael, G.G. (2001) Quality control of mRNA function. *Cell*, **104**, 173-176.
- Martinez, J., Ren, Y.G., Nilsson, P., Ehrenberg, M. and Virtanen, A. (2001) The mRNA cap structure stimulates rate of poly(A) removal and amplifies processivity of degradation. *J. Biol. Chem.*, **276**, 27923-27929.
- 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.
- Matsuura, Y. and Stewart, M. (2004) Structural basis for the assembly of a nuclear export complex. *Nature*, **432**, 872-877.
- Mattaj, I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905-911.
- Mattaj, I.W. and Englmeier, L. (1998) Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.*, **67**, 265-306.
- Maxwell, E.S. and Fournier, M.J. (1995) The small nucleolar RNAs. *Annu. Rev. Biochem.*, **64**, 897-934.
- Mazza, C., Segref, A., Mattaj, I.W. and Cusack, S. (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.*, **21**, 5548-5557.
- McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 32-41.
- Mingot, J.M., Kostka, S., Kraft, R., Hartmann, E. and Gorlich, D. (2001) Importin 13: a novel mediator of nuclear import and export. *EMBO J.*, **20**, 3685-3694.
- Misra, P., Qi, C., Yu, S., Shah, S.H., Cao, W.Q., Rao, M.S., Thimmapaya, B., Zhu, Y. and Reddy, J.K. (2002) Interaction of PIMT with transcriptional coactivators CBP, p300, and PBP differential role in transcriptional regulation. *J. Biol. Chem.*, **277**, 20011-20019.
- Mitchell, P. and Tollervey, D. (2001) mRNA turnover. *Curr. Opin. Cell Biol.*, **13**, 320-325.

- Mitrousis, G., Olia, A.S., Walker-Kopp, N. and Cingolani, G. (2008) Molecular basis for the recognition of snurportin 1 by importin beta. *J. Biol. Chem.*, **283**, 7877-7884.
- Moore, J.D., Yang, J., Truant, R. and Kornbluth, S. (1999) Nuclear import of Cdk/cyclin complexes: identification of distinct mechanisms for import of Cdk2/cyclin E and Cdc2/cyclin B1. *J. Cell Biol.*, **144**, 213-224.
- Mouaikel, J., Bujnicki, J.M., Tazi, J. and Bordonne, R. (2003a) Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res.*, **31**, 4899-4909.
- Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J. and Bordonne, R. (2003b) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep*, **4**, 616-622.
- 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.
- Moy, T.I. and Silver, P.A. (2002) Requirements for the nuclear export of the small ribosomal subunit. *J. Cell Sci.*, **115**, 2985-2995.
- Muhlhauser, P., Muller, E.C., Otto, A. and Kutay, U. (2001) Multiple pathways contribute to nuclear import of core histones. *EMBO Rep*, **2**, 690-696.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240-255.
- Nagata, T., Suzuki, S., Endo, R., Shirouzu, M., Terada, T., Inoue, M., Kigawa, T., Kobayashi, N., Guntert, P., Tanaka, A., Hayashizaki, Y., Muto, Y. and Yokoyama, S. (2008) The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Res.*, **36**, 4754-4767.
- Nemergut, M.E., Mizzen, C.A., Stukenberg, T., Allis, C.D. and Macara, I.G. (2001) Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. *Science*, **292**, 1540-1543.
- Nilsson, P., Henriksson, N., Niedzwiecka, A., Balatsos, N.A., Kokkoris, K., Eriksson, J. and Virtanen, A. (2007) A multifunctional RNA recognition motif in poly(A)-specific ribonuclease with cap and poly(A) binding properties. *J. Biol. Chem.*, **282**, 32902-32911.
- Ohno, M., Kataoka, N. and Shimura, Y. (1990) A nuclear cap binding protein from HeLa cells. *Nucleic Acids Res.*, **18**, 6989-6995.
- 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.
- Ospina, J.K., Gonsalvez, G.B., Bednenko, J., Darzynkiewicz, E., Gerace, L. and Matera, A.G. (2005) Cross-talk between snurportin1 subdomains. *Mol. Biol. Cell*, **16**, 4660-4671.
- Ossareh-Nazari, B., Bachelier, F. and Dargemont, C. (1997) Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science*, **278**, 141-144.

- Paine, P.L., Moore, L.C. and Horowitz, S.B. (1975) Nuclear envelope permeability. *Nature*, **254**, 109-114.
- 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.
- Paraskeva, E., Izaurralde, E., Bischoff, F.R., Huber, J., Kutay, U., Hartmann, E., Luhrmann, R. and Gorlich, D. (1999) CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J. Cell Biol.*, **145**, 255-264.
- Parker, R. and Song, H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.*, **11**, 121-127.
- Peters, R. (2006) Introduction to nucleocytoplasmic transport: molecules and mechanisms. *Methods Mol. Biol.*, **322**, 235-258.
- Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., Soler-Lopez, M., Baudin, F., Mattaj, I.W. and Muller, C.W. (2004) Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol. Cell*, **16**, 761-775.
- Plafker, S.M. and Macara, I.G. (2000) Importin-11, a nuclear import receptor for the ubiquitin-conjugating enzyme, UbcM2. *EMBO J.*, **19**, 5502-5513.
- 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.
- Pollard, V.W., Michael, W.M., Nakielnny, S., Siomi, M.C., Wang, F. and Dreyfuss, G. (1996) A novel receptor-mediated nuclear protein import pathway. *Cell*, **86**, 985-994.
- Quiocho, F.A., Hu, G. and Gershon, P.D. (2000) Structural basis of mRNA cap recognition by proteins. *Curr. Opin. Struct. Biol.*, **10**, 78-86.
- 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.
- Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell*, **64**, 615-623.
- Ruan, J.P., Ullu, E. and Tschudi, C. (2007) Characterization of the Trypanosoma brucei cap hypermethylase Tgs1. *Mol. Biochem. Parasitol.*, **155**, 66-69.
- Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J.E., Jr. (1980) The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell*, **19**, 69-78.
- Schubert, H.L., Blumenthal, R.M. and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.*, **28**, 329-335.
- Schwartz, T.U. (2005) Modularity within the architecture of the nuclear pore complex. *Curr. Opin. Struct. Biol.*, **15**, 221-226.

- 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.
- Seraphin, B. and Rosbash, M. (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, **59**, 349-358.
- Shatkin, A.J. (1976) Capping of eucaryotic mRNAs. *Cell*, **9**, 645-653.
- Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112-122.
- Shuman, S. (2002) What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.*, **3**, 619-625.
- Siomi, M.C., Eder, P.S., Kataoka, N., Wan, L., Liu, Q. and Dreyfuss, G. (1997) Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. *J. Cell Biol.*, **138**, 1181-1192.
- Sleeman, J.E., Ajuh, P. and Lamond, A.I. (2001) snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *J. Cell Sci.*, **114**, 4407-4419.
- Sleeman, J.E. and Lamond, A.I. (1999) Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.*, **9**, 1065-1074.
- Stade, K., Ford, C.S., Guthrie, C. and Weis, K. (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell*, **90**, 1041-1050.
- Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315-326.
- Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D. and Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell*, **9**, 31-44.
- Stewart, M. (2007) Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell Biol.*, **8**, 195-208.
- Strasser, A., Dickmanns, A., Luhrmann, R. and Ficner, R. (2005) Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J.*, **24**, 2235-2243.
- Strom, A.C. and Weis, K. (2001) Importin-beta-like nuclear transport receptors. *Genome Biol.*, **2**, reviews3008.1-3008.9.
- Suntharalingam, M. and Wente, S.R. (2003) Peering through the pore: nuclear pore complex structure, assembly, and function. *Dev. Cell*, **4**, 775-789.
- Takizawa, C.G., Weis, K. and Morgan, D.O. (1999) Ran-independent nuclear import of cyclin B1-Cdc2 by importin beta. *Proc. Natl. Acad. Sci. U S A*, **96**, 7938-7943.
- Tarn, W.Y. and Steitz, J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev.*, **8**, 2704-2717.
- Tiganis, T., Flint, A.J., Adam, S.A. and Tonks, N.K. (1997) Association of the T-cell protein tyrosine phosphatase with nuclear import factor p97. *J. Biol. Chem.*, **272**, 21548-21557.

- Tran, E.J. and Wente, S.R. (2006) Dynamic nuclear pore complexes: life on the edge. *Cell*, **125**, 1041-1053.
- Truant, R. and Cullen, B.R. (1999) The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol. Cell. Biol.*, **19**, 1210-1217.
- Vetter, I.R., Arndt, A., Kutay, U., Gorlich, D. and Wittinghofer, A. (1999) Structural view of the Ran-Importin beta interaction at 2.3 Å resolution. *Cell*, **97**, 635-646.
- Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701-718.
- Weis, K. (2002) Nucleocytoplasmic transport: cargo trafficking across the border. *Curr. Opin. Cell Biol.*, **14**, 328-335.
- Weis, K. (2003) Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell*, **112**, 441-451.
- Wen, W., Meinkoth, J.L., Tsien, R.Y. and Taylor, S.S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*, **82**, 463-473.
- Will, C.L., and Luhrmann, R. (2006) Spliceosome Structure and Function. In *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 369-400.
- Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290-301.
- Wohlwend, D., Strasser, A., Dickmanns, A. and Ficner, R. (2007) Structural basis for RanGTP independent entry of spliceosomal U snRNPs into the nucleus. *J. Mol. Biol.*, **374**, 1129-1138.
- Wu, M., Nilsson, P., Henriksson, N., Niedzwiecka, A., Lim, M.K., Cheng, Z., Kokkoris, K., Virtanen, A. and Song, H. (2009) Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease. *Structure*, **17**, 276-286.
- Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E. and Song, H. (2005) Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.*, **24**, 4082-4093.
- Xiao, Z., Liu, X. and Lodish, H.F. (2000) Importin beta mediates nuclear translocation of Smad 3. *J. Biol. Chem.*, **275**, 23425-23428.
- Yean, S.L. and Lin, R.J. (1991) U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction. *Mol. Cell. Biol.*, **11**, 5571-5577.
- Zhu, Y., Qi, C., Cao, W.Q., Yeldandi, A.V., Rao, M.S. and Reddy, J.K. (2001) Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc. Natl. Acad. Sci. U S A*, **98**, 10380-10385.

Appendix I • Abbreviations

Å	Ångström
aa	amino acid
AdoHcy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
AL	acidic loop
ATP	adenosine triphosphate
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
bp	base pairs
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CAS	cellular apoptosis susceptibility
CBC	cap binding complex
CBD	cap binding domain
CBP	CREB-binding protein
CBP20	cap binding protein 20
CBP80	cap binding protein 80
CCR4	carbon catabolite repressor protein 4
CRM1	chromosome region maintenance 1
Cse1	chromosome segregation 1
CV	column volume
d	day
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
Da	Dalton (1g/mol)
Dcp1	decapping protein 1
Dcp2	decapping protein 2
DcpS	scavenger decapping enzyme
DEDD	abbreviation for the four amino acids Asp, Glu, Asp, Asp
DESY	Deutsches Elektronen-Synchrotron
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTL	drosophila-tat-like
DTT	dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
eIF4E	eucaryotic initiation factor 4E
ESRF	european synchrotron radiation facility
F _{calc}	calculated structure factor
F _{obs}	observed structure factor
G	guanine
<i>G. lamblia</i>	<i>Giardia lamblia</i>
GSH	reduced glutathion
GST	glutathione-S-transferase
GTP	guanosine triphosphate
h	hour
<i>H. sapiens</i>	<i>Homo sapiens</i>
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A and TOR
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography

IBB	importin β binding domain
IgG	immunoglobulin G
Imp	importin
IPTG	isopropyl β -D-thiogalactopyranoside
Kap	karyopherin
K _d	dissociation constant
KsgA	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethyltransferase
LB	lysogeny broth
LMB	leptomycin B
M	molar
m	meter
<i>M. janaschii</i>	<i>Methanococcus janaschii</i>
<i>M. musculus</i>	<i>Mus musculus</i>
m ^{2,2,7} ₍₃₎ GTP	N2, N2-dimethyl, N7-monomethyl guanosine triphosphate
m ⁷ GTP	N7-monomethyl guanosine triphosphate
MAD	multiple anomalous dispersion/diffraction
MES	2-(N-morpholino) ethanesulfonic acid
MR	molecular replacement
mRNA	messenger RNA
MTase	methyltransferase
MW	molecular weight
MWCO	molecular weight cut off
NES	nuclear export signal
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NPC	nuclear pore complex
NTE	N-terminal extension
NTF2	nuclear transport factor 2
OD	optical density
<i>P. horikoshii</i>	<i>Pyrococcus horikoshii</i>
p300	adenovirus E1A-binding protein p300
PAGE	polyacrylamide gel electrophoresis
Pan2/3	poly(A)-specific ribonuclease subunit homolog isoform 2/3
PARN	poly(A)-specific ribonuclease
PBP	PPAR-binding protein
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PHAX	phosphorylated adapter of RNA export
PIMT	PRIP-interacting protein with methyltransferase domain
PKI	protein kinase inhibitor
PPAR	peroxisome proliferator-activated receptor
PRIP	peroxisome proliferator-activated receptor interacting protein
R3H	RNA binding motif
Ran	ras-related nuclear antigen
Rev	regulator of virion
RMSD	root mean square deviation
RNA	ribonucleic acid
RNP	ribonucleoprotein particle
RRM	RNA recognition motif
rRNA	ribosomal RNA
RsmC	ribosomal RNA small subunit methyltransferase C
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>

<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
SDS	sodium dodecyl sulfate
SeMet	selenomethionine
SMN	survival of motor neuron
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SPN1	snurportin 1
<i>T. brucei</i>	<i>Trypanosoma brucei</i>
TAR	trans-activation response element
TEV	tobacco etch virus
TGS1	Trimethylguanosine Synthase 1
TLC1	telomerase component 1
TRIS	tris (hydroxymethyl) aminomethane
Trm1	tRNA methyltransferase 1
tRNA	transfer RNA
UsnRNP	uridyl-rich small nuclear ribonucleoprotein particle
v/v	volume-volume percentage
VP39	viral protein 39
w/v	mass-volume percentage
<i>X. laevis</i>	<i>Xenopus laevis</i>
Xpo1	exportin 1
XRN1	exoribonuclease 1
λ	wavelength

Appendix II • Danksagungen

An diesem Punkt der Arbeit möchte ich gern den Leuten zu danken, die maßgeblich zum guten Gelingen dieser Arbeit beigetragen haben.

Zunächst und vor allem möchte ich mich bei Herrn **Prof. Dr. Ralf Ficner** bedanken, der als Initiator aller Projekte diese gesamte Arbeit erst ermöglicht hat. Für sein mir entgegengebrachtes großes Vertrauen und seine kontinuierliche Unterstützung bin ich ihm außerordentlich dankbar.

Herrn **Prof. Dr. Holger Stark** bin ich für die Übernahme des Korreferates und die damit verbundene Begutachtung der Arbeit sehr dankbar.

Dr. Achim Dickmanns gebührt besonderer Dank. Er erwies sich stets als Helfer in der Not, außerordentlich ideenreicher Diskussionspartner, sowie unermüdlicher Motivator nach wissenschaftlichen Katastrophen. Als „Kopf“ der Transportgruppe und stets kritischer Beobachter aller Experimente hat er maßgeblich zum Erfolg aller Teilprojekte dieser Arbeit beigetragen.

Allen Mitgliedern des gegenwärtigen und ehemaligen „Transport-Labors“ bin ich außerordentlich dankbar für eine sehr angenehme Arbeitsatmosphäre. Besonders **Dr. Anja Strasser**, die durch die hervorragende Betreuung meiner Diplomarbeit den Grundstein für die Promotion legte und **Dr. Daniel Wohlwend** sowie **Jens Brinkmann** möchte ich hier ganz herzlich danken. Danke, nicht nur für wissenschaftliche Diskussionen und zahlreichen Hilfen, sondern auch für die unzählig vielen schönen Abende und wunderbaren Freundschaften.

Ein ganz besonderer Dank gebührt auch **Dr. Denis Kudlinzki**, der mich über das komplette Studium vom vierten Tag an, kontinuierlich begleitet hat. Herzlichen Dank nicht nur für viele gemeinsame schöne Abende sondern auch für unzählige, professionelle Tipps und Rat, wenn es mal wieder lichterloh brannte.

Dr. Piotr Neumann bin ich ausdrücklich für das rechtzeitige Erscheinen in der Abteilung und die damit verbundenen Lösungen großer Probleme sehr dankbar.

Dr. Kristina Lakomek sowie **Eike Schulz** danke ich für das Korrekturlesen der Arbeit und zusammen mit **Andreas Schmitt** für viele schöne und unterhaltsame Kaffeepausen.

Bernhard Kuhle, **Stephanie Schell** und **Tolga Soykan** haben im Rahmen von sehr erfolgreichen Laborpraktika ihre Qualifikation und Leidenschaft unter Beweis gestellt und ebenfalls maßgeblich zum Gelingen der Arbeit beigetragen. Stephanie Schell sei hier außerdem für ihr außergewöhnliches Engagement im Rahmen ihrer Diplomarbeit gedankt.

Schließlich bin ich jedem einzelnen gegenwärtigen sowie früheren Mitglied der Abteilung für Molekulare Strukturbiologie sehr dankbar für eine wunderbare leise und laute Arbeitsatmosphäre und viele, viele hilfreiche Tipps und Tricks.

Herrn **Prof. Dr. Dirk Görlich** sowie **Thomas Güttler** bin ich für eine sehr fruchtbare Kooperation dankbar, die sich nicht zuletzt in der zusammen erreichten Publikation widerspiegelt.

Mein bester Dank gilt an dieser Stelle auch Herrn **Wolfgang Benisch**, ohne dessen Ermutigung ich den universitären Weg des Biologiestudiums wohl nie gegangen wäre. Als mein Biologielehrer wusste er schon damals, dass ein Biologiestudium ohne fundierte Chemiekenntnisse unmöglich ist, glücklicherweise hat er es mir jedoch stets verschwiegen.

Ebenfalls ganz besonders möchte ich mich bei meinen Eltern, **Gundolf** und **Waltraud Monecke** und meiner gesamten Familie bedanken. Die kontinuierliche Unterstützung, die nicht nur finanzieller sondern auch moralischer und seelischer Natur war, ist und bleibt undenkbar wichtig.

Ein ganz besonderer Dank an ganz besonderer Stelle gebührt **Mareicke**, die mich seit nunmehr fast drei Jahren begleitet. Danke für das Verständnis, wenn ich zu ungünstigen Zeitpunkten an den unmöglichsten Orten sein musste. Sie hat dabei sicherlich mehr als jeder und jede andere alle Höhen und Tiefen intensiv miterlebt und trotzdem nie aufgehört mich immer wieder aufzubauen. Für ihre Unterstützung in allen Phasen der Arbeit bin ich ihr unendlich dankbar.

Appendix III • Curriculum Vitae

Thomas Monecke
Otto-Wels-Weg 1a
37077 Göttingen
E-mail tmoneck@uni-goettingen.de

Geburtsdatum	19.11.1979
Geburtsort	Leinefelde
Familienstand	ledig
Staatsangehörigkeit	deutsch

Schulbildung

1986-1992	Polytechnische Oberschule IV Leinefelde
1992-1998	Staatliches Gymnasium Worbis
07/1998	Abitur

Wehrdienst

1998-1999	Grundwehrdienst (1. Panzergrenadierbataillon 381)
-----------	---

Studium

1999-2005	Biologiestudium (Georg-August-Universität Göttingen)
02/2005	Diplom (Georg-August-Universität Göttingen)

Promotion

seit 03/2005	Promotion (Georg-August-Universität Göttingen)
--------------	--
