

# Molecular Determinants of Snurportin 1 Ligand Affinity and Structural Response upon Binding

Maik Goette,<sup>†</sup> Martin C. Stumpe,<sup>†</sup> Ralf Ficner,<sup>‡</sup> and Helmut Grubmüller<sup>†\*</sup>

<sup>†</sup>Theoretical and Computational Biophysics Department, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany; and <sup>‡</sup>Abteilung für Molekulare Strukturbioogie, Institut für Mikrobiologie und Genetik and GZMB, Georg-August-Universität Göttingen, Göttingen, Germany

**ABSTRACT** The transport of large biomolecules such as proteins and RNA across nuclear pore complexes is a field of strong interest and research. Although the basic mechanisms are fairly well understood, the details of the underlying intermolecular interaction within these transport complexes are still unclear. The recognition dynamics and energetics of cargo binding to the transport receptor are not yet resolved. Here, the binding of dimethylated RNA-caps to snurportin 1 is studied by molecular-dynamics simulations. The simulations reveal a strong structural response of the protein upon RNA-cap release. In particular, major rearrangements occur in regions already intrinsically flexible in the holo structure. Additionally, the difference in free energy of binding to snurportin 1 between the two methylation states of the RNA-cap, responsible for the directionality of the transport is quantified. In particular, desolvation of the ligand is revealed as the key-step in binding to snurportin 1. These findings suggest that the binding of m<sub>3</sub>G-capped RNA is mainly driven by the enhanced water entropy gain of the solvation shell.

## INTRODUCTION

Transporting macromolecules between cellular compartments is one of the major differences between eukaryotes and prokaryotes. The transport of molecules across nuclear pore complexes is required for the proper regulation and function of eukaryotic cells. The underlying transport processes have been intensively investigated and are currently an area of particularly active research (1,2).

The spliceosome is a large machinery consisting of proteins and small nuclear RNA molecules, the snRNAs. It removes noncoding sequences, i.e., introns, from pre-mRNA while fusing the exon sequences required for proper translation in the nucleus. The spliceosome itself is formed by several ribonucleoprotein subunits called uridine-rich small nuclear ribonucleoproteins (UsnRNPs). The biogenesis of UsnRNPs requires a cytoplasmic maturation step (3) (see Fig. 1). The UsnRNA is recognized by the 5'-end and the export complex consists of the phosphorylated adaptor for RNA export (PHAX), the export receptor chromosome region maintenance-1 (CRM1), the GTP-bound form of the GTPase Ran, and the cap-binding complex, which recognizes and binds 7-methyl-guanosine(m<sup>7</sup>G)-capped RNA.

In the cytoplasm, the export complex is released, and after assembly of the seven Sm proteins the m<sup>7</sup>G cap is hypermethylated by TGS1 to a 2,2,7-trimethyl-guanosine (m<sub>3</sub>G) cap. This hypermethylation triggers the reimport of the pre-

UsnRNP. The import complex consists of the survival motor neuron (SMN) complex acting as an adaptor to the actual receptor importin  $\beta$  and snurportin 1 (SPN1), specifically binding to m<sub>3</sub>G-capped RNA (2).

Recent experiments (4), as well as the crystal structure of snurportin 1 (5) (Fig. 2 C), suggest that the 2,2,7-trimethyl-guanosine-cap dinucleotide (m<sub>3</sub>GpppG) is sufficient to prevent the binding of m<sub>3</sub>G-capped UsnRNA to snurportin 1 with a similar affinity, and therefore is an intrinsic inhibitor candidate. From the molecular-dynamics (MD) point of view, it is an interesting model system for investigating the effects of hypermethylation of the m<sup>7</sup>G-capped RNA on binding to snurportin 1.

Interestingly, the m<sub>3</sub>G cap binds better to snurportin 1 than the mono-methylated m<sup>7</sup>G cap (6). Despite the ability of an amino group to function as both donor and acceptor for hydrogen bonding, the binding affinity of m<sup>7</sup>GpppG (Fig. 2 B) to SPN1 could not be measured accurately, in contrast to m<sub>3</sub>GpppG (Fig. 2 A). Strasser et al. (5) suggested the entropic penalty of the watershell near the free ligand to be the driving force of ligand-binding in the case of snurportin 1. The effective shielding of the hypermethylated guanosine-cap by a tryptophane residue of the protein is therefore a plausible explanation for the observed behavior of ligand binding and is supported by mutation experiments (5). Due to the low binding affinity of m<sup>7</sup>GpppG to snurportin 1, however, no crystal structure of this complex could be obtained yet, thus we still lack the definite proof for this hypothesis on the structural level. As for the m<sup>7</sup>GpppG/snurportin 1 complex, no crystal structure is available for the ligand-free SPN1 either. The failure to obtain crystals was attributed to an effect of the dinucleotide on the structural integrity of the protein, required for stable interactions in crystal packing. Additionally, an unusual highly twisted conformation of the  $\beta$ -strand 1 (Fig. 2 C) containing the cap-shielding tryptophane residue

Submitted November 3, 2008, and accepted for publication April 27, 2009.

\*Correspondence: hgrubmu@gwdg.de

Martin C. Stumpe's present address is Stanford University, 318 Campus Drive West, Stanford, CA 94305.

This is an Open Access article distributed under the terms of the Creative Commons-Attribution Noncommercial License (<http://creativecommons.org/licenses/by-nc/2.0/>), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editor: Alexandre M. J. J. Bonvin.

© 2009 by the Biophysical Society

0006-3495/09/07/0581/9 \$2.00

doi: 10.1016/j.bpj.2009.04.049

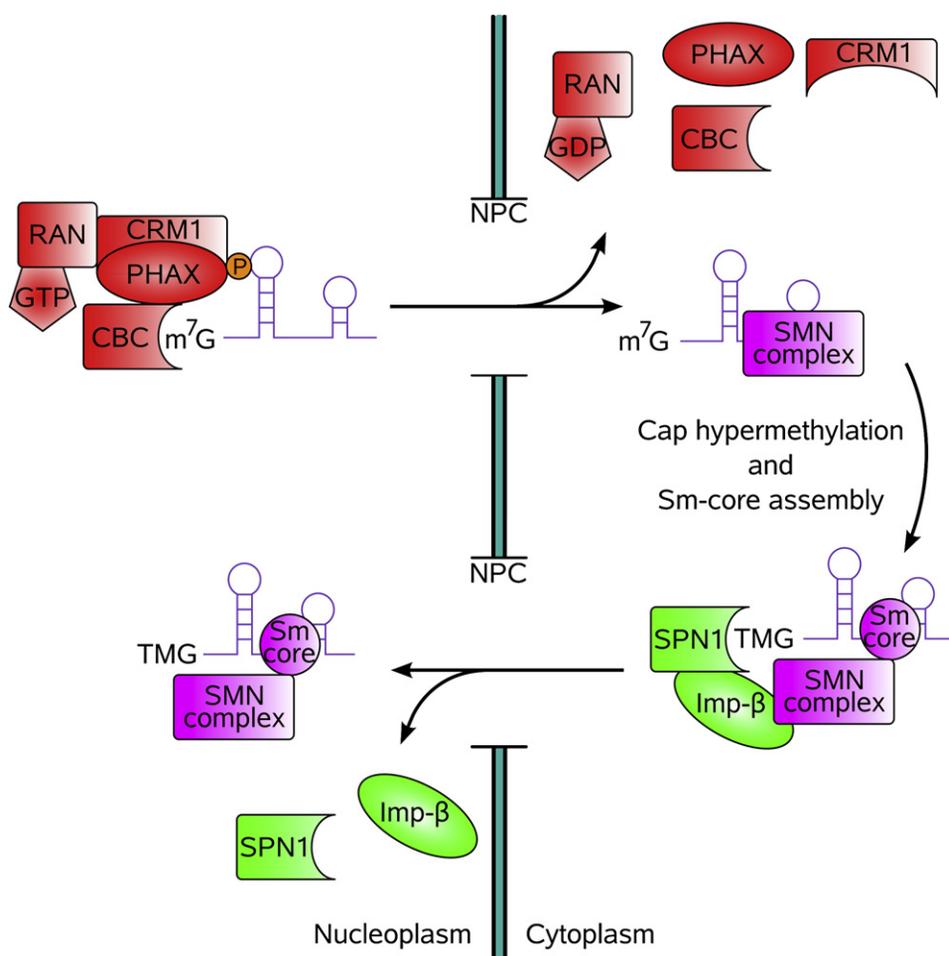


FIGURE 1 Basic nucleocytoplasmic shuttling cycle of snRNA. In the nucleus, an export complex (red) is formed, which binds to the  $m^7G$ -capped snRNA and transports it through the nuclear pore complex (NPC). After dissociation, Sm-core assembly, and cap hypermethylation, the import complex (purple and green) is formed by association of the UsnRNP complex with importin- $\beta$  and snurportin 1 (SPN1), which binds the  $m^7G$ -capped snRNA and the survival motor neuron complex. After transport into the nucleus, the import complex dissociates, and the UsnRNP is fully assembled.

was observed, which supports the assumption of an enhanced dynamics of snurportin 1 in the absence of ligands (5). Insights into the dynamics of the protein upon ligand release would be very interesting and helpful for the understanding of these issues, and would aid in crystallization attempts.

In this work, the dynamics of SPN1 upon ligand release was investigated with MD simulations. To gain insights into the dynamics and overall structural changes of the ligand-free snurportin 1, we computed multiple trajectories of the protein in the absence of a ligand. From this trajectory, the overall global motions as well as the dynamics of several specific amino acids in the binding pocket and the C-terminal region of SPN1 were investigated in more detail. Analysis of the watershell near the two methyl groups either in solvent or when bound to SPN1 was intended to help us gain insight into the contribution of the protein as a shielding factor of water from the ligand. We estimate the difference in binding free energy of  $m_3GpppG$  and  $m^7GpppG$  together with the enthalpic contributions to obtain evidence whether the binding process is driven either enthalpically or entropically.

## METHODS

We used the snurportin 1 structure 1XK5 (5) from the Protein Data Bank (PDB) (7) (Fig. 2 C) as starting structure and the AMBER99 force field

(8,9) for our MD simulations. Due to the lack of force-field parameters for  $m_3GpppG$  (resolved in the crystal structure) and  $m^7GpppG$  (not resolved in the crystal structure), we used the standard AMBER99 values for guanosine and ribose. The  $m_3G$ - and  $m^7G$ -nucleoside parameters were taken from Aduri et al. (10), and the parameters for triphosphate, connecting the two nucleotides, from Meagher et al. (11). Additionally, we scaled the charges of the molecule at the connecting phosphates such that the resulting net charge was  $-2$ .

The electrostatics in our simulations was treated with a cutoff of 1.0 nm for short-range Coulomb as well as Lennard-Jones and with particle-mesh Ewald with a grid spacing of 0.12 nm and an interpolation order of 4 for the long-range electrostatics (12). All simulations were carried out in an  $NpT$  ensemble using Berendsen pressure and temperature coupling (13) at 1 bar with a pressure coupling coefficient of  $\tau_p = 1$  ps and at 300 K with a coupling coefficient of  $\tau_T = 0.1$  ps. A 2-fs time step was used while constraining all bond lengths with the LINCS algorithm (14). All simulations were carried out in explicit solvent with the TIP4P water model (15) and a 150 mM NaCl salt-concentration to mimic physiological conditions. All simulations were carried out with the GROMACS software-package (Ver. 3) (16). We performed MD simulations of  $m_3GpppG$  bound to snurportin 1 with a total length of 650 ns, as well as ten 50-ns simulations of  $m^7GpppG$ . Additionally, 100 ns of each ligand in solvent and six trajectories of the SPN1 structure without ligand with varying length of 634, 640, 527, 641, 557, and 551 ns for the simulations 1–6 were computed.

To obtain information about the overall stability and the changes in amino-acid mobility of the protein between the ligand-bound and -unbound systems, we calculated the root mean-square deviation (RMSD) of the ligand-free protein, as well as with  $m_3GpppG$  and  $m^7GpppG$  bound to snurportin 1 along the respective trajectories. Furthermore, the backbone RMSD of every single amino acid was calculated to characterize relaxation motions upon ligand removal.

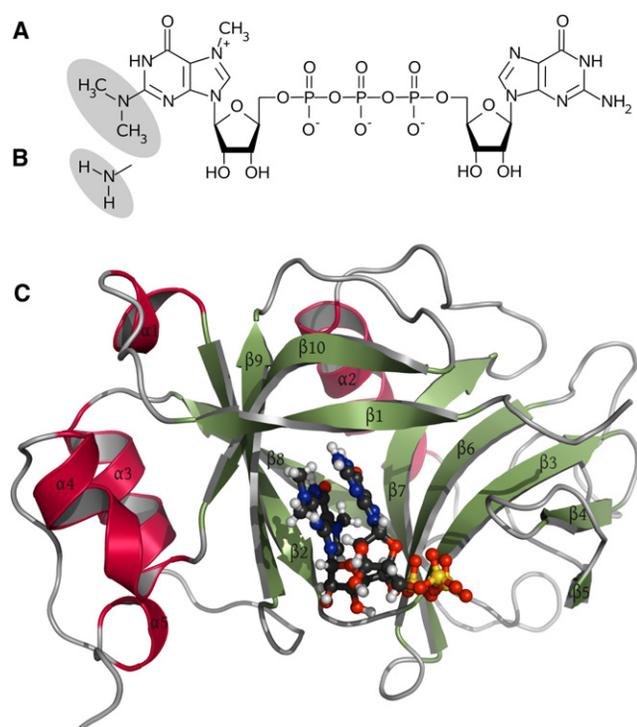


FIGURE 2 Chemical and crystal structures. (A)  $m_3$ GpppG (state A), (B)  $m^7$ GpppG (state B). The difference of both molecules is shown in the gray ellipsoids, where A represents the N2-nitrogen in the hypermethylated and B in the nonmethylated state. (C) Human snurportin 1 with bound  $m_3$ GpppG ligand (PDB ID:1XK5). The  $\alpha$ -helices are colored in red,  $\beta$ -sheets in green and loop-regions are colored in gray.  $m_3$ GpppG is shown in a ball-and-stick model.

Quantification of the structural changes of the ligand-free structure of snurportin 1 was performed by principal component analysis (PCA) (17) on the 650-ns equilibration trajectory of the snurportin 1-complex structure as well as on the trajectories 1, 2, and 6 of the protein without ligand. The PCA was carried out over the backbone atoms, and frames separated by 200 ps were used for the averages. Because of the high and presumably functionally irrelevant fluctuations of the truncated termini, 10 residues from both the N- and C-terminus were excluded from the PCA. All four trajectories mentioned above were subjected to one common PCA.

Furthermore, the distribution of water molecules in the vicinity of  $m_3$ G and  $m^7$ G, which is the only chemical difference in the ligand molecules, was analyzed in solution and when bound to SPN1. Therefore, we extracted the water molecules from the trajectories of both ligands in pure solvent (100 ns each) and in solvated protein environment (500 ns for state A and 500 ns for state B) in a sphere with a radius of 1 nm around the N2-atom

of the mono- and the trimethylated guanine-nucleoside (see Fig. 2). The density distribution of water molecules was obtained using a three-dimensional grid, consisting of 100 bins in each dimension that was laid upon the spatial coordinates of the oxygen atoms of the water molecules and smoothed with a three-dimensional Gaussian function of 0.01-nm width, which was chosen to trade-off resolution and statistical noise (18).

To compute the binding free energy difference, we equilibrated the original crystal structure with the  $m_3$ GpppG ligand (state A) for 50 ns, placed the  $m^7$ GpppG (state B) in the equilibrated protein, and let the system equilibrate for another 5 ns. We calculated the binding free energy differences as well as their statistical uncertainty using a new method, Crooks Gaussian Intersection (see Fig. 3 A) (19), which rests on fast growth thermodynamic integration (FGTI) simulations,

$$W_\tau = \int_0^1 \frac{\delta H_\lambda}{\delta \lambda} d\lambda, \quad (1)$$

where  $W_\tau$  is the work, computed for the switching process from state A ( $\lambda = 0$ ) to state B ( $\lambda = 1$ ), and  $\lambda$  is the coupling parameter which switches the system during a simulation of length  $\tau$  from state A to state B (defined by Hamiltonians  $H_A$  and  $H_B$ , respectively), i.e., via  $H_\lambda = (1 - \lambda)H_A + \lambda H_B$ .

In the free energy calculations,  $m_3$ GpppG represents the ligand in the A-state ( $\lambda = 0$ ) and  $m^7$ GpppG in the B-state ( $\lambda = 1$ ). The last 100 ps from each of the four trajectories ( $m_3$ GpppG and  $m^7$ GpppG in solvent and bound to the protein) were taken to generate four times 50 statistically independent snapshots, which were subsequently used in the starting structures for the FGTI-runs for the switching process from state A to B ( $\lambda_{0 \rightarrow 1}$ ) and B to A ( $\lambda_{1 \rightarrow 0}$ ). Statistical independence was assessed via an autocorrelation analysis of  $dH/d\lambda$ , which yielded an autocorrelation time well below 100 ps/50 = 2 ps.

The resulting free energy differences  $\Delta G$  of  $m_3$ GpppG and  $m^7$ GpppG either bound to SPN1 or in solvent were used in a thermodynamic cycle (20,21) (see Fig. 3 B) to compute the difference in binding free energy ( $\Delta\Delta G$ ) for the two ligands to SPN1.

The question whether the binding of  $m_3$ GpppG is mainly driven either enthalpically or entropically, was addressed by estimating  $\Delta H$  from the average total energies derived from the equilibrium simulations of  $m_3$ GpppG and  $m^7$ GpppG in pure solvent and bound to SPN1. From this estimate, together with the Gibbs' free energy ( $\Delta G$ ) from the free energy calculations, the entropic contribution ( $T\Delta S$ ) was estimated. An error estimate for  $\Delta H$  was obtained by computing the standard error via block averaging of ten 50-ns blocks and ten 10-ns blocks from the computed trajectories of the complexes and ligands, respectively.

## RESULTS

### Root mean-square deviations

The computed equilibrium trajectories contain information about the dynamics of SPN1 either in the ligand-bound or in the ligand-free state. To obtain quantitative information about

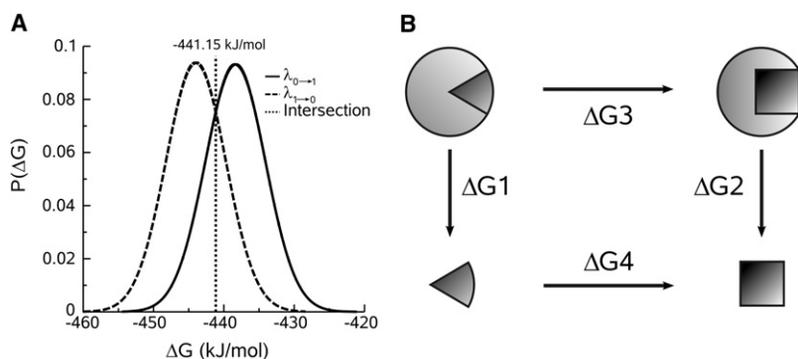


FIGURE 3 (A) Gaussian work distribution of 50 calculated work values for morphing  $m_3$ GpppG into  $m^7$ GpppG ( $\lambda_{0 \rightarrow 1}$ ) and  $m^7$ GpppG into  $m_3$ GpppG ( $\lambda_{1 \rightarrow 0}$ ). The intersection point of both Gaussians describes the value with zero dissipative work. (B) Thermodynamic cycle.  $\Delta G_3$  and  $\Delta G_4$  are calculated via thermodynamic integration simulations.  $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$ .

the structural changes of SPN1 either bound to  $m_3$ GpppG or without a ligand bound, the RMSD was calculated for the ligand-free and the  $m_3$ GpppG-bound trajectories (Fig. 4 A). After the usual fast increase within the first few nanoseconds due to thermal fluctuations, the RMSD of the ligand-bound and of five independent ligand-free trajectories stays  $<3$  Å.

Interestingly, as can be seen in Fig. 4 A, in one trajectory (blue curve), the system rapidly escapes from the initial minimum toward a different minimum with an RMSD of 4 Å, whereas in another trajectory (red curve), the system stays in the initial minimum for 300 ns. This raised the question of whether SPN1 changes its overall conformation or if the increase in RMSD is related to small areas within SPN1. To identify the regions in SPN1 that are mainly involved in the destabilization motions upon ligand removal, the backbone RMSD for each amino acid was calculated. Fig. 4, B and C, shows the time-resolved backbone RMSD for each amino acid in the structure of snurportin 1 bound to  $m_3$ GpppG and without ligand. Since the termini exhibit an intrinsically high RMSD, 10 amino acids from each terminus were excluded from this analysis. An improved statistics was obtained by averaging the RMSD values from the six ligand free trajectories.

Surprisingly, as can be seen in Fig. 4, B and C, only a few local regions contribute markedly to the observed structural changes. Fig. 5 highlights in color these regions in the structure. In the protein without a ligand, they are much more pronounced, but in similar regions.

To test whether these structural changes correlate with flexible regions observed already in the ligand-bound state, root mean-square fluctuations (RMSF) of the amino acids in the ligand-bound state were compared with the structural change (RMSD) induced by ligand removal. Although the RMSF profiles of ligand-bound and ligand-free structures are rather similar (Fig. 4, B and C), no significant correlation was seen for the structural change (data not shown).

The largest destabilization motions are seen in the C-terminal domain (Fig. 5, blue). Indeed, closer inspection of the trajectories reveals a structural rearrangement in this part of the protein upon ligand removal. Furthermore, a region of the  $\beta$ 10-strand and adjacent loops significantly rearrange (Fig. 5, red). One further region of enhanced mobility is a solvent-exposed loop built up from residues 161–167, shown in green in Fig. 5. In contrast to the other, less-stable regions, which show increased motions upon ligand removal, this loop shows a mobility similar to that in the ligand-bound structure. Furthermore, a small loop region, containing Lys<sup>144</sup> (Fig. 5, yellow), exhibits a larger structural change upon removal of the ligand. In the bound state, Lys<sup>144</sup> interacts with the phosphate backbone of  $m_3$ GpppG via a salt bridge.

### Principal component analysis

Principal component analysis (PCA) was used to compare the global motions of ligand-free and  $m_3$ GpppG-bound snurportin 1 in a common subspace. Fig. 6 shows the projection

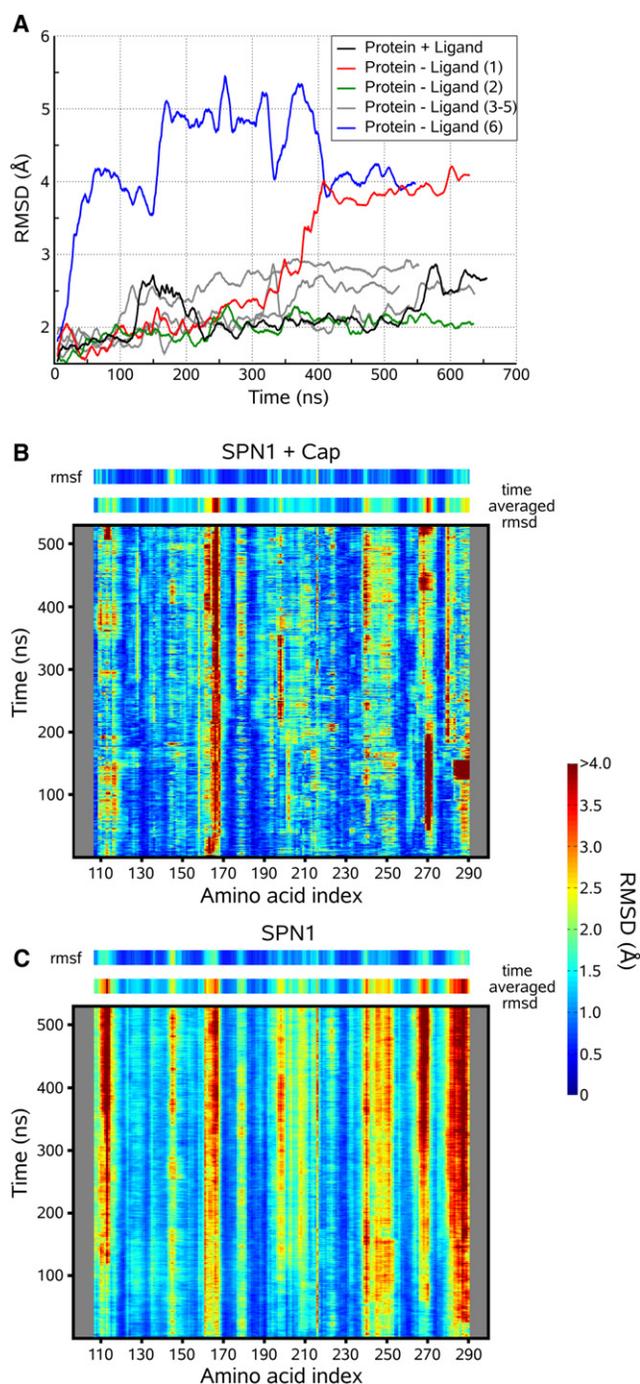


FIGURE 4 Root mean-square deviations (RMSD). (A) Backbone RMSD of SPN1 with and without  $m_3$ GpppG as ligand. The black curve denotes the RMSD of ligand-bound SPN1. The remaining curves represent the RMSDs of six independent trajectories of snurportin 1 in absence of the ligand. The trajectories with the highest RMSD are shown in red and blue; the most stable trajectory is shown in green. (B) Time-resolved amino acid RMSDs of SPN1 bound to  $m_3$ GpppG-cap. (C) The ligand free protein, where the RMSD from all six trajectories was averaged. The first 10 amino acids from each terminus have been removed in this analysis. For a better resolution in the lower RMSD regions, all values above 4 Å have been truncated to this value. RMSF and time-averaged RMSD are shown for comparison above the respective plots.

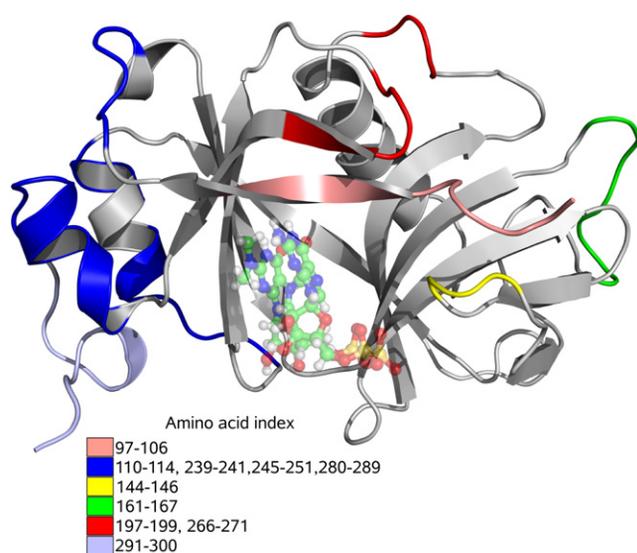


FIGURE 5 Color-coded structure of SNP1. Selected high RMSD amino acids from the ligand free SNP1 trajectories are colored according to their position in the structure. (Blue) C-terminal domain; (red)  $\beta$  10-related region; (green) loop region; (yellow) Lys<sup>144</sup>-loop; (Pink) N-terminus; (light blue) C-terminus; and (transparent) m<sub>3</sub>GpppG (for guidance).

of four trajectories onto eigenvectors 1 and 2 of this subspace. As can be seen (Fig. 6, inset), these eigenvectors describe already 49% of the atomic motion.

The system with m<sub>3</sub>GpppG bound to SNP1 (black cloud) remains close to the x-ray structure (yellow square), with rare transient transitions to an adjacent shallow minimum. In contrast, removal of the ligand from the original structure leads to an extensive sampling of phase space until the

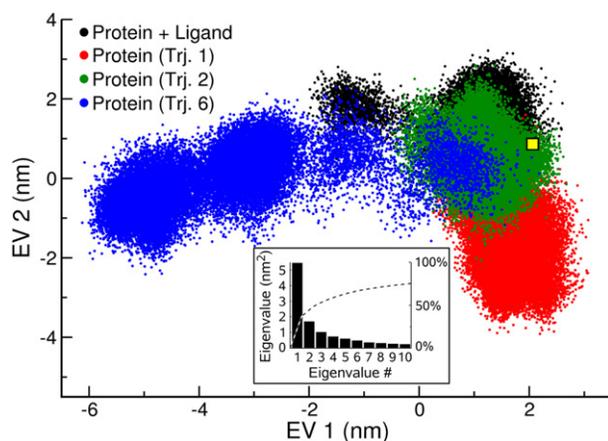


FIGURE 6 PCA of snurportin 1. The black cloud represents the 650-ns trajectory of the protein with m<sub>3</sub>GpppG ligand bound, projected onto the first two eigenvectors, and the yellow square as the general starting configuration. The clouds colored in red, green, and blue display the trajectories 1, 2, and 6 of the protein without ligand as in Fig. 4 A. Every 10th frame of the respective trajectories has been used in the projections. (Inset) The first 10 eigenvalues of the covariance matrix. The dashed line is the cumulative sum of the contribution to the total fluctuations. The first two eigenvectors describe 49% of the main global motion.

system reaches different local minima on the energy landscape (red and blue dots in Fig. 6). In particular, as can be seen from the blue cloud (Fig. 6), the sampling of the system is sufficient to visit two distinct minima, separated by a small energy barrier. This drift motion was observed for two out of six trajectories. For a better overview, only one trajectory of the remaining four, which differed only slightly, is shown (green dots in Fig. 6).

The projection of the ligand-free trajectory of SPN1 onto its first two principal components was used to select structures for detailed analysis. The largest motions were seen for trajectory 6 (blue cloud), which, therefore, was chosen for closer analysis. Accordingly, two further snapshots from the trajectory were chosen. These snapshots have been selected because they are close to the center of the respective substate.

The amino acids in direct interaction with the m<sub>3</sub>GpppG ligand were investigated first. In the bound state  $t = 0$  ns, the N7-methyl-group of the m<sub>3</sub>G-nucleobase is buried in a hydrophobic pocket, built by the residues Cys<sup>124</sup>, Ile<sup>175</sup>, Leu<sup>186</sup>, and Leu<sup>264</sup>. After removal of the ligand at 79.7 ns, this hydrophobic pocket is exposed to the surrounding water. Trp<sup>276</sup> moves toward the pocket, undergoing a local hydrophobic collapse. Additionally, the mobility of Lys<sup>144</sup> increases due to the lack of the ligand as interaction partner (Fig. 4, B and C, and Fig. 5). By moving closer to Asp<sup>173</sup>, Lys<sup>144</sup> weakens the ionic interaction between Arg<sup>129</sup> and Asp<sup>173</sup>. As a consequence, Arg<sup>129</sup> can detach from Asp<sup>173</sup> and form a new,  $\pi$ -stacking interaction with Trp<sup>276</sup>. This structural rearrangement is supported by a motion of Trp<sup>276</sup> into the binding pocket (Fig. 7 A).

Further rearrangement is seen for the N-terminal  $\beta$ -strand  $\beta$ 1 (Fig. 7 A, purple). After removal of the ligand from the binding pocket, a relaxation of this originally twisted  $\beta$ -strand is observed. Trp<sup>107</sup> moves into the now unoccupied binding pocket, loosening the strain on  $\beta$ -strand  $\beta$ 1. This movement leads to a loss of structural stability of the  $\beta$ -sheet ( $\beta$ -strands  $\beta$ 1 and  $\beta$ 10), resulting in a further distortion of  $\beta$ 10. This finding confirms the suggestion by Strasser et al. (5) that this  $\beta$ -sheet should untwist upon ligand removal.

After 148 ns, no major structural changes are seen near the binding pocket. One exception is Lys<sup>144</sup>, which moves to a purely solvent interacting position, which leads to a loss of the salt bridge between Lys<sup>144</sup> and Asp<sup>173</sup>. This is compensated by reformation of the ionic interaction of Arg<sup>129</sup> and Asp<sup>173</sup>, after moving away from its former cation- $\pi$ -interaction-partner Trp<sup>276</sup>. In summary, a highly dynamic structure near the empty binding pocket is seen.

A second, quite flexible region is located near the C-terminus (Fig. 5, blue, and Fig. 7 B). The previously described movement of Trp<sup>107</sup> and Trp<sup>276</sup> toward the binding pocket coincides with a shift of the hydrophobic residues Val<sup>111</sup>, Leu<sup>115</sup>, Val<sup>282</sup>, Val<sup>285</sup>, and Leu<sup>286</sup>. These motions destabilize the hydrophobic region, which connects the C-terminal part to the rest of the protein, resulting in a higher

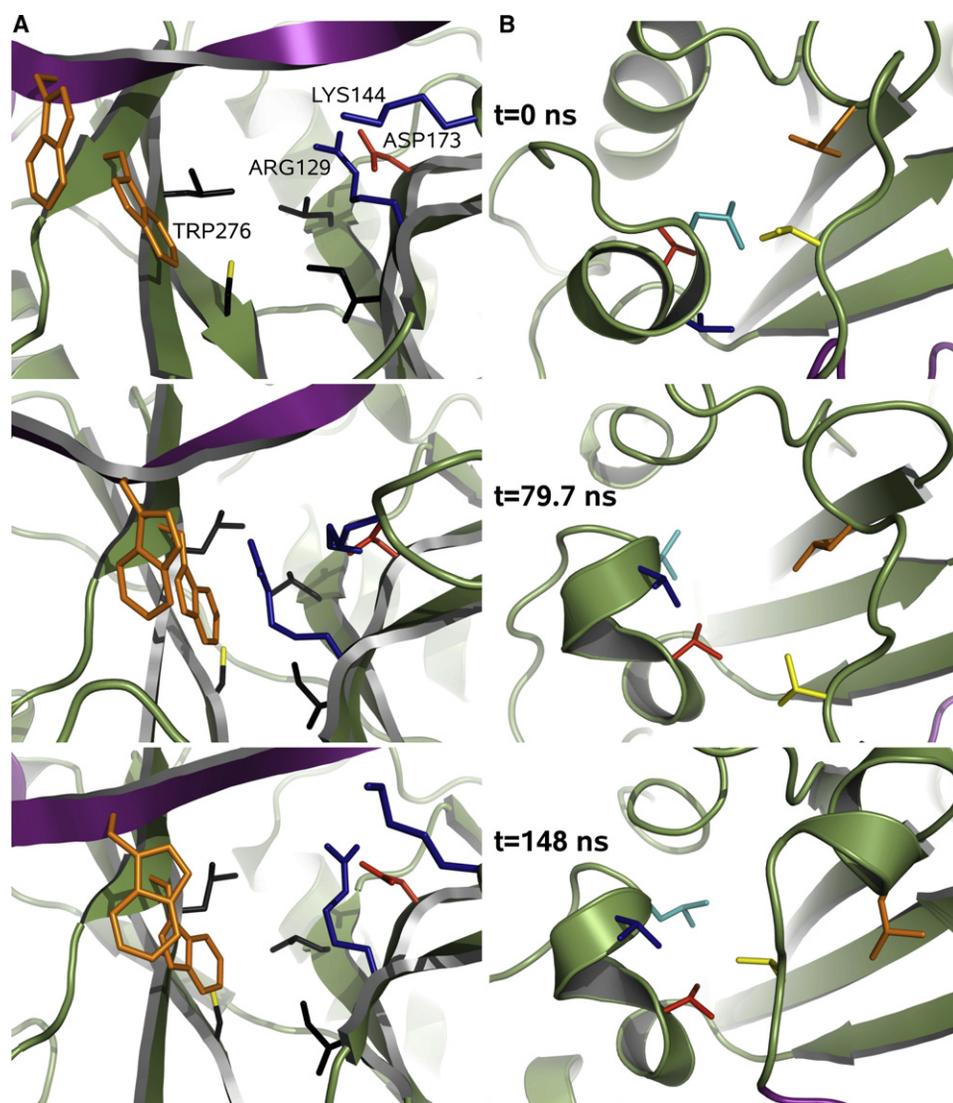


FIGURE 7 Snapshots of the binding pocket and C-terminus of SPN1 at 0, 79.7, and 148 ns. (A) Trp<sup>107</sup> and Trp<sup>276</sup> are shown in orange; the residues Cys<sup>124</sup>, Ile<sup>175</sup>, Leu<sup>186</sup>, and Leu<sup>258</sup>, building a hydrophobic pocket, in black; Arg<sup>129</sup> and Lys<sup>144</sup> in blue; and Asp<sup>173</sup> in red. The  $\beta$ -strand 1 is colored purple. (B) Val<sup>111</sup> (yellow), Leu<sup>115</sup> (orange), Val<sup>282</sup> (red), Val<sup>285</sup> (blue), and Leu<sup>286</sup> (light blue).

flexibility of the C-terminal region. Remarkably, an  $\alpha$ -helical structure of residues 113–118 is formed, followed by a reorientation of Leu<sup>115</sup>, turning away from the hydrophobic interaction region. After this transient rearrangement, the initial hydrophobic cluster is reformed, albeit with reduced stability.

### Water shell

The results so far reveal structural rearrangements of SPN1 but do not explain the unexpected strong binding of the hypermethylated m<sub>3</sub>GpppG cap. It has been suggested (5) that the entropic penalty of the watershell near the free ligand is the driving force for binding to SPN1. To test this hypothesis, we compared the solvation shell around the two ligands in solvent and bound to SPN1.

Upon binding of m<sub>3</sub>GpppG to SPN1, the volume of the solvation shell is significantly reduced (Fig. 8, A and B). This release of water molecules from the shell to the bulk is entropically favorable. m<sup>7</sup>GpppG (Fig. 8, C and D) also

exhibits a decrease of the volume of the solvation shell upon binding, but to a markedly smaller extent than for m<sub>3</sub>GpppG. Accordingly, the entropy gain due to water release is larger for m<sub>3</sub>GpppG than for m<sup>7</sup>GpppG, rendering m<sub>3</sub>GpppG binding more favorable.

Moreover, one hydrogen bond of the N2 amino group of m<sup>7</sup>GpppG to water molecules is lost upon binding (Fig. 8 D). This is also evident from the density plots, which show two high-density peaks for m<sup>7</sup>GpppG in solvent (Fig. 8 C), but only one for m<sup>7</sup>GpppG bound to SPN1 (Fig. 8 D). These high-density peaks indicate the positions of the hydrogen-bonded water molecules. In contrast to free m<sup>7</sup>GpppG, m<sup>7</sup>GpppG bound to the protein exhibits only one stable hydrogen bond with water molecules, indicated by the high-density peak in Fig. 8 D. The second hydrogen bond is formed with the protein, which is less stable and, therefore, entails a small enthalpic loss upon ligand binding.

Both effects—the larger entropy gain due to water release and the enthalpic loss of m<sup>7</sup>GpppG binding—combine to

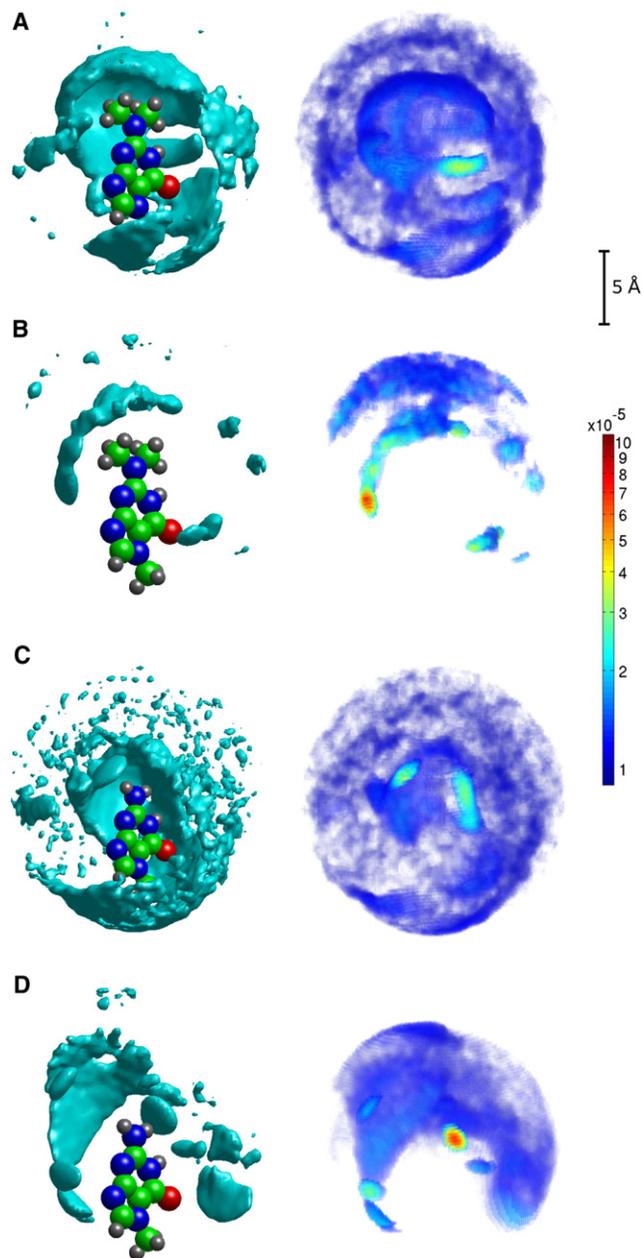


FIGURE 8 Three-dimensional density distribution of water molecules around the  $m_3$ GpppG and  $m_7$ GpppG ligands. (A and B) The  $m_3$ G-nucleoside in water and complexed to SPN1; (C and D) the  $m_7$ G-nucleoside. (Left column) Fifteen-percent of the total solvation shell, starting from the highest to lower densities, as a qualitative isosurface view. (Right column) Sixty-percent of the density in a quantitative plot. For clarity of presentation, a logarithmic color scale was chosen.

explain the unexpectedly strong binding of  $m_3$ GpppG as opposed to  $m_7$ GpppG. This result is in line with the recent mutation experiments (5), which have shown that the binding affinity of SPN1 to  $m_3$ GpppG is significantly reduced when Trp<sup>107</sup> is mutated into Ala<sup>107</sup>. This finding confirms the crucial role of Trp<sup>107</sup> as key residue for the desolvation of the dimethylated N2-nitrogen of  $m_3$ GpppG. Structural rearrangements of Trp<sup>107</sup> should therefore be

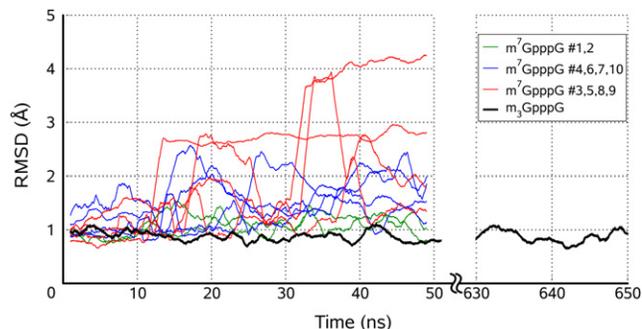


FIGURE 9 Root mean-square deviations of Trp<sup>107</sup>. The RMSD curves were smoothed with a running average, where 100 data points were used for averaging. The coloring indicates the deviation of Trp<sup>107</sup> in trajectories with  $m^7$ GpppG, where green represents a small, blue a medium, and red a strong deviation. Black shows the deviation of Trp<sup>107</sup> with  $m_3$ GpppG.

observable for bound  $m^7$ GpppG where shielding is less pronounced.

We tested this prediction by comparing the RMSD of Trp<sup>107</sup> from ten 50-ns trajectories of  $m^7$ GpppG bound to SPN1 with the RMSD of the trajectory of  $m_3$ GpppG bound to SPN1 (Fig. 9). Indeed, 8 out of 10 show a significant increase of the RMSD (blue and red), with four of these being very large increases (red). These large deviations thus indicate the loss of the hydrophobic interaction between the ligand and Trp<sup>107</sup>.

Interestingly, the RMSF of Trp<sup>107</sup> seems to be unaffected by the ligand type ( $m_3$ GpppG: 0.3 Å,  $m^7$ GpppG:  $0.4 \pm 0.1$  Å).

### Binding thermodynamics

In an attempt to confirm that the observed effects in the simulations actually cause the selective SPN1 binding affinity, we have calculated the binding free energy difference for the two ligands  $m_3$ GpppG and  $m^7$ GpppG (Table 1). We calculated via Crooks Gaussian Intersection simulations the free energy differences between  $m_3$ GpppG and  $m^7$ GpppG bound to the protein ( $\Delta G_b$ ) and in solvent ( $\Delta G_u$ ), respectively (as described in Methods), yielding  $\Delta G_b = -430.3 \pm 0.9$  kJ/mol and  $\Delta G_u = -441.2 \pm 0.5$  kJ/mol, respectively. Using the depicted thermodynamic cycle (Fig. 3 B), a binding free energy difference  $\Delta \Delta G = 10.8 \pm 1.0$  kJ/mol is obtained. Using the known dissociation constant (5) for  $m_3$ GpppG/SPN1,  $K_D = 1.00 \pm 0.03$   $\mu$ M, this free energy difference translates into  $K_D = 72\text{--}152$   $\mu$ M for  $m^7$ GpppG/SPN1. This value is

TABLE 1 Enthalpy, free energy, and entropic contribution to ligand binding and respective differences

System	$\Delta H$	$\sigma$	$\Delta G$	$\sigma$	$T\Delta S$	$\sigma$
Bound	418.7	23.3	-430.3	0.9	849.0	23.3
Unbound	464.0	15.9	-441.2	0.5	905.2	15.9
Diff. ( $\Delta\Delta$ )	-45.3	28.2	10.9	1.0	-56.2	28.2

All units are in kJ/mol, temperature  $T$  is at 300 K, and  $\sigma$  is the standard error. Differences refer to the two ligands, i.e.,  $m_3$ GpppG- $m^7$ GpppG.

consistent with the estimate of 100–1000  $\mu\text{M}$  derived from ultraviolet cross-linking studies (6).

With this result at hand, the simulations serve to dissect the binding free energies into the corresponding enthalpic and entropic contributions. The enthalpic contributions  $\Delta H$  were estimated from the averaged total energies of the respective simulations, and the entropic contributions  $T\Delta S$  from  $T\Delta S = \Delta H - \Delta G$ . The obtained  $\Delta\Delta H = -45.3 \pm 28.2$  kJ/mol and  $\Delta(T\Delta S) = -56.2 \pm 28.2$  kJ/mol (Table 1) suggest that the entropic contribution to the binding affinity difference between the two ligands is slightly larger than the enthalpic contribution within the obtained accuracy.

## DISCUSSION

Several open questions concerning SPN1 have been addressed in this work. The  $m_3\text{G}$ -cap-binding domain of human SPN1 bound to the inhibitor  $m_3\text{GpppG}$  was chosen to investigate the effects of RNA-cap hypermethylation on binding to SPN1. Extended molecular dynamics simulations of SPN1 were performed for the ligand-free protein as well as for the protein with bound ligands  $m_3\text{GpppG}$  and  $m^7\text{GpppG}$ .

PCA, RMSD, and RMSF calculations were carried out on the trajectories with  $m_3\text{GpppG}$  bound to SPN1, and with the ligand-free structure to reveal the ligand-dependent structural changes of the systems. Comparison of the complexes with simulations of the solvated two ligands showed that the solvation shell plays a crucial role for binding selectivity. In addition,  $\text{Trp}^{107}$  was shown to be crucial for binding.

Furthermore, our simulations served to study possible structural changes upon ligand removal for the apo protein SPN1, whereas the complex structure SPN1/ $m_3\text{GpppG}$  remained stable. Remarkably, a large fraction of these structural destabilizations seems to be already contained in the equilibrium motions of the stable complex.

Unexpectedly, the pattern of structural changes upon ligand removal was not already encoded within the equilibrium fluctuations of the ligand-bound state. Together with the nearly unchanged fluctuations, this suggests that the driving forces for the structural change are not dominated by entropy changes of the binding pocket, which agrees with the role of the solvent discussed below. The largest structural changes were seen within the C-terminal domain, for several residues next to the N-terminus, a solvent-exposed loop, and a small loop containing  $\text{Lys}^{144}$ .

The large structural deviations also provide a likely explanation as to why crystallization attempts of the apo protein have been unsuccessful so far (5), and suggest possible constructs for further crystallization attempts. In particular, due to the observed motion of the amino acids in the binding pocket and in the C-terminal region, the K144A mutation as well as mutations of hydrophobic residues into polar ones in the C-terminal cluster may lead to a more stable ligand-free protein construct.

PCA served to characterize the observed structural changes. Compared to the  $m_3\text{GpppG}$ -bound dynamics of SPN1, the ligand-free trajectories showed extensive sampling, additionally testifying to its drastically, and collectively, enhanced dynamics. Closer analysis of the ligand-free trajectory enabled us to characterize the amino-acid rearrangements in the binding pocket and the C-terminal domain in more detail.

The observed destabilization of a hydrophobic cluster in the C-terminal region suggests a major conformational change upon ligand release. Recent experiments have shown that the export receptor chromosome region maintenance-1 (CRM1) is highly competitive with  $m_3\text{G}$ -capped RNA in binding to SPN1. Although the binding affinity of CRM1 to SPN1 was strongly dependent on the existence of the full SPN1 N-terminal domain, the deletion of the C-terminus beyond residue 285 resulted in an affinity decrease of 60% (22). Furthermore, the effect of several point mutations in CRM1 on the binding to SPN1 was studied (23). It was suggested that the recognition of SPN1 involves multiple sites widely separated on the CRM1 surface. The structural rearrangement observed here in the C-terminal domain upon ligand release thus agrees well with both the observed binding affinities of CRM1 as well as the competitive behavior to  $m_3\text{G}$ -capped RNA. Whether the observed structural rearrangements in the C-terminal domain occur in the available truncated protein only, or also occur in the full-length protein, cannot, in the absence of the full-length structure, rigorously be decided. Overall, these rearrangements suggest a possible explanation for the in vitro binding mechanism leading to the cluster formation of an export complex consisting of Ran-GTP, CRM1, and snurportin 1.

The role of the solvation shell in the binding thermodynamics of the caps is indeed remarkable and unexpected. Both ligands,  $m_3\text{GpppG}$  and  $m^7\text{GpppG}$ , are surrounded by a highly ordered water shell around the N2-nitrogen, differing in its methylation state. In contrast, when bound to SPN1, the volumes of the remaining water shells are quite different for the two ligands. As a result, the protein shields  $m_3\text{GpppG}$  to a much larger extent from the surrounding water than  $m^7\text{GpppG}$ .

As revealed by its differential dynamics,  $\text{Trp}^{107}$  appears to be crucial for this shielding. Indeed, the W107A mutant has been shown to bind  $m_3\text{GpppG}$  with markedly reduced binding affinity (5). Taken together, we suggest  $\text{Trp}^{107}$  as the key residue for the shielding of the two N2-methyl-groups in  $m_3\text{GpppG}$ .

To validate our simulations, binding free energy differences between  $m_3\text{GpppG}$  and  $m^7\text{GpppG}$  to SPN1 were computed with a newly developed method (19). From these calculations, the binding affinity of  $m^7\text{GpppG}$  to SPN1, quantified by the equilibrium dissociation constant  $K_D$ , is reduced by approximately two orders of magnitude with respect to  $m_3\text{GpppG}$ . This result agrees well with estimates from ultraviolet cross-linking experiments (6), for which a decrease by 2–3 orders of magnitude is reported. Further splitting into

entropy and enthalpy shows that the entropically driven desolvation of the dimethylated N2-nitrogen is indeed the main driving force for the better affinity of m<sub>3</sub>GpppG to SPN1.

We note that the AMBER99 force field overstabilizes  $\alpha$ -helical peptide conformations (24), and it was not parameterized with all bonds constrained. However, the overstabilization due to backbone dihedral parameters does not impede structural rearrangements in the C-terminus of the SPN1 structure that is mainly  $\alpha$ -helical, and it has been shown that bond-constraining has no significant influence on free energy calculations (25).

In summary, our free energy calculations support the experimental findings of Huber et al. (6) as well as Strasser et al. (5) for the selective binding of m<sub>3</sub>GpppG, which mimics the m<sub>3</sub>G-cap as an important part of the nuclear localization signal specific for UsnRNP nuclear import (26,27).

We thank Achim Dickmanns for helpful discussions.

## REFERENCES

- Conti, E., C. Müller, and M. Stewart. 2006. Karyopherin flexibility in nucleocytoplasmic transport. *Curr. Opin. Struct. Biol.* 16:237–244.
- Matera, A., R. Terns, and M. Terns. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* 8:209–220.
- Will, C., and R. Lührmann. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* 13:290–301.
- Bahia, D., M. Bach-Elias, A. Aviñó, R. Eritja, and E. Darzynkiewicz. 2006. Trimethylguanosine nucleoside inhibits cross-linking between Snurportin 1 and m<sub>3</sub>G-capped U1 snRNA. *Nucleosides Nucleotides Nucleic Acids.* 25:909–923.
- Strasser, A., A. Dickmanns, R. Lührmann, and R. Ficner. 2005. Structural basis for m<sub>3</sub>G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J.* 24:2235–2243.
- Huber, J., U. Cronshagen, M. Kadokura, C. Marshallsay, T. Wada, et al. 1998. Snurportin1, an m<sub>3</sub>G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.* 17:4114–4126.
- Berman, H., J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Wang, J., P. Cieplak, and P. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–1074.
- Sorin, E., and V. Pande. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* 88:2472–2493.
- Aduri, R., B. Psciuk, P. Saro, H. Taniga, H. Schlegel, et al. 2007. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.* 3:1464–1475.
- Meagher, K., L. Redman, and H. Carlson. 2003. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* 24:1016–1025.
- Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
- Berendsen, H., J. Postma, W. van Gunsteren, A. DiNola, and J. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
- Hess, B., H. Bekker, H. Berendsen, and J. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
- Jorgensen, W., J. Chandrasekhar, J. Madura, R. Impey, and M. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
- van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. Mark, et al. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
- Amadei, A., A. Linssen, and H. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.
- Stumpe, M., and H. Grubmüller. 2007. Aqueous urea solutions: structure, energetics, and urea aggregation. *J. Phys. Chem. B.* 111:6220–6228.
- Goette, M., and H. Grubmüller. 2009. Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *J. Comput. Chem.* 30:447–456.
- Jorgensen, W., and C. Ravimohan. 1985. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* 83:3050–3054.
- van Gunsteren, W. 1988. The role of computer simulation techniques in protein engineering. *Protein Eng. Des. Sel.* 2:5–12.
- Paraskeva, E., E. Izaurralde, F. Bischoff, J. Huber, U. Kutay, et al. 1999. CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J. Cell Biol.* 145:255–264.
- Petosa, C., G. Schoehn, P. Askjaer, U. Bauer, M. Moulin, et al. 2004. Architecture of CRM1/exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol. Cell.* 16:761–775.
- Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, et al. 2006. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinformatics.* 65:712–725.
- Christen, M., C. Christ, and W. van Gunsteren. 2007. Free energy calculations using flexible-constrained, hard-constrained and non-constrained molecular dynamics simulations. *ChemPhysChem.* 8:1557–1564.
- Fischer, U., and R. Lührmann. 1990. An essential signaling role for the m<sub>3</sub>G cap in the transport of U1 snRNP to the nucleus. *Science.* 249:786–790.
- Fischer, U. 1991. Diversity in the signals required for nuclear accumulation of UsnRNPs and variety in the pathways of nuclear transport. *J. Cell Biol.* 113:705–714.