Structural and functional studies of the Scavenger Decapping Enzyme DcpS

Untersuchungen zur Struktur und Funktion des Scavenger Decapping Enzyms DcpS

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Dipl. Biochem. Anna-Lisa Fuchs

aus Karlsruhe

Tübingen

2019

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	25.07.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Remco Sprangers, Universität Regensburg
2. Berichterstatter:	Prof. Dr. Dirk Schwarzer

Table of Contents

Abbreviations

A	Adenine/Adenosine
Å	Ångström
C	Cytosine
C. t. / C. thermophilum	*Chaetomium thermophilum*
CBC	Cap binding complex
CPF	Cleavage and polyadenylation factor
CoA	Coenzyme A
CSP	Chemical shift perturbation
DcpS	Scavenger decapping enzyme
DNA	Desoxyribonucleic acid
E. coli	*Escherichia coli*
G	Guanine / guanosine
H.s. / H. sapiens	*Homo sapiens*
HIT	Histidine triade
I	Inosine
IRES	Internal ribosome entry site
ITC	Isothermal titration calorimetry
M.m. / M. musculus	*Mus musculus*
m^6A	N6-methyl-adeninosine
m^7G	N7-methyl-guaninosine
miRNA	Micro RNA
mRNA	Messenger RNA
MST	Microscale thermophoresis
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NCIN	Non-canonical initiating nucleotide
ncRNA	Non-coding RNA
NMD	Nonsense mediated decay
NMP	Nucleotide monophosphate
NMR	Nuclear magnetic resonance
nt	nucleotide

ORF	Open reading frame
pre-mRNA	Precursor messenger RNA
PTC	Premature termination codon
rmsd	Root mean square deviation
RNA	Ribonucleic acid
RNA Pol II / III	RNA Polymerase II / III
RNAi	RNA interference
S. c. / S. cerevisiae	*Saccharomyces cerevisiae*
SAH	S-adenosylhomocystein
SAM	S-adenosylmethionine
SMA	Spinal muscular atrophy
SMN	Survival motorneuron
TMG	2,2,7-trimethylguanine
TROSY	Transverse relaxation optimised spectroscopy
U	Uracil
Urea-PAGE	Urea polyacrylamide gel electrophoresis
UTR	Untranslated region
WT	Wild type

Amino acids are abbreviated with their one or three letter code.

Summary

Messenger RNA (mRNA) is the working copy of a gene containing the information for the production of a specific protein. The amount of any given protein has to be adjusted to the current needs of a cell. One mechanism for the regulation of protein production is the targeted degradation of the corresponding mRNA. A key element of an mRNA molecule is its 5' cap structure that serves as a binding hub for interacting proteins and as a protection against premature degradation. During regulated mRNA degradation, the RNA body can be degraded by the exosome complex from the 3' end, which leaves short capped RNA fragments. The cap structure of these fragments is removed by the Scavenger Decapping Enzyme (DcpS). This dimeric enzyme has a bipartite active site which is constituted from residues of two domains. Substrate binding induces a see-saw like conformational change that closes the active site around the substrate. In this thesis, we show that this domain motion is essential for the catalytic activity of DcpS, but that under substrate excess these motions can be too fast to allow for efficient substrate turnover, resulting in a unique way of substrate inhibition.

To date, the information on the structure and protein dynamics of DcpS during catalysis rely on the usage of cap analogues, representing only the cap structure itself and one nucleotide of the RNA body. The products of the exosomal degradation of mRNA and thus substrates of DcpS, however, are short capped RNA fragments. The synthesis of such capped RNA was not possible until now, and we therefore established a method for the large scale enzymatic capping of *in vitro* transcribed RNA that employs the capping enzyme of vaccinia virus. This allows us to produce homogeneous, capped RNA of different length enabling in-depth biochemical and biophysical studies of the DcpS enzyme.

Activity assays of DcpS with capped RNA of different lengths showed a clear influence of substrate length on catalytic activity with a strong preference for substrates of up to two nucleotides. We show that longer RNA is incompatible with the catalytically required domain motions of DcpS, due to steric hindrance between the enzyme and the mRNA body. We show that the preference of DcpS for capped mono- and di-nucleotides is conserved in various species and that this length dependence is in accordance with our preliminary data on the product length of the exosome. From this we conclude that the biological function of the threshold in substrate usage of DcpS is based on a direct handover of RNA fragments from the exosome to DcpS to prevent the decapping of actively translated transcripts.

Zusammenfassung

Messenger RNA (mRNA) ist die Arbeitskopie eines Gens, welche die Information zur Herstellung eines spezifischen Proteins trägt. Die Menge eines jeden Proteins muss an den ständig wechselnden Bedarf einer Zelle angepasst werden. Ein Mechanismus, die Proteinproduktion zu regulieren, ist der gezielte Abbau der entsprechenden mRNA. Ein wichtiges Element eines mRNA Moleküls ist seine 5‘-Cap-Struktur, welche als Plattform für Protein-Interaktionen, sowie als Schutz vor vorzeitigem Abbau dient. Während des regulierten Abbaus der mRNA, kann der RNA-Körper durch den Exosom-Komplex vom 3‘-Ende her abgebaut werden, was in kurzen gecappten RNA-Fragmenten resultiert. Die Cap-Struktur dieser Fragmente wird durch das Scavenger Decapping Enzym (DcpS) entfernt. Das dimere Enzym besitzt ein zweigeteiltes aktives Zentrum, welches sich aus Resten zweier Domänen zusammensetzt. Durch Substratbindung wird eine wippenartige Konformationsänderung ausgelöst, welche das aktive Zentrum um das Substrat schließt. Wir zeigen, dass diese Bewegung der Domänen für die katalytische Aktivität von DcpS essentiell ist. Allerdings kann die Bewegung unter Substratüberschuss zu schnell sein, um effektiven Substratumsatz zu gestatten, was in einer einzigartigen Weise der Substrat-Inhibierung resultiert.

Bis heute beruhen alle Informationen über Struktur und Dynamik von DcpS auf Cap-Analoga, welche lediglich die Cap-Struktur selbst und das erste Nucleotid des RNA-Körpers repräsentieren. Allerdings sind die Produkte des exosomalen Abbaus der mRNA, und damit die Substrate von DcpS, kurze gecappte RNA-Fragmente. Da die Synthese solcher gecappter RNA bislang nicht möglich war, haben wir eine Methode für das enzymatische Capping von *in vitro* transkribierter RNA in großem Maßstab, beruhend auf dem Capping Enzym des Vaccinia Virus, entwickelt. Diese erlaubt es uns, homogene, gecappte RNA unterschiedlicher Länge herzustellen, welche eingehende biochemische und biophysikalische Untersuchungen des DcpS-Enzyms ermöglicht.

Aktivitäts-Assays von DcpS mit gecappter RNA unterschiedlicher Länge zeigen einen klaren Einfluss der Substratlänge auf die katalytische Aktivität mit starker Präferenz für Substrate von bis zu zwei Nukleotiden. Wir zeigen, dass längere RNA die Bewegung der Domänen, welche für die erfolgreiche Katalyse notwendig ist, nicht induziert. Aus Strukturdaten können wir schließen, dass der Schließmechanismus des aktiven Zentrums aufgrund eines längerer RNA-Körpers sterisch behindert wird. Die Untersuchung von Enzymen verschiedener Spezies ergibt, dass die

Länge der effektiv umgesetzten Substrate und die strukturelle Grundlage für die Substrat-Präferenz speziesunabhängig konserviert sind. Die Präferenz für gecappte Mono- und Dinukleotide von DcpS stimmt mit unseren vorläufigen Daten für die Länge der Exosom-Produkte überein. Daraus schließen wir, dass die biologische Funktion der Beschränkung von DcpS auf kurze Substrate auf einer direkten Übergabe von RNA-Fragmenten von Exosom zu DcpS beruht. Dies verhindert, dass von Transkripten, welche aktiv translatiert werden, die Cap-Struktur durch DcpS entfernt wird.

List of publications

Published work:

- Neu, A., Neu, U., **Fuchs, A.-L.**, Schlager, B., & Sprangers, R. (2015). An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nature Chemical Biology*

- **Fuchs, A.-L.***, Neu, A.*, & Sprangers, R. (2016). A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA*
  * shared first authorship

Prepared manuscripts:

- **Fuchs, A.-L.**, Neu, A., & Sprangers, R.*;* Molecular basis for mRNA length sensing by the DcpS decapping enzyme

Contributions

My contributions to the work performed in the studies included in this thesis are stated hereafter:

Neu, A., Neu, U., **Fuchs, A.-L.**, Schlager, B., & Sprangers, R. (2015). An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nature Chemical Biology*

- Protein purification
- Measurement of turnover rates by ^{31}P NMR
- ITC measurements

Fuchs, A.-L.*, Neu, A.*, & Sprangers, R. (2016). A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA*

- Establishing of purification protocols
- Production of RNA
- Capping reactions for initial characterization
- Capping reactions of long RNAs, structures RNAs and sequence specific RNA
- Decapping assays
- Sample preparation for mass spectrometry

Fuchs, A.-L., Neu, A., & Sprangers, R.; Molecular basis for mRNA length sensing by the DcpS decapping enzyme

- I performed all experiments
- I was supported by Dr. Ancilla Neu in crystallographic analysis

1 Introduction

1.1 The Cellular Function of mRNA

Messenger RNA (mRNA) is the connecting element between the coding molecule DNA and the performing molecule protein. In the nucleus of eukaryotes, the information encoded in the DNA is transcribed into mRNA by the RNA Polymerase II (Pol II). Simultaneously, processing and modification of this initial transcript, the precursor mRNA (pre-mRNA), takes place. The 5' and 3' ends of the RNA are modified for stability, non-coding regions are removed by splicing and single bases are modified specifically [1]. Only after successful processing of the pre-mRNA into mature mRNA, the transcript is exported into the cytoplasm. Transport across the nuclear membrane takes place through the nuclear pore complex, a structure that allows for regulated transport of large molecules across the nuclear membrane. In the nucleus, mRNA is bound by export receptors that convey it into the cytoplasm. In the cytoplasm, the RNA is released and further modifications, as for example base modification or hypermethylation of the cap might take place [2–4]. The mRNA is then translated into a polypeptide chain by the ribosome. The nascent polypeptide chain folds into its functional conformation and can fulfil cellular functions as enzymes, regulatory proteins or structural proteins [5] (**Figure 1.1**).

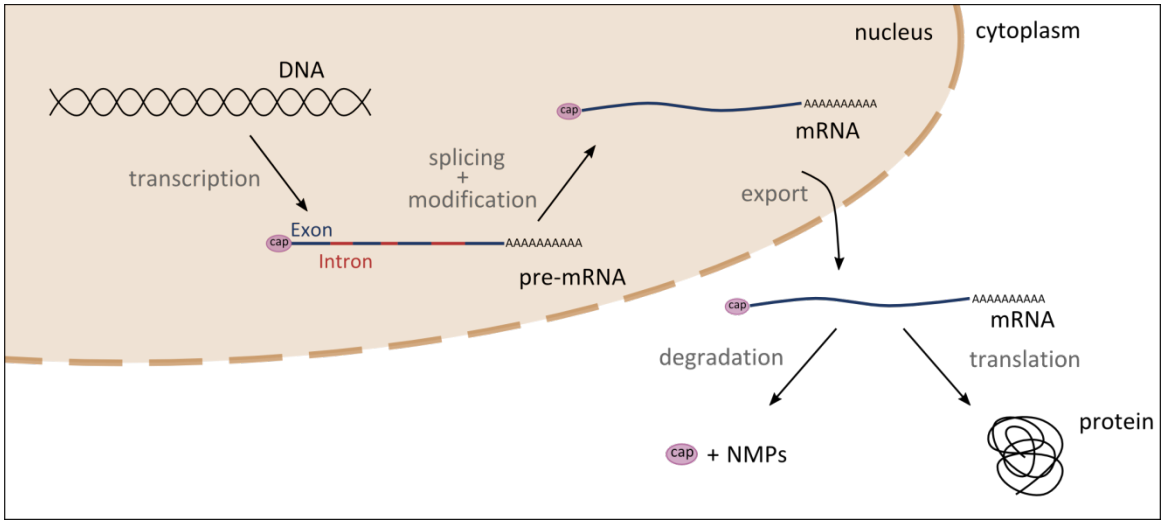


Figure 1.1 | Lifecycle of mRNA. In the nucleus, DNA is transcribed into pre-mRNA, which contains exons as well as introns. Co-transcriptionally, the pre-mRNA is modified with a 5' cap and additional base modifications, and introns are removed by splicing. The mature mRNA is exported to the cytoplasm through the nuclear pore complex and translated into protein. Obsolete mRNA molecules are degraded.

The concentration of any given protein is crucial for the maintenance of cellular homeostasis and provides the cell with a means to react to environmental changes. Therefore it is important that the production rate of a protein is constantly adapted to current cellular needs. Thus all processes leading to protein production that are described above, are heavily regulated at each step. One aspect of this regulation machinery is the targeted degradation of mRNA molecules [6].

1.2 Structural Characteristics of Eukaryotic mRNA

Eukaryotic mRNA exhibits several characteristic features: The coding region that is translated into protein, flanking 5' and 3' untranslated regions (UTRs), the 3' poly(A)-tail and the 5' cap structure (**Figure 1.2 A**).

The coding sequence of mRNA is also called the open reading frame (ORF). Initially it consists of exons interrupted by none-coding introns, which are removed co-transcriptionally by the splicing machinery. Alternative splicing events allow for multiple proteins to be expressed from the same starting pre-mRNA molecules [7–9].

The coding sequence of an mRNA is flanked by untranslated regions (UTR) with regulatory functions. The 5' UTR can form intricate secondary structures which affect the scanning efficiency of the translation initiation complex both positively and negatively, and possess sequences that are recognized by regulatory non-coding RNA [10]. The longer 3' UTR is a platform for RNA-binding proteins that interact sequence or structure specific and mediate mRNA stability, translation efficiency and mRNA localisation [11].

Posttranscriptional modifications of the RNA body, also called the epitranscriptome, are modifications that are not directly encoded in the DNA. Methylated bases and ribose moieties are a frequent modification, as for example methylation of the N1 and N6 position of adenine and the N3 and C5 position of cytosine. The degree and biological function of these base methylations is an active field of research [12]. In most cases, in which adenine is the 5' base of the RNA body, it is present in the dimethylated form *N6*-methyl-2'-*O*-methyladenosine (m^6A_m) [13]. m^6A enhances nuclear export efficiency [14] and regulates mRNA decay by localizing mRNA to P-bodies [15]. Another modification is base editing, the most common being A to I editing, where adenine is deaminated to inosine. Inosine is read as guanine by the ribosome and thus changes

the genetic code in certain positions [16]. Most posttranscriptional modifications are catalysed by enzymes located in the nucleus and are performed early in the mRNA lifecycle [17].

The 3' end of mature mRNA consists of a stretch of adenosine residues. In yeast, this poly(A)-tail has a length of about 60 adenosines, whereas in higher eukaryotes this is up to 250 adenosines [3]. The poly(A) tail prevents premature degradation of the mRNA and is involved in nuclear export and translational regulation [18]. When Pol II reaches the recognition sequence for Poly(A)-tail formation, the trimodular cleavage and polyadenylation factor (CPF) induces transcription termination, cleavage of the transcript ~100 nt upstream of the polyadenylation signal and formation of the poly(A)-tail [19,20].

After transcription of 20 to 30 nucleotides by RNA Pol II, the nascent pre-mRNA molecule is modified at the 5'-end by addition of the 5' cap structure [21,22]. The simplest cap structure is the cap0, which comprises an N7 methylated guanosine moiety attached to the 5' end of the mRNA molecule by a 5'-5' triphosphate bridge [23–25]. Additional methylation at the 2' OH group of the ribose moiety of the 5' base of the RNA body results in a cap1 structure, further 2' O-methylation of subsequent ribose moieties in cap2 up to cap4 (**Figure 1.2 B**). From cap2 on, the modification is no longer co-transcriptionally and takes place in the cytoplasm. The abundance of each cap structure differs between species, with a general trend to higher methylated cap structures in higher eukaryotes [13]. In some species like nematodes or certain RNA viruses like Sindbis virus, a hypermethylated variant of the cap is found. Here, the cap guanine is additionally di-methylated at the N2 position, resulting in the $m^{2,2,7}G$ trimethylated cap (TMG) [26,27]. SnRNA, snoRNA and telomere RNA can also exhibit hypermethylated caps, which are generated in the cytoplasm by Trimethylguanosine Synthase 1 (Tgs1) [28,29]. Direct methylation of the 5' γ-phosphate is also possible, as in the U6 snRNA transcribed by RNA Pol III [30].

The 5' cap structure is necessary for several cellular processes, including mRNA splicing [31–33], nuclear export of mRNA to the cytoplasm [34], and translation initiation [35] by acting as a binding platform for proteins that facilitate further protein-protein interactions, for example the cap binding complex CBC [31] and the translation initiation factor eIF4E [36]. Furthermore the cap structure protects the mRNA from premature degradation from the 5' end [37]. The absence of a cap structure is a trigger for the innate immune system that recognises uncapped RNA as non-self and starts the immune defence programme [38].

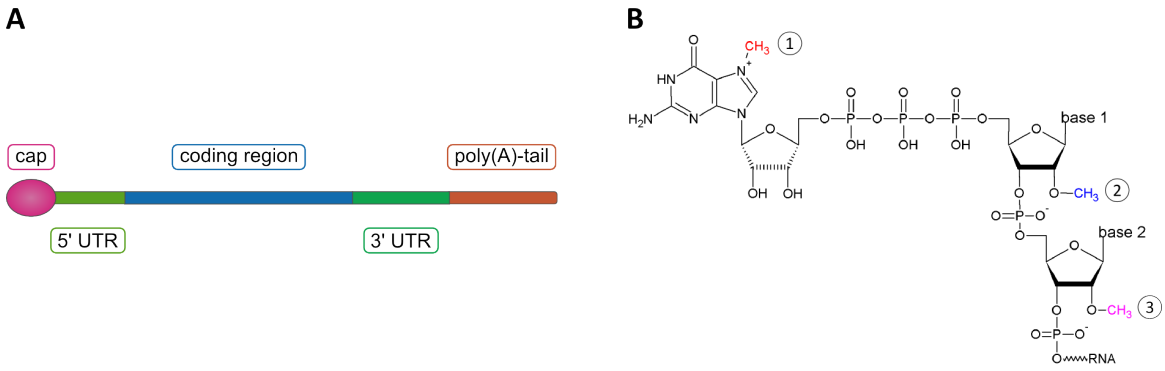


Figure 1.2 | Characteristic features of mRNA molecules. A: The region coding for a protein sequence (blue) is flanked by untranslated regions (UTR, green) and exhibits a poly(A)-tail at the 3' end (orange) and a cap structure at the 5' end (pink). **B:** The cap structure of mRNA has species dependent methylation patterns. The simplest cap is cap0 methylated only at the N7 position of the cap guanine (red, 1). Cap1 has an additional methylation at the 2' O position of the first base of the RNA 5' end (blue, 2) and cap2 has a further methylation at the ribose 2' O position of the second base of the RNA body (pink, 3).

Splicing events and modifications that are performed along the way from the pre-mRNA to the mature mRNA are essential for proper protein production. They play an important role in transport of the mRNA to its target location and its regulated translation into a functional protein sequence, encoded in the open reading frame. Throughout the following paragraphs, the focus will be on the cap structure, how it is attached to the 5' end and how it is removed in the course of mRNA degradation.

1.3 Mechanisms of mRNA Capping

RNA cap structures are an essential feature of eukaryotic mRNA and are therefore found in all eukaryotes and also the transcripts of their viruses. As a cap structure is necessary for the recognition of the mRNA by the cellular translation machinery, viruses that make use of the host translation machinery need to exhibit the same 5' end modification as their host. Most eukaryotes and their viruses, use the same general mechanism for the attachment of a cap structure to the 5' end of mRNA. Nascent RNA possesses a triphosphorylated 5' end, which is modified in a three step reaction to a cap0 structure (**Figure 1.3 A**). First, the triphosphate is dephosphorylated by an mRNA triphosphatase resulting in a diphosphate and inorganic phosphate. Second, a guanylyltransferase transfers the guanosine moiety of GTP first to the enzyme itself and then to the RNA via a 5'-5' triphosphate linkage by attaching the α-phosphate of GTP to the β-phosphate of the diphosphorylated RNA body. Lastly, a methyltransferase uses

S-adenosylmethionine as a donor for the methylation of the N7 position of the cap guanine and releases S-adenosylhomocystein [38].

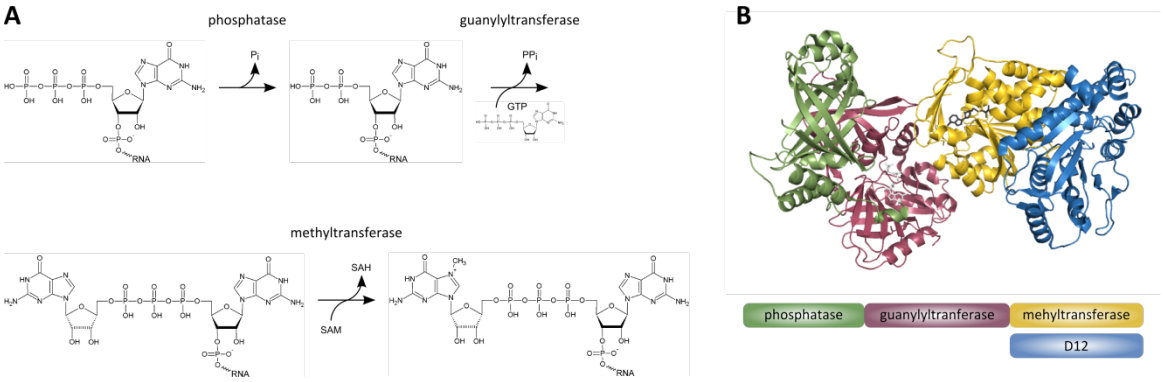


Figure 1.3 | Reaction mechanism of the vaccinia capping enzyme. A: 5' Triphosphorylated RNA is dephosphorylated by the phosphatase. GTP is attached to the resulting diphosphate-RNA by a guanylyltransferase via a 5'-5' triphosphate linkage. The N7-position of the cap-guanine is methylated by a methyltransferase which used SAM as the methyl donor. **B:** The vaccinia capping enzyme is a complex of two proteins. The D1 protein harbours all three catalytic activities required for the capping reaction, whereas the D12 protein is associated with the methyltransferase domain of D12 and stimulates its activity.

The enzymatic functions for cap0 formation are conserved among species that use the reaction strategy described above. However, the number of enzymes involved in the capping process differs. Some species have one enzyme per catalytic function, e.g. yeast [39], others developed bifunctional enzymes by merging the triphosphatase and guanylyltransferase activity into one peptide chain, e.g. mouse [40], or even developed trifunctional enzymes. Two examples of trifunctional enzymes are the capping enzymes of the large DNA viruses Mimivirus and Poxvirus [41,42]. In this study, we made use of the capping enzyme of the poxvirus vaccinia. Unlike the fully autonomous single-chain capping enzyme of Mimivirus, the trifunctional 97 kDa enzyme D1 of the poxvirus vaccinia requires the stimulatory 33 kDa protein D12 for full activity of the methyltransferase domain (**Figure 1.3 B**) [43,44].

Interestingly, the order of reactions in capping is not conserved. For example, some virus-families exploit pathways, in which for example GTP is methylated in a first step and attached to triphosphorylated RNA in a second step [38]. Orthomyxoviridae like Influenza use a completely unrelated mechanism of capping their RNA, called cap snatching, by which the 5' end of cellular RNA is cleaved by a viral enzyme and reattached to the 5' end of a viral RNA [45].

Until recently, the general idea has been that RNA capping is only present in eukaryotes but not in bacteria. In the past years however, bacterial RNA was found, that exhibits cap-like structures that resemble NAD and CoA residues. Their cellular function, however, remains controversial [46–48]. Both modifications posses a terminal nucleotide in their free form and thus

can be incorporated as non-canonical initiating nucleotides (NCIN) both by the bacterial RNA polymerase, but also by the eukaryotic RNA Pol II [49].

The production of m^7G-capped RNA on laboratory scale for structural or functional experiments has been a challenge for a long time. Different methods were proposed, including addition of an excess of commercially available cap analogue to *in vitro* transcription reactions [33,50–52]. This, however, results in partial incorporation of the cap analogue in the reverse orientation [53]. For this reason, anti-reverse cap analogues were developed that are incorporated only in one orientation but have chemical modifications at the 3' OH group [54]. Another approach is the chemical synthesis of capped RNA, which faces several challenges, as the methylated guanine moiety is labile to acidic and basic conditions. Short non-methylated capped RNA can be chemically synthesised and methylated in a subsequent enzymatic reaction [55] or very short RNAs of 1-3 nt can be fully chemically synthesised with complex methylation patterns [56]. During the course of this thesis, we established a protocol for the capping of *in vitro* transcribed RNA employing the capping machinery of vaccinia virus.

1.4 Pathways of mRNA Degradation

To maintain cellular homeostasis, the concentration of mRNA is tightly regulated on the level of transcription, but also on the level of targeted degradation. mRNA decay is largely influenced by interactions with proteins and ncRNAs, which can inhibit or promote degradation depending on the cellular status [6] The main degradation pathways of RNA molecules are 3'-5' decay or 5'-3' decay (**Figure 1.4**). Both pathways are present in eukaryotes; however, their abundance is species dependent. 5'-3' decay is the major degradation pathway in yeast, whereas in mammals 3'-5' decay is the primary pathway [57]. The initial step of both major degradation pathways is deadenylation of the 3' end of the mRNA by the deadenylase complexes PAN2:PAN3 or CCR4:NOT [58–60]. Hereafter the directionality of the degradation is different between the two pathways.

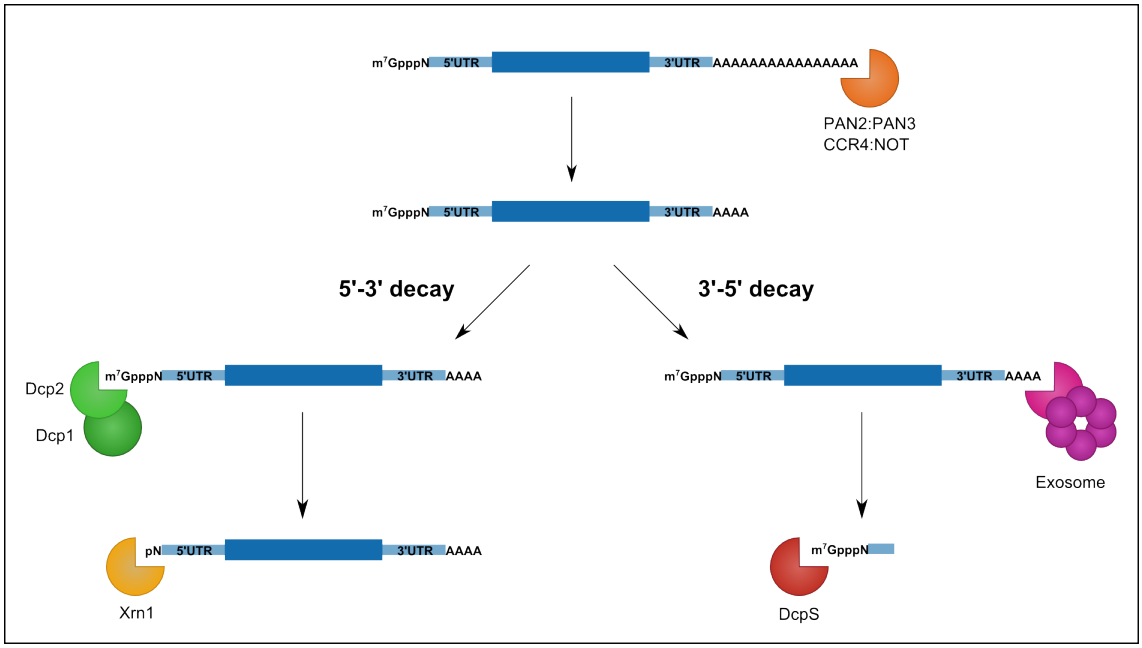


Figure 1.4 | mRNA is degraded via two major pathways. The poly(A)-tail is removed by the PAN2:Pan3 or CCR4:NOT complex and the resulting deadenylated RNA is subject to either 5'-3' or 3'-5' decay. In 5'-3' degradation the cap structure is removed from the 5' end by Dcp1:Dcp2 and the RNA subsequently hydrolysed by Xrn1. In 3'-5' decay, the RNA is hydrolysed from the 3' end by the exosome complex and the remaining short capped fragments are decapped by DcpS.

In the 5'-3' degradation pathway, the next step is the removal of the 5' cap structure by the decapping enzyme Dcp2. The enzyme is recruited to the 5' end by the Pat1:Lsm1-7 complex that is interacting with the deadenylated 3' end. Dcp2 itself has only low intrinsic decapping activity facilitated by its Nudix-domain. However, the activity of Dcp2 is vastly enhanced by complex formation with decapping enhancers, foremost Dcp1, but also the Edc proteins, Pat1, Dhh1 and others [61–70]. Dcp2 releases m^7GDP and 5' monophosphorylated RNA [68,71], the latter of which is processively degraded by the exoribonuclease Xrn1 [72]. The proteins of the degradation network assemble in P-bodies, membrane-less cellular compartments that are formed by liquid-liquid phase separation [73].

In 3'-5' decay, after deadenylation of the 3' end, the RNA is hydrolysed by the exosome complex [74]. The eukaryotic exosome is a complex of at least 10 proteins. Nine proteins that show no catalytic activity form a barrel like structure with a central channel through which the RNA enters the complex. Rrp44, the active subunit of the cytoplasmic exosome complex, is located at the bottom of the channel and can adopt two different conformations with respect to the rest of the complex. The Rrp44 protein has exo- as well as endonucleolytic activities, located in the RNB domain and PIN domain, respectively [75–78]. Exosomal degradation is processive and releases nucleotides and a small capped RNA fragment [74,79]. The cap structure of these small RNAs is

removed by the Scavenger Decapping Enzyme DcpS (Dcs1p in yeast). The reaction products of the DcpS decapping reaction are m^7GMP and a 5' diphosphorylated RNA fragment [80].

Beyond the 3'-5' and the 5'-3' mRNA degradation pathways, mRNA can be degraded in different manners, partly in the context of quality control to prevent the translation of faulty mRNA. In the case of premature termination codons (PTC) in an mRNA, nonsense mediated decay (NMD) takes place. Here, during the first round of translation, the main NMD factor UPF1 interacts with uncommonly placed Exon Junction Complexes (EJC), and UPF1 is consequently phosphorylated by the kinase SMG1 which induces endonuclealytic cleavage of the RNA by SMG6. The two RNA fragments are then degraded by the exosome and Xrn1 respectively [6,81,82]. No-go decay takes place at stalled ribosomes and is also initiated by endonucleolytic cleavage, whereas RNA without a stop codon is degraded by non-stop decay [6]. These mRNA quality control pathways take place co-translationally [83]. Furthermore, mRNA is a subject to degradation that is initiated by RNA interference. Here, a miRNA is used as a guide in the Argonaut:RISC complex to bind complementary or partly complementary sequences in mRNA and slows down translation, induces cleavage of the RNA duplex or recruits CCR4, leading to deadenylation and thus to degradation [82,84,85].

What all pathways have in common is the degradation of the mRNA by specialized enzymes that each recognise and degrade a specific feature of the mRNA. Each degradation event includes removal of the poly(A)-tail, removal of the cap structure and degradation of the RNA body. The order of these events and the responsible enzymes vary with the given pathway.

1.5 The Scavenger Decapping Enzyme DcpS

In 3'-5' decay, after deadenylation of the 3' end, the mRNA body is degraded by the exosome complex. The product of exosomal degradation is short capped RNA of supposedly 4-5 nucleotides [86-88]. DcpS (Dcs1p in yeast) hydrolyses the cap structure of these mRNA fragments and releases m^7GMP and dephosphorylated RNA [57,80,89]. DcpS was first found in 1977 by Nuss and Furuichi by analysis of HeLa cell extracts, where it is the major RNA decapping enzyme [57,80].

The 80 kDa homodimer DcpS [57,89], has a domain swapped fold with a distinct N- and C-terminal domain, which are connected by a flexible hinge region [90,91]. DcpS is a member of the

HIT protein family, which is characterized by its histidine triade (H-X-H-X-H) in the active site and that is located in the C-terminal domain in DcpS. However, DcpS is only catalytically active when both N- and C-terminal domain are present and the catalytically competent active site is formed by residues of both domains alike. [89,91–94].

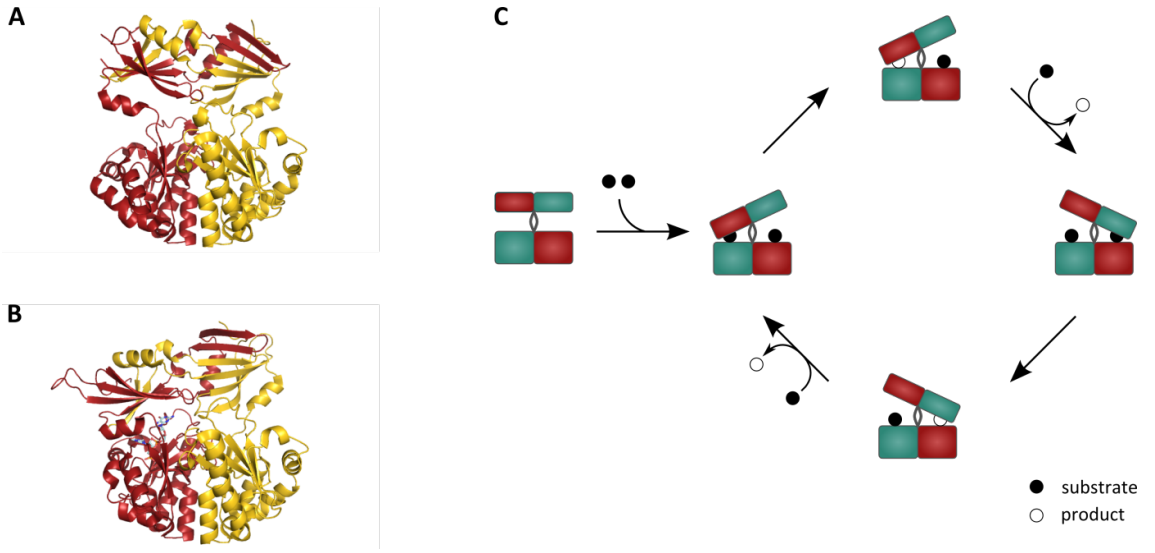


Figure 1.5 | Structure of *H.s.* DcpS and dynamic reaction cycle. A, B: DcpS is a homodimer with a domain swap fold. In the apo form, DcpS is symmetrical (**A,** pdb excession code 1XML)) whereas in the ligand bound form it adopts an asymmetric conformation with the ligand bound in the active site between the N- and C-terminal domain (**B,** pdb excession code 1ST0). **C:** The DcpS dimer has two active sites. When substrate is bound in one site, the enzyme closes this active site while simultaneously opening the second binding site. When catalysis is performed in the closed site, it opens to release the products and thereby closes the second site where then catalysis can take place.

Structural data is known for the human [90,91,95,96] and mouse enzymes in both the apo form and in complex with different cap analogues, m^7GDP and medicinal inhibitors [90,91,95–99]. The structural data shows that DcpS exists in two distinct conformations (**Figure 1.5 A, B**). In the apo form, the homodimer is completely symmetrical, whereas ligand binding induces a conformational change in which the N-terminal domain flips over upon the C-terminal domain to enclose the ligand in the occupied binding site while at the same time opening the second binding site (**Figure 1.5 C**). Thereby the N-terminal domain performs a movement of 37 ° around the central axis or 30 Å [90,91]. The side chain orientation of the residues of the open site is quite similar to the one in the apo protein, while the closed site reorients to form the catalytically competent active site [90]

The hydrolysis reaction of DcpS is an S_N2-type reaction in which the central histidine residue of the HIT motif acts as a nucleophile attacking the α-phosphate of the triphosphate linkage [90,93]. The rate limiting step of the reaction is depending on the concentration of substrate available. Under single turnover conditions, binding of the cap is rate limiting, while under multiple

turnover conditions, it is the sidechain rearrangement to the catalytically competent conformation [100].

In terms of specificity, DcpS is indifferent to the first base of the RNA body being G or A, although the molecular interaction with the base is slightly different in the two cases [91]. In contrast, the m^7G-cap structure is bound by specific sidechain interactions. Unmethylated or trimethylated guanine or methylated inosine are not recognised as substrates by the human enzyme [80,89]. The activity of DcpS is dependent on substrate length. Liu *et al.* showed that catalytic activity is drastically reduced when the substrate length increases from 1 to 10 nucleotides [80,89]. On long substrates DcpS shows residual activity, albeit 2500 fold reduced compared to short substrates. This difference in activity correlates with the difference in affinity for short (K_d=75 nM) and long (K_d=1.25 µM) substrates [94].

Several interactions between DcpS and other proteins have been observed. DcpS was found in complex with a series of proteins involved in mRNA degradation, including components of the exosome complex [57]. Furthermore, an interaction with Xrn1 was shown [101] and an influence of DcpS on the regulation of the 5'-3' decay and mRNA stability mediated by protein-protein interactions [102,103]. Moreover, DcpS was shown to be involved in miRNA processing [104,105], to have an influence in pre-mRNA splicing [106] and was found to have a phosphorylation dependent role in nutrient availability under stress conditions [107]. An inactive homologue of DcpS, Dcs2p, is only found in *S. cerevisiae* and both proteins can form heterodimers as a means of activity regulation [107,108].

The importance of decapping by DcpS is underlined by two diseases that are associated with DcpS. A mutation in the *DCPS* gene that leads to a loss of function variant of DcpS is responsible for the Al-Raqad syndrome, which manifests in craniofacial anomalies, intellectual disabilities and neuromuscular defects [109]. Spinal muscular atrophy (SMA) is caused by deletion or mutation of the survival motorneuron 1 (SMN1) gene. SMN2, a SMN1 homologue that is lacking the last exon of SMN1 can partially compensate for SMN1 and reduce the severity of SMA. C5-substituted quinazolines like RG3039 are enhancing the promoter of SMN2 by an unknown mechanism. These drugs were found to bind to DcpS and inhibit it by keeping it in an open conformation [95,110–112].

In summary, DcpS is an enzyme with a clear catalytic function and diverse regulatory functions beyond that. Its main function is the decapping of short mRNA fragments in the 3'-5' mRNA decay. Despite the insights into the cellular function of DcpS and the structural

information on two distinct conformations, little is known about the mechanism by which the conformational changes influence the catalytic efficiency. Furthermore, it remains elusive, how the preference of DcpS for short mRNA substrates is achieved. This work focuses on these two aspects of the regulatory mechanisms important for the efficient functioning of the Scavenger Decapping Enzyme.

2 Aims and Significance

Tight regulation of mRNA levels is a key aspect of maintaining the exact levels of protein production required for proper cellular function. An essential part of this regulation is the degradation of mRNA, which is conducted in three steps: deadenylation, hydrolysis of the mRNA body and removal of the cap structure. Here we study the Scavenger Decapping Enzyme DcpS, for which a motion between two conformational states is required to perform the decapping reaction in the 3'-5' mRNA decay pathway. For a long time, proteins have been studied as static, rigid molecules and their abundant and often functionally necessary dynamics were neglected. In this work, we use special NMR techniques that are able to analyse the dynamics of the protein and broaden the understanding of DcpS to a dynamic model of the catalytic cycle. Furthermore, we took a look at how these catalytically required motions impact the efficiency of the reaction. Our results are not only important to understand the mechanism of DcpS, but are also valuable for the general understanding of how protein dynamics influence catalytic mechanisms. When developing new drugs for example, taking protein dynamics of the substrate binding site into account will provide a further level of interference with protein function. Drugs could be designed to modulate enzyme activity through interference with functionally important dynamics.

In Neu *et al* we use a cap analogue, only representing the cap structure and one base of the RNA body. To study the interaction of DcpS with longer RNA substrates, a method for the production of longer capped RNA in sufficient quantity was a prerequisite. To that end, we established a robust protocol for the capping of *in vitro* transcribed RNA based on the capping enzyme of vaccinia virus. This enabled us to produce any capped RNA substrate for subsequent experiments with DcpS to unravel how the length of the RNA influences DcpS activity. This method is not only essential for the work presented here, but proofed also useful for other researchers in the RNA field. So far, we have distributed the material for the recombinant expression of the capping enzyme to over 30 laboratories worldwide.

With the method at hand to produce capped RNA of different length, we focused on the question of why DcpS preferably acts on short substrates and if this feature could be explained by structural characteristics. Employing a combination of enzymatic assays, NMR experiments and protein crystallography, we elucidated the molecular details regarding the interaction of

DcpS with eligible substrates, how the enzyme recognises the length of the substrate and reveal that conformational changes are essential for efficient substrate turnover.

Our results give insights into the way in which RNA is transferred from the exosome to DcpS. We elucidated the mechanism that ensures the substrate preference of DcpS for very short capped RNA and that thus prevents random decapping of longer cellular RNA that is still engaged in translation.

3 Results

3.1 An Excess of Catalytically Required Motions Inhibits the Scavenger Decapping Enzyme

The Scavenger Decapping Enzyme (DcpS) is a central enzyme in mRNA degradation and catalyses the removal of the 5' cap structure of short mRNA fragments. In previous structural studies of the human enzyme, two distinct conformations of DcpS were observed: a symmetrical apo form and a substrate bound asymmetrical form (see **Figure 1.5 A, B** in Chapter 1.5) [90,91]. These two distinct structures suggest a domain motion of the enzyme upon substrate binding. However, this dynamic model is based only on the observation of two static structures of the catalytic cycle and molecular dynamics simulations [113] and no data on the actual dynamics are available (**Figure 3.1 A**). To fill this gap, we employed NMR spectroscopy methods that are unique in their ability to record dynamic protein motions with per-residue resolution. For extensive NMR measurements protein samples have to be stable at room temperature for several days. Due to a lack of stability for the human protein, all experiments were performed with DcpS from *S. cerevisiae* unless otherwise stated. Furthermore, we used a catalytically inactive mutant for dynamics studies of DcpS that is mutated in the histidine triade, DcpS^{H268N}, to prevent degradation of the substrate. As a model substrate, the cap analogue m^{7}GpppG was used, which resembles the methylated cap guanosine linked to a single guanosine. This is the shortest possible substrate for DcpS and the reaction products are m^{7}GMP and GDP (**Figure 3.1 B**).

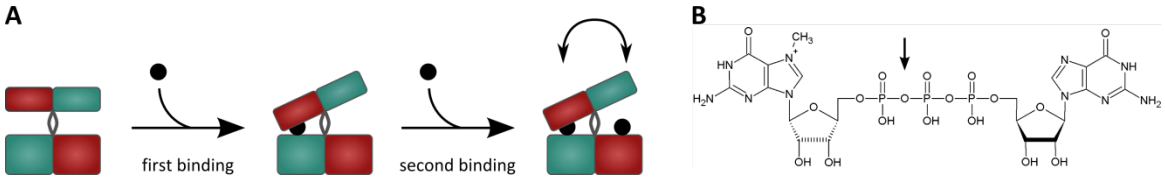


Figure 3.1 | Binding mode and cleavage of DcpS. A: DcpS has two binding sites located between the N- and C-terminal domain. First, one binding site is occupied, which induces a conformational change, and second, the open binding site is successively occupied by a second substrate, inducing the constant flipping motion of the N-terminal domain. **B:** DcpS cleaves between the α- and β-phosphate of the triphosphate bridge seen from the cap structure. In the case of the cap analogue m^{7}GpppG as the substrate, the products are m^{7}GMP and GDP.

DcpS is a complex of 80 kDa and thus conventional TROSY (transverse relaxation-optimised spectroscopy [114]) NMR methods with ^{1}H-^{15}N-labelling of the protein cannot be used, due to spectral overlap and slow tumbling of the molecule, resulting in severe peak broadening.

Instead, we applied ^{13}C-methyl group labelling of isoleucine (Ile) and methionine (Met) residues. These residues are distributed across the protein and thus allow for observation of conformational changes in all parts of the protein. We recorded methyl-TROSY experiments that are especially useful for NMR studies of high molecular weight proteins. Tugarinov *et al.* found that for ^{1}H-^{13}C-labelled methyl groups, the standard HMQC experiment in itself shows a TROSY effect [115]. In addition, signal intensity of methyl-TROSY spectra benefit from the fast rotating methyl groups with their three chemically identical protons.

For DcpS in the apo state, we observed one set of resonances in NMR, where every signal corresponds to two residues in an identical chemical environment, indicating that the dimer is perfectly symmetrical without substrate (**Figure 3.2 B**, black spectrum). The resonances were assigned by a mutational approach. To that end, every Ile and Met residue is successively mutated to a residue invisible in methyl-TROSY spectra. Comparison of the spectrum of the mutant to that of the WT protein allows the assignment of the residue changed in the mutant. Furthermore, spectra of the isolated dimeric N- and C-terminal domains were recorded. They overlap well with the spectra of the full length protein, which indicates that little interaction is present between the two domains in the symmetric conformation.

Titration of the catalytically inactive mutant DcpS^{H268N} with cap analogue as a substrate results in spectra where some resonances of the spectrum in the apo state split into two resonances (**Figure 3.2 B**, red spectrum). As the substrate is bound, the enzyme undergoes the conformational change to the asymmetric conformation. In this conformation, the residues of the two protein chains of the dimer are not in the same chemical environment which is indicated by splitting of the corresponding resonances (**Figure 3.2 A**). This agrees with the crystal structures of human protein [90,91]. From this we conclude that for split resonances, the one with a smaller chemical shift perturbation (CSP) corresponds to the open site, while the residues in the closed site show larger CSP. Further addition of cap analogue to DcpS^{H268N} reveals a two-step binding mechanism. At equimolar concentrations of enzyme and substrate, approximately one binding site per dimer is occupied as detected by peak splitting. With an excess concentration of substrate, this is subsequently bound by the second binding site, which can be derived from the CSP of the peaks corresponding to the open site (**Figure 3.2 D-E)**.

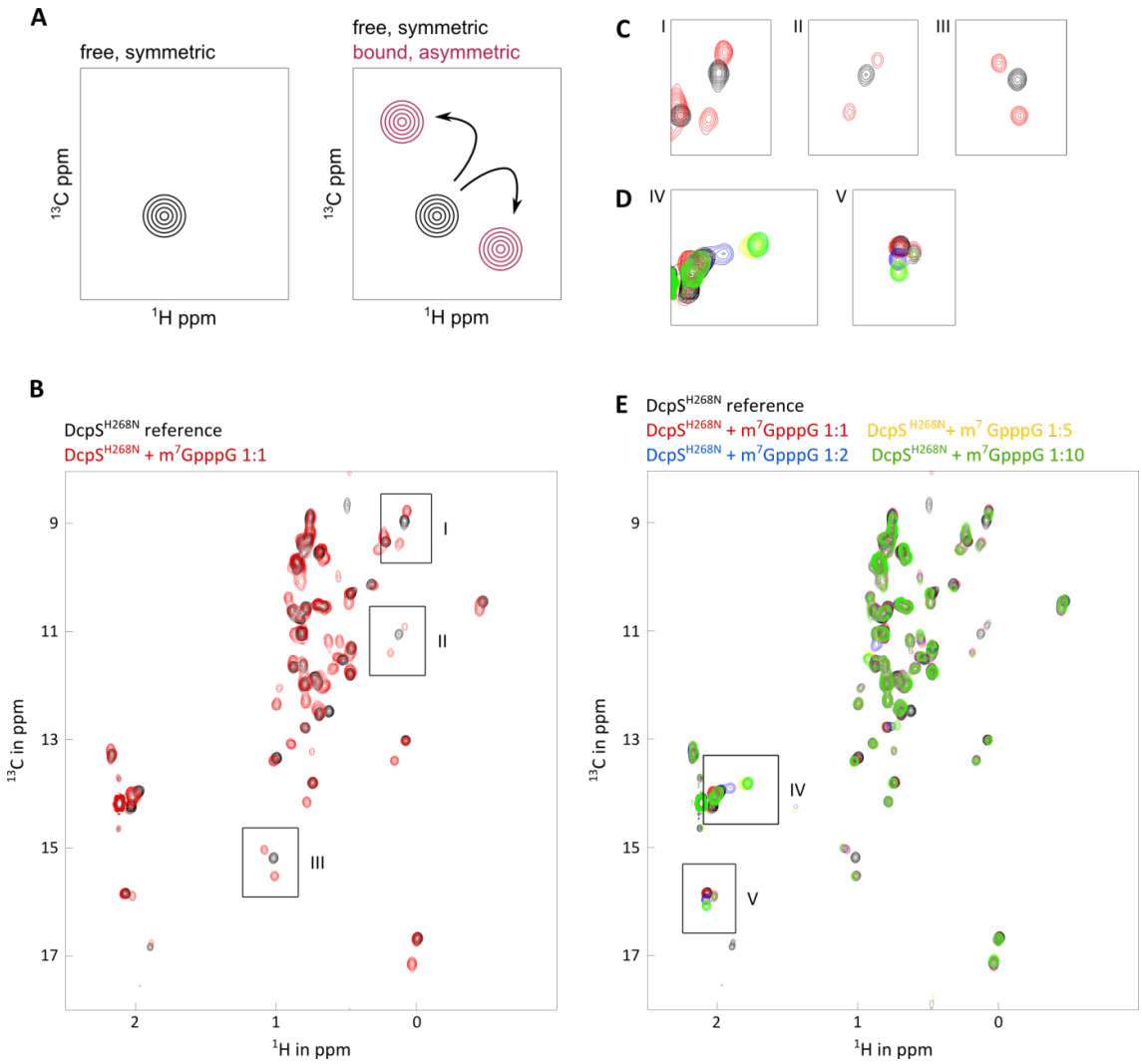


Figure 3.2 | The symmetry state of DcpS can be detected by resonance splitting in methyl-TROSY NMR spectra. **A:** For a symmetrical homodimer (black resonances), the equivalent residues of both protein chains result in the same resonance as they have an identical chemical environment. If the dimer adopts an asymmetric conformation upon substrate binding, residues that are affected by the conformational change no longer have an identical chemical environment and thus give rise to different resonance signals. **B:** NMR titration experiment of DcpS (apo, black) with an equimolar amount of m^{7}GpppG shows resonance splitting (red, zooms in **C**) and thus the change from the symmetrical apo conformation to the asymmetrical substrate bound conformation. **E:** Titration of DcpS with an excess of m^{7}GpppG shows CSP for residues sensing the binding of substrate in the open binding site (zooms in **D**).

From precise NMR titration experiments of DcpS^{H268N} with cap analogue and fitting of the resonance intensities with a sequential binding model, we could extract binding affinities for the two binding sides. The first binding site with a K_d = 68 ± 255 nM has a significantly higher affinity to the cap analogue than the second binding site with a K_d = 105 ± 20 µM. The values derived from NMR experiments were independently confirmed by isothermal titration calorimetry (ITC). For the single domains, the affinity of cap analogue to the C-terminal domain is similar to that of the second binding site which confirms that there are few interdomain contacts and that the second substrate interacts mainly with residues from the C-terminal domain upon initial binding.

Furthermore, the affinity of the product m^7GMP to DcpS was determined using NMR and ITC. To the closed binding site, the product binds with a K_d = 1.1 ± 0.6 µM, which is four times weaker than the affinity of the substrate to the closed site. To the open binding site, no interaction of the product could be detected. This difference in affinity enables that after the N-domain flipping over to enclose the substrate in the second binding site, the affinity to the product in the now open, first site is low enough for product release and a new substrate molecule can bind.

To obtain detailed atomic resolution insights into the yeast protein for interpretation of the NMR data, we determined the structure of $DcpS^{H268N}$ in complex with the inhibitor m^7GDP by X-ray crystallography (**Figure 3.3 A, B**). As for the substrate bound structure of inactive *H.s.* $DcpS^{H277N}$, the inhibitor bound structure of *S.c.* DcpS is also in the asymmetric conformation. The structure was determined with a resolution of 2.3 Å and superposes well with the structure from human protein, with the exception of loop regions that are not conserved. In contrast to the structure of *H.s.* DcpS, the yeast structure is in complex with m^7GDP instead of m^7GpppG. Despite of this difference, m^7GDP in *S.c.* DcpS has the same interaction of the cap guanine with W161 as the substrate in *H.s.* DcpS (**Figure 3.3 C, D**).

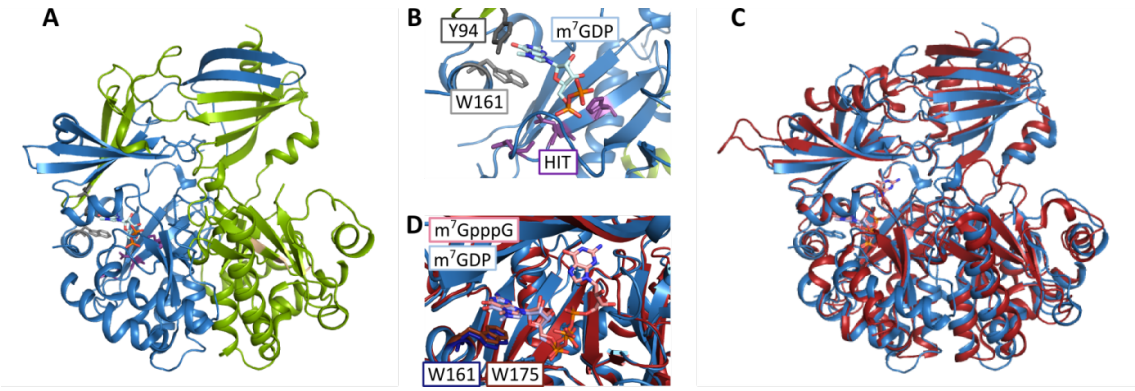


Figure 3.3 | The structure of *S.c.* DcpS in complex with m^7GDP is homologous to the structure of human enzyme. **A:** structure of *S.c.* DcpS in complex with m^7GDP determined in this study. The complex crystallised in the asymmetrical conformation. **B:** The cap guanine forms a stacking interaction with W161 and Y94 interacts with the N7 methylation. **C:** *S.c.* DcpS superposes with the human enzyme in complex with m^7GpppG (pdb excession code 1ST0) with an rmsd of 1.275. **D:** The orientation of the cap structure with DcpS is the same for human and yeast enzyme.

From the structure of $DcpS^{H268}$ in complex with m^7GDP it is apparent that there is a close contact between the cap-methylation and the aromatic ring of residue Y94, which seems essential for substrate recognition. To further investigate the role of the interaction between Y94 and the N7-methyl group and its possible involvement in the closing mechanism of the active site, we employed two strategies. To evaluate the importance of Y94 for the catalytic competence of DcpS, we recorded methyl-TROSY NMR spectra of Ile and Met ^{13}C-labelled $DcpS^{H268N\ Y94A}$ titrated with m^7GpppG. Here, no resonances of the asymmetric conformation

could be detected. For the reverse experiment of a titration of DcpS^{H268N} with the unmethylated cap analogue GpppG, the enzyme also remains in the symmetric conformation. This shows that the interaction of the cap methylation withY94 is indeed essential for the domain motion.

To determine catalytic turnover rates of DcpS, we turned to ^{31}P NMR spectroscopy. Catalytic turnover of cap analogue substrates can be monitored by the distinct 1d ^{31}P spectra of substrate and product (**Figure 3.4 A, B**). To determine the turnover rate of DcpS with m^{7}GpppG, successive ^{31}P NMR spectra were recorded during substrate turnover. From the signal intensity of substrate and product, a catalytic turnover rate of 0.27 s^{-1} was extracted (**Figure 3.4 C**). In contrast to WT DcpS with m^{7}GpppG, WT DcpS showed no catalytic activity for GpppG nor did DcpS^{Y94A} for m^{7}GpppG. This confirms that the interaction of the N7 methyl group of the cap with Y94 is essential for the formation of the closed active site.

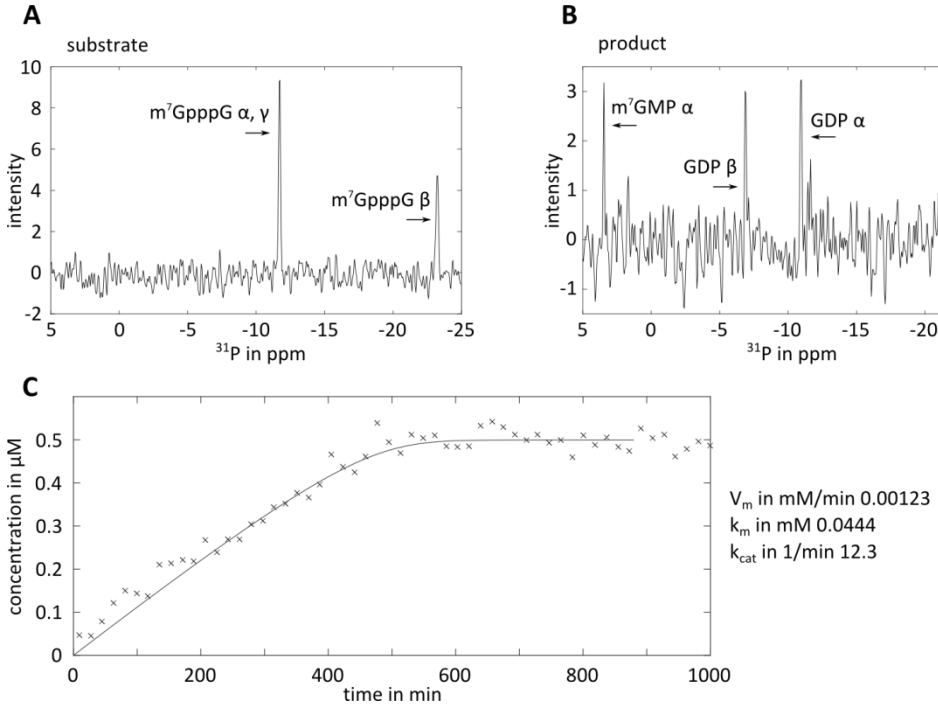


Figure 3.4 | Activity assays of DcpS by ^{31}P NMR. A, B: Cleavage of m^{7}GpppG by DcpS can be monitored by ^{31}P NMR due to different 1d spectra of substrate (**A**) and products (**B**). **C:** Turnover rates are determined by fitting of intensity decrease of substrate signals and increase of product signals. The turnover rate of DcpS for m^{7}GpppG is 12.3 min^{-1}.

Titration experiments are a powerful tool to determine the conformational state of proteins upon interaction with binding partners or substrates. However, no exact information about the actual dynamic rates of the domain motions can be extracted from these experiments for DcpS. Depending on the exchange rate of dynamic states, different NMR techniques are used to observe protein dynamics. In the case of DcpS, longitudinal exchange experiments were used,

which can probe motions in the ms to s range [116]. In longitudinal exchange experiments with ^{13}C-labelling of Ile and Met residues, the magnetisation is transferred from ^{1}H to ^{13}C, evolves in this dimension during a delay time, is then transferred back to ^{1}H and recorded. If the conformational state of the protein changes during the delay time between ^{13}C and ^{1}H dimension, this results in cross peaks where the ^{13}C-resonance of state A and the ^{1}H-resonance of state B are connected or vice versa (**Figure 3.5 A**). With this experiment, we could analyse the domain motions of DcpS induced by interaction with $m^{7}GpppG$. From experiments with different delay times, flipping rates and population ratios were extracted on the basis of resonance intensities of the cross peaks. As expected, both the symmetrical and asymmetrical conformation are populated to the same extend. The domain motions were determined to a flipping rate of 35.4 ± 4.8 s^{-1} at a 20-fold substrate excess (**Figure 3.5 B**). Consistent with a global domain motion, all residues that were analysed showed the same exchange rates. Interestingly, the domain flipping motion is two orders of magnitude faster than the turnover rate and influenced by the occupancy of the second binding site. Moreover, longitudinal exchange experiments with varying substrate concentration shows that with decreasing occupation of the second binding site, the flipping rate decreases. This indicates an allosteric pathway connecting both binding sites where binding of substrate in the open binding site induces a conformational change that closes the open binding site and vice versa.

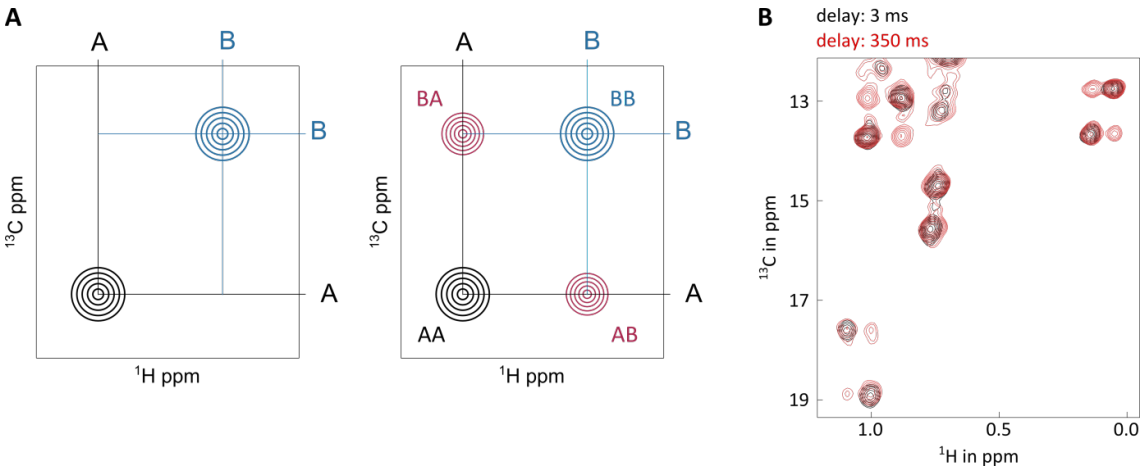


Figure 3.5 | Determination of dynamics rates of DcpS by longitudinal exchange NMR experiments. A: For a protein with two different conformational states A and B whose residues give rise to distinct resonances (black and blue), longitudinal exchange experiments result in cross peaks (red) for conditions where the magnetisation is transferred from the carbon dimension to the proton dimension of different states during the delay time of the experiment. **B:** DcpS shows dynamics induced by the substrate $m^{7}GpppG$ as indicated by the appearance of cross peaks (red). The flipping rate was determined to 35.4 ± 4.8 s^{-1}. (Spectrum adapted from reference Neu *et al.* in list of published work.)

The hinge region of DcpS exhibits several positively charged residues, e.g. K126, that are likely to interact with the negatively charged phosphate bridge of the substrate and presumably are involved in the allosteric pathway connecting the N- and C-terminal domains. To investigate their influence on the domain flipping motion, the hyperactive mutant $DcpS^{K126A}$, designed to alter the rates of the domain motion, was used for further longitudinal exchange experiments. The turnover rate for $DcpS^{K126A}$, determined by ^{31}P NMR, is twofold increased compared to WT enzyme, with 0.48 s^{-1}. No domain motions could be detected for $DcpS^{K126A}$ and thus must exceed the resolution range of longitudinal exchange NMR experiments with a flipping rate of < 1 s^{-1}, while the affinity for the substrate did not change substantially compared to WT enzyme. This counterintuitive behaviour of faster turnover with slower domain motions shows that the turnover rate of DcpS is directly inversely dependent on domain flipping rates. In hyperactive mutants, higher turnover rates are achieved by slower domain motions, as here the active site is closed long enough to allow for catalysis.

Taken together, in this study we could show that the conformational changes necessary for catalytic activity of DcpS are influenced by an allosteric pathway. In high excess of substrate, the enzyme is substrate inhibited by the increased speed of the domain motions leading to the formation of a closed, catalytically competent conformation for too short a time to allow for catalysis. In the inversion of the argument, the hyperactive mutant $DcpS^{K126A}$ in fact shows slower domain motions. This is an unprecedented mechanism of the involvement of domain motions that are necessary for catalysis but also inversely proportional to the catalytic rate of an enzyme.

3.2 A General Method for Rapid and Cost-Efficient Large-Scale Production of 5′-Capped RNA

All experiments described above were performed with a commercial cap analogue, m^7GpppG, which resembles the cap structure and the first base of the RNA body. DcpS was previously reported to catalyse most efficient on substrates in the range between one and ten nucleotides in the RNA body [89], however, only the two extremes of this range have been looked at. To address the activity of DcpS with capped RNA between one and ten nucleotides, we aimed at performing activity assays with capped RNA substrates of different lengths. Such RNA is, however, not commercially available nor are there protocols available to produce capped RNA in amounts necessary for enzymatic assays and structural studies. Thus we here established the prerequisite method for the production of capped RNA in the amounts required for biochemical and biophysical experiments by enzymatic capping of *in vitro* transcribed RNA with capping enzyme of vaccinia virus. The vaccinia capping enzyme is a complex of two proteins, wherein the 97 kDa D1 protein harbours the three catalytic activities necessary for the formation of a cap0 structure, being a triphosphatase, a guanylyltransferase and a N7-methyltransferase activity. The 33 kDa D12 protein is an accessory protein relevant for the stability of D1 and enhances the activity of the methyltransferase domain [43].

Genes coding for the D1 and D12 proteins of vaccinia were codon optimised for expression in *E. coli* and both proteins were co-expressed from one plasmid. To that end both gene sequences were cloned into a pRSF-duet vector, a plasmid with a promoter for each ORF. The complete expression cassette of this vector was subsequently transferred into a modified pet vector which resulted in higher expression yields. BL21 Gold (DE3) pLysS cells were used for expression to avoid protein expression prior to induction due to the toxicity of capping enzyme activity for bacterial cells. As bacterial mRNA is not capped, *E.coli* cannot cope with large amounts of capping enzyme during growth phase. For purification purposes, a hexa-histidine tag was attached to the N-terminus of D1. Ni-affinity chromatography was used as a first purification step and afterwards size exclusion chromatography was performed for further removal of impurities as well as the removal of cellular RNase that would impede the stability of RNA in capping reactions performed with this enzyme.

RNA substrates for the production of capped RNA were made by *in vitro* transcription with T7 RNA polymerase and plasmid DNA or DNA primers as templates. To produce RNA starting with G

or A, the φ 6.5 or φ 2.5 promoter of T7 were used, respectively [117]. To ensure the homogeneity of the final product, RNA transcripts were separated from abortive fragments and reaction side products by ion exchange chromatography under denaturing conditions. In small scale test reactions, purified capping enzyme was used in different ratios to determine the efficiency of the capping reaction and reaction products were monitored by Urea-PAGE (**Figure 3.6 A**). The most efficient ratio of enzyme to RNA, regarding amounts of enzyme, depends on the RNA substrate and varies between 1:10 to 1:500 for reactions of 1.5 hours. The capped RNA was further purified by removal of the capping enzyme by heat precipitation and removal of remaining reaction and buffer components by precipitation with ethanol and a subsequent desalting step. To analyse the reaction products, mass spectrometry was performed on samples before and after a capping reaction with a mixture of mono-, di- and triphosphorylated RNA as substrates for the reaction. From mass analysis we could not only verify that the vaccinia capping enzyme efficiently produces cap0 modifications, but also that tri- as well as di-phosphorylated RNA is a substrate for the capping enzyme, but not monophosphorylated RNA (**Figure 3.6 B**). This is in agreement with the reaction pathway, which involves a 5' diphosphorylated RNA as a reaction intermediate.

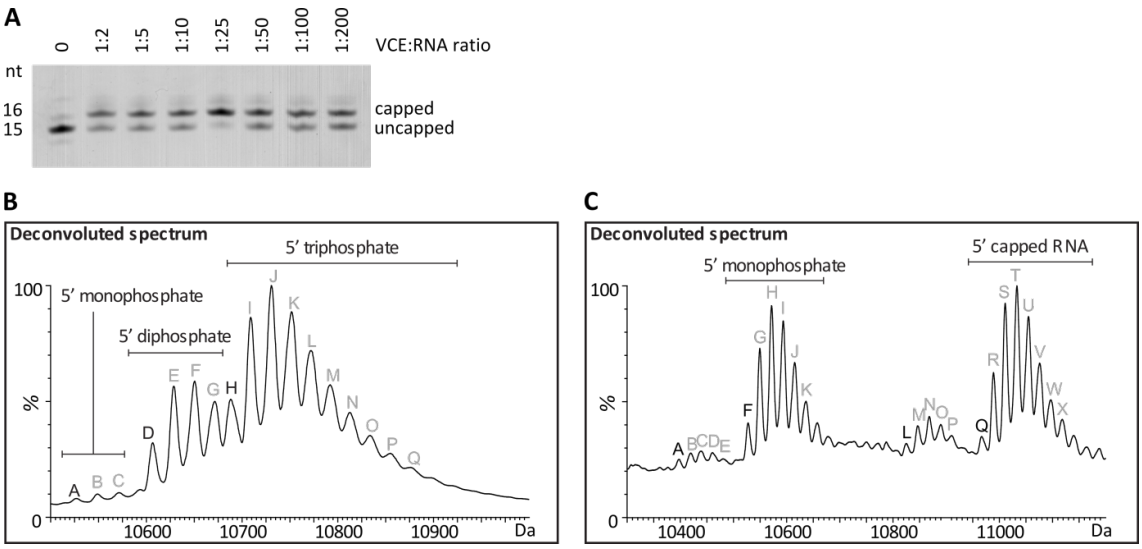


Figure 3.6 | Reaction of *in vitro* transcribed RNA with vaccinia capping enzyme results in a 5' cap0 structure. **A:** Reaction of vaccinia capping enzyme with 15 nt RNA in different ratios analysed by Urea-PAGE shows different reaction efficiency. **B:** A mixture of mono-, di- and triphosphorylated RNA was incubated with vaccinia capping enzyme and products analysed by mass spectrometry. Vaccinia capping enzyme can efficiently cap di- and triphosphorylated RNA but not monophosphorylated RNA. (Mass spectra adapted from reference Fuchs and Neu *et al.* in list of published work.)

Although the vaccinia capping enzyme has been studied for a long time and the reaction mechanism is known [118–124], information about substrate specificity of the capping reaction is lacking. To characterise on what RNA species the enzyme is active, we performed capping

reactions with RNA substrates of different length, sequence, and secondary structure. In summary, RNA in different length from 2 nucleotides on could be capped efficiently, showing that the enzyme performs independent of the substrate length (**Figure 3.7 A**). To test for a potential sequence preference, reactions with GA and CU rich sequences of 15 nucleotides were performed and showed equal efficiency (**Figure 3.7 B**). Also RNA starting with G as well as A at the 5' end were tested to exclude a specificity for the first base, and both could be capped efficiently by the vaccinia capping enzyme. To test for the influence of secondary structures on the capping efficiency, we designed RNA with a stable hairpin structure and varying length of 5' overhang adjacent to the hairpin of 0, 2, and 4 nucleotides overhang. RNA with 2 and 4 nucleotides overhang at the 5' end are capped efficiently. A completely basepaired 5' end cannot be capped showing that the minimal requirement of substrate for the vaccinia capping enzyme is an unpaired overhang at the 5' end of 1 to 2 nucleotides (**Figure 3.7 C**). The capping protocol can also be modified for the production of very short capped RNA of up to 5 nucleotides. As the transcription yield for short RNA is usually low, long RNAs are transcribed, capped and afterwards cleaved to the desired length with sequence specific RNase like RNase A, which cleaves after C and U bases (**Figure 3.8**).

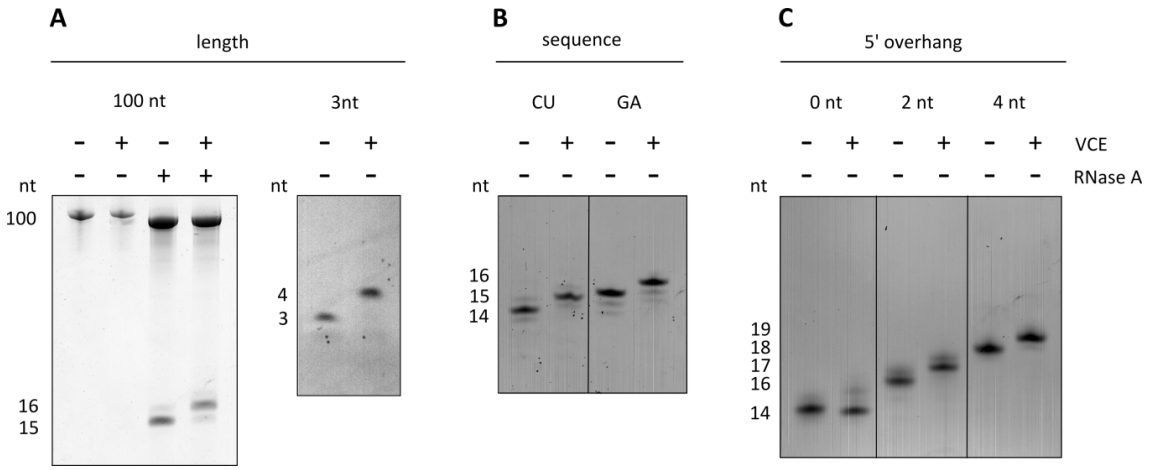


Figure 3.7 | Characterisation of possible substrates for the vaccinia capping enzyme. RNA is capped efficiently independent on length or base composition, but needs a 5' overhang. A: RNA of 100 nt and 3 nt length was efficiently capped by vaccinia capping enzyme as shown by Urea-PAGE analysis. 100 nt long RNA was cleaved with RNase A at position 15 for analysis purposes to be able to distinguish uncapped and capped RNA on the gel. **B:** RNA with CU- or GA-rich sequence is capped efficiently. **C:** Hairpin RNA with a 5' overhang of 2 or 4 nt is capped efficiently, whereas poor capping efficiency is detected for hairpin RNA without a 5' overhang.

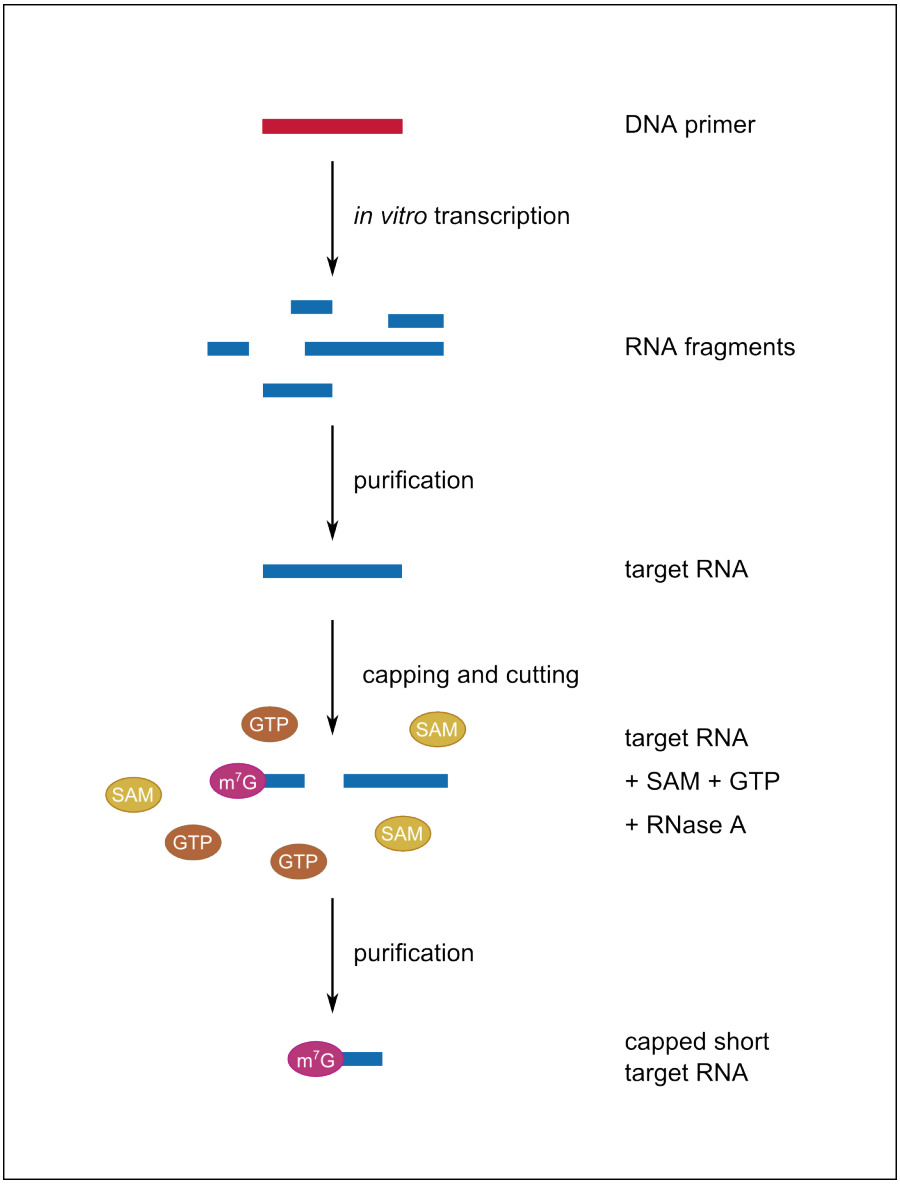


Figure 3.8 | Schematic of the protocol for the production of short capped RNA in large quantities. A longer RNA than the target RNA is produced by *in vitro* transcription using DNA primers as templates. After purification by ion exchange chromatography, the RNA is capped with vaccinia capping enzyme and cleaved to the target length with RNase at a specific cleavage site. Short capped RNA and cleaved 3' ends are separated by ion exchange chromatography.

We used interaction partners of the cap structure in a series of experiments to benchmark the applicability of RNA capped with the protocol presented here. First, translation efficiency was tested in an *in vitro* translation assay using wheat germ extract, where capped RNA showed a significantly higher translation rate compared to uncapped RNA. In assays with the decapping enzyme Dcp2, the cap structure attached by the vaccinia capping enzyme is efficiently recognised and removed, confirming that the protocol produces the physiological cap structure. Furthermore, the eukaryotic translation initiation factor eIF4E that binds strongly to the cap structure during translation can be used to separate capped from uncapped RNA by a specific

and tight interaction with capped RNA only. eIF4E was further used in exemplary binding studies by fluorescence anisotropy combining the capping protocol with fluorescence labelling of RNA. In fluorescence anisotropy measurements we determined a K_d = 294 ± 42 nM of eIF4E to capped RNA of 30 nucleotides, which is in agreement with previously reported data on the binding affinity of eIF4E to m^7GDP of K_d = 726 ± 122 nM determined by ITC [125]. The capping protocol can also be combined with stable isotope labelling for NMR studies by using for example ^{13}C-labelled SAM for the capping reaction. This results in a cap0 structure specifically labelled at the N7-methyl group, which enables observation of interactions of the cap structure specifically. Furthermore, we show that the protocol can be expanded to a cap1 modification by a subsequent reaction with the vaccinia VP39 protein, the methyltransferase responsible for 2' OH methylation of the first ribose. We performed exemplary NMR titration experiments of a cap1-RNA, ^{13}C-labelled at both methylgroups of the cap1 structure, with NMR-invisible eIF4E. Spectra of the RNA alone showed two resonances, one for each methyl group of the cap1 structure. Upon interaction with eIF4E, only the resonance corresponding to the 2'-O methylation remains, while the one corresponding to the N7-methylation is broadened beyond detection by the interaction with the cap-binding protein (**Figure 3.9 A**). This shows that the interaction of eIF4E with capped RNA is solely facilitated by the N7-methylation of the cap-guanosine, whereas the 2'-O methyl group is not affected by binding of eIF4E. The experiment was also recorded vice versa with unlabelled cap1-RNA and ^{15}N-labelled eIF4E. Upon titration with capped RNA, the spectrum of eIF4E shows CSP in a set of resonances indicating the interaction of the protein with the RNA (**Figure 3.9 B**). A combination of biochemical and biophysical experiments can thus elucidate the interaction specificity of potential cap binding proteins.

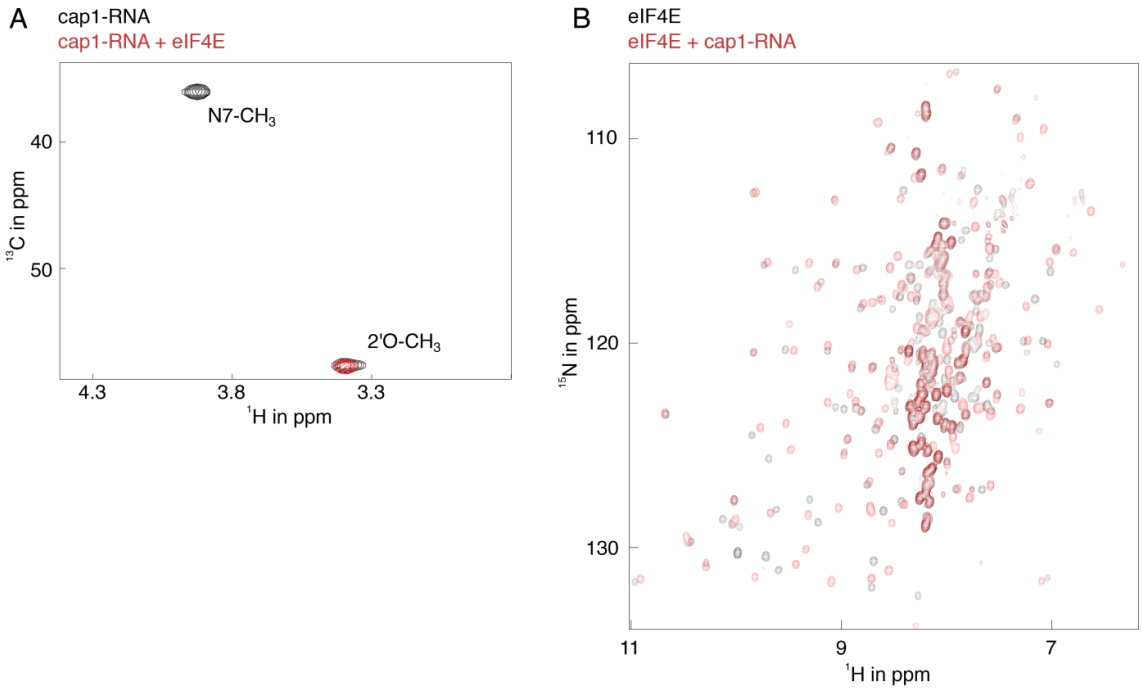


Figure 3.9 | NMR titration experiments of capped RNA with eIF4E. A: RNA with a 5' cap1 ^{13}C-labelled at N7- and 2' O-methylation (black) is titrated with eIF4E (red). The interaction is mediated solely through the cap guanine and the resonance signal for the N7-methylation is broadened beyond detection upon interaction with eIF4E. **B:** ^{15}N-labelled eIF4E (black) is titrated with cap1-RNA (red). CSP in the spectrum of eIF4E show binding of the RNA to the protein. (Spectra adapted from reference Fuchs and Neu *et al.* in list of published work.)

In summary, we here established a method for the large scale production of capped RNA from *in vitro* transcribed RNA. The protocol reliably adds a cap0 modification to the 5' end of RNA largely independent on RNA length, base composition, and secondary structures of the RNA. The method can be combined with other protocols for fluorescent labelling of the RNA or isotope labelling for use in NMR spectroscopy. The yields are solely dependent on the amount of starting RNA and large enough for the use of the capped RNA in biochemical and structural applications like crystallography or NMR spectroscopy.

3.3 Molecular Basis for the length sensing by the Scavenger Decapping Enzyme DcpS

The turnover efficiency of DcpS was previously shown to be dependent on the length of the RNA substrate. Liu *et al.* performed degradation experiments with RNA of different length ^{32}P-labelled at the α-phosphate of the triphosphate linkage (m^7G*ppp-RNA, with 1, 10, 15 and 20 nucleotides in the RNA body). They found the highest activity for cap analogue, decreased activity for 10 nucleotides and only residual activity for 15 nucleotides. For 20 nucleotides no activity was detected [89].

We here study the apparent substrate length dependency of DcpS by interaction studies with capped RNA of different length. To that end, we made use of our method for the production of capped RNA of any length. We prepared capped RNA of 2, 3, 5, 15 and 31 nucleotides and performed activity assays under substrate excess conditions (**Figure 3.10**). Substrates were chosen to cover the range of length between 1 and 15 nucleotides and also expand the range of the previous experiment to test the behaviour of DcpS on longer RNA. We then compared those to the turnover rate we observed for cap analogue, which is rated as the model substrate for DcpS in many studies [91,94,100,102,126].

The highest activity was observed for cap analogue with a turnover rate of 26.9 ± 2.3 min^{-1}, which is in agreement with the data we previously determined by ^{31}P-NMR methods. With only one nucleotide more in the RNA body, the activity drops to 6.4 ± 2.2 min^{-1} for a capped dinucleotide, which signifies a loss in activity by 76 %. For three nucleotides in the RNA body the activity was reduced to <1 min^{-1}, what we regard as basal activity. For all longer RNA substrates tested, we found the same unchanged low activity as for three nucleotides. Thus the activity of DcpS shows a clear threshold between two and three nucleotides. However, the residual activity we found for substrates longer than two nucleotides demonstrates that the enzyme is in general able to perform catalysis also on RNA longer than two nucleotides. As the residual activity is unchanged from three to 31 nucleotides, we reasoned that the mechanism by which DcpS is inhibited by longer substrates relies on the third nucleotide of the RNA and that nucleotides further apart do not contribute to the inhibition. The reason for the difference in results from Liu *et al.* regarding the threshold of activity is most likely due to different reaction conditions. Under enzyme excess as in the experiments by Liu *et al.*, DcpS is not substrate inhibited (see Chapter 3.1) and could accept substrates that are not used under excess of more favoured substrates.

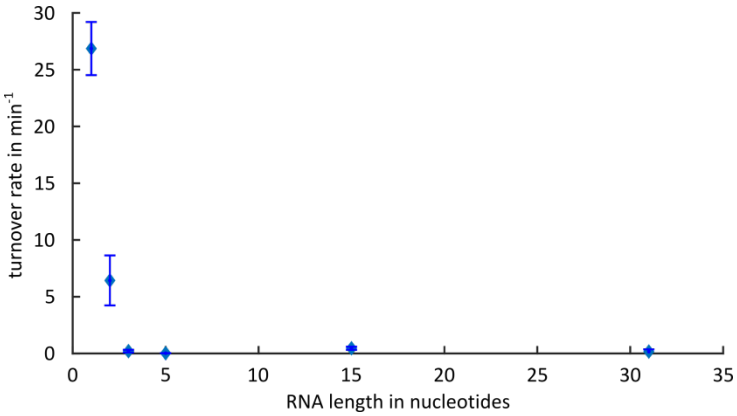


Figure 3.10 | Length dependency of DcpS turnover rates. Activity assays of DcpS with RNA of 1, 2, 3, 5, 15 and 31 nt in length was performed and analysed by HPLC or Urea-PAGE methods depending on the size of the substrate. The turnover rate of DcpS is dependent on the substrate length and decreases with increasing length of the RNA. From 3 nt on, the enzyme shows only basal activity.

As shown in Chapter 3.1, the formation of an asymmetric conformation is a prerequisite for catalysis by DcpS. Interference with this domain motion might be the molecular reason for the inhibition of catalytic activity by RNA substrates longer than two nucleotides. Thus, we analysed the conformational state of DcpS upon interaction with the respective substrates by performing methyl-TROSY NMR titrations with capped RNA of two and three nucleotides. Spectra of DcpS in complex with cap analogue served as a comparison, as the enzyme is known to adopt a fully asymmetric and stable conformation with this substrate (**Figure 3.2 B** in Chapter 3.1). Titration with a capped dinucleotide results in spectra that overlap well with those of titrations with cap analogue, albeit with weaker intensities for residues in the closed site (**Figure 3.11 A**). This shows that DcpS adopts an asymmetric conformation upon binding of capped dinucleotide, but the conformation is probably less stable compared to cap analogue. The same experiment was repeated for capped trinucleotide. Here the CSP differ from the titration with cap analogue. Analysis of the CSP shows no presence of a stable asymmetric conformation. However, CSP that indicate binding of the cap structure in the open active site could be observed (**Figure 3.11 B**). This shows that the capped trinucleotide is able to bind to the enzyme, but does not induce a conformational change. Without closing of the active site, however, catalysis is not possible. The lack of catalytically competent closed conformation with a capped trinucleotide as substrate explains the drastic decrease in activity with longer substrates. However, residual activity is observed also for RNA substrates of three nucleotides and longer as stated before. Because DcpS is only catalytically competent in the closed conformation, this conformation has to exist at least transiently. However, the population of the closed conformation is below the detection limit of NMR and no peaks for the asymmetric conformation can be detected. The threshold in length of RNA that induces a stable asymmetric conformation between two and three nucleotides is in

agreement with the results of the activity assays which also showed a threshold in substrate turnover between two and three nucleotides in the RNA body.

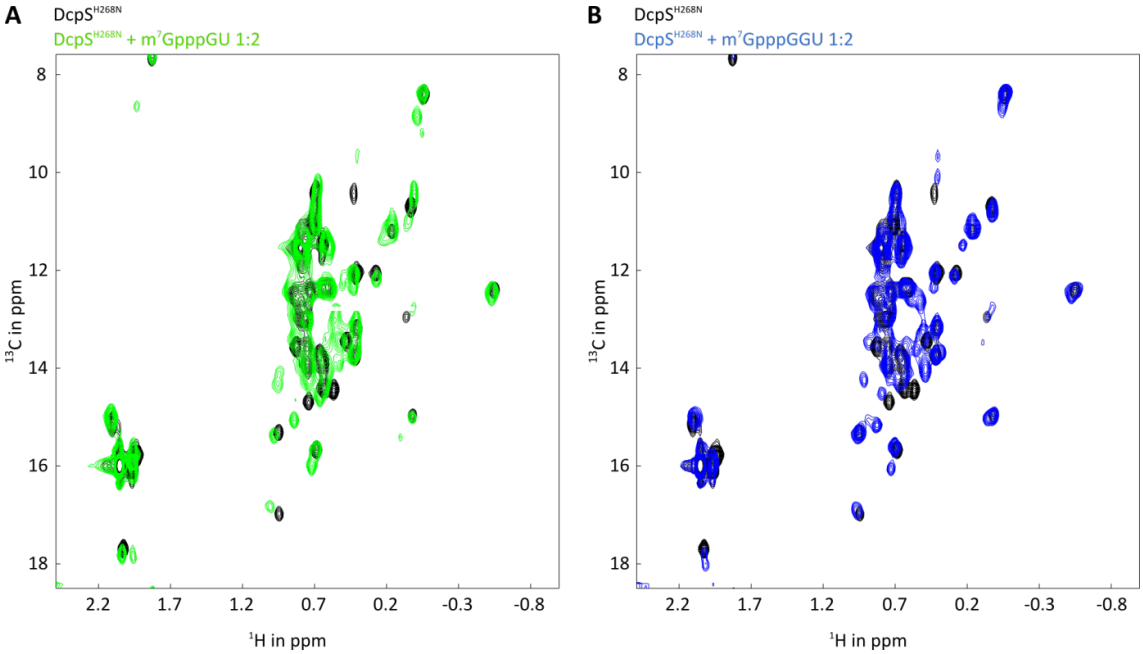


Figure 3.11 | NMR titrations of DcpS^{H268N} with capped RNA of two and three nucleotides. A: Titration of DcpS with capped dinucleotide shows resonance splitting, which indicates the adoption of an asymmetric conformation upon interaction with the substrate. **B:** Titration of DcpS with capped trinucleotide shows CSP indicating binding of the RNA in the active site but no resonance splitting and thus no asymmetric conformation.

As a capped dinucleotide is the longest RNA that is still able to induce the asymmetric conformation of DcpS, we used this complex for crystallisation assays. The complex of DcpS^{H268N} and capped dinucleotide could be co-crystallised and the structure determined with a resolution of 2.9 Å (**Figure 3.12 A**). The capped dinucleotide is located in the closed site of the dimer, while the open site is not occupied. The orientation of the cap structure is the same as compared to the previously determined structure of DcpS in complex with m^{7}GDP and also as in the structures of human enzyme in complex with different cap analogues with an rmsd of 1.275 Å from *S.c.* DcpS to *H.s.* DcpS. The first base of the RNA body is oriented similar to the one in the human structure of DcpS in complex with m^{7}GpppG. The second base points towards a void between the N- and C-terminal domains that has been proposed as a possible exit channel for longer RNA molecules based on the structure of *H.s.* DcpS with m^{7}GpppG [91] (**Figure 3.12 B**). The capped dinucleotide occupies the complete active site cavity of *S.c.* DcpS and leaves no space for a further nucleotide within the boundaries of the active site. A third nucleotide would thus interact with the proposed exit channel, which Gu *et al*. proposed would be wide enough to fit single stranded RNA. The channel is mainly lined by flexible residues that could adopt alternative rotameric states upon the need for more space due to larger substrates [91]. Using the structure of

DcpS in complex with a capped dinucleotide as a basis, we used the programme CYANA [127] to model a further nucleotide into the binding site to probe a possible interaction surface. All calculated structures with more than two nucleotides in the RNA body showed steric clashes of the RNA not only with side chains of residues in the channel, but also atoms of the protein backbone. Some residues that are affected are likely flexible enough to move aside for longer substrates but an energy barrier clearly has to be overcome in order to close the binding site around a substrate longer than two nucleotides (**Figure 3.12 C-E**). Thus we hypothesise that the mechanism of the length sensing in DcpS is based on the protein structure by steric hindrance of the conformational change by longer substrates. This finding also agrees with the data previously obtained by activity and dynamics experiments. DcpS is most active for cap analogue which fits into the active site without sterical hindrance. For decapping of RNA with a second nucleotide in the RNA body the activity is decreased about four times. The closing of the two domains of DcpS around a capped trinucleotide is energetically not favoured due to necessary conformational rearrangements and thus the catalytic activity drops to only a basal activity of $< 1\ min^{-1}$. The energy needed to overcome the steric hindrance and side chain conformational changes are apparently the same for all RNAs longer than two nucleotides, as the turnover rates remain unchanged for longer RNA. Thus although the proposed exit channel is the interaction surface of long RNA with DcpS as based on our modelling data, the channel is not wide enough to allow for interaction with RNA without conformational changes in the channel.

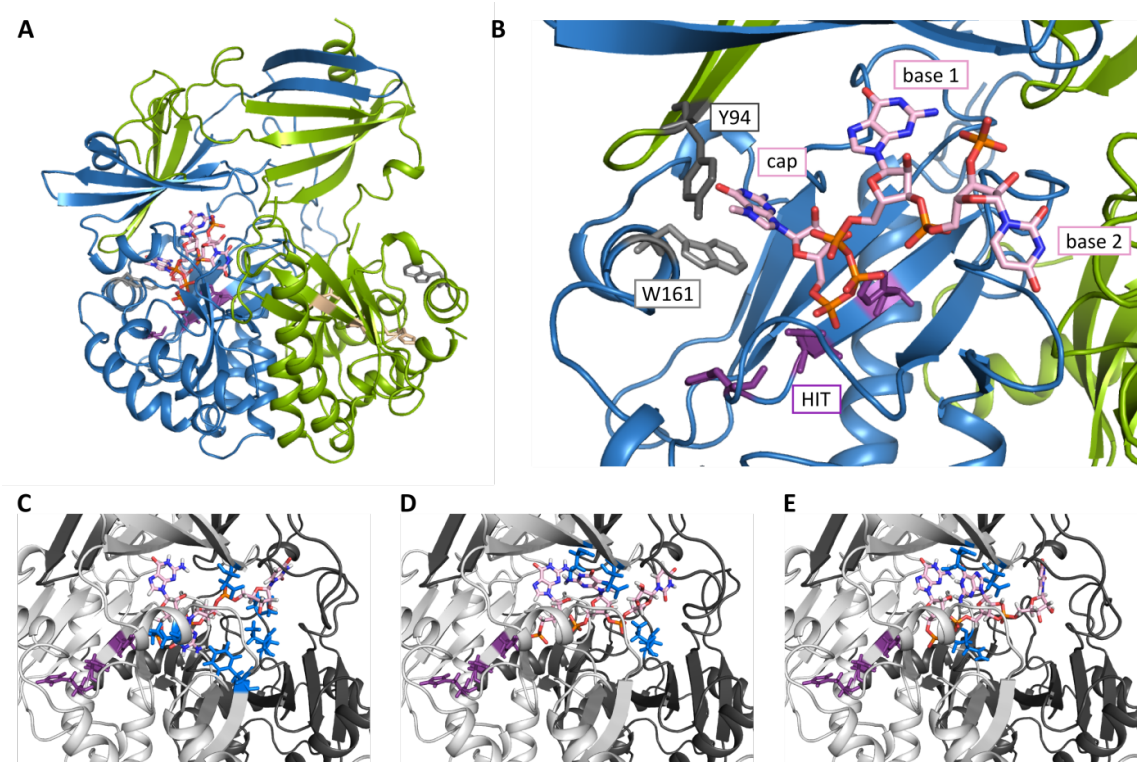


Figure 3.12 | RNA longer than two nucleotides interferes with the closing mechanism of the active site. A, B: Structure of DcpS in complex with a capped dinucleotide. The cap structure of the RNA shows stacking interaction with W161 and interaction of the N7 methylation with Y94. The RNA body is pointing towards a possible exit channel in between the N- and C-terminal domains. **C-E:** Three best models of computational modelling of three nucleotide long RNA into the active site based on the structure shown in **A** with CYANA. The third nucleotide experiences several clashes (blue residues) with the protein.

For eukaryotes, DcpS is not very highly conserved on a sequence level. Compared to the human enzyme, *S.cerevisiae* derived enzyme shows only 55 % sequence similarity (33 % identity) and *C. thermophilum* derived enzyme 51 % similarity (30 % identity). To test if the protein is nevertheless conserved on a structural level, we determined the structure of *C.t.* DcpS. The apo protein crystallised in the symmetrical conformation as the apo proteins of human and mouse and superposes well (**Figure 3.13**). The fold of the core protein, which is decisive for substrate interaction and includes the binding and catalytic site, is highly conserved among all species studied to date. To see whether DcpS from different species distinguish substrate length by the same mechanism, NMR titration experiments with RNA substrates of increasing length were performed with DcpS from *H. sapiens* and *C. thermophilum*. Interestingly, both enzymes showed the same substrate threshold as the *S.c.* DcpS enzyme: between two and three nucleotides (**Figure 3.14**). Although the intrinsic activity for *H.s.* DcpS, with a turnover rate of 60 min^{-1} for m^7GpppG, is different from DcpS, the mechanism by which an eligible substrate is recognised, seems to be a conserved mechanism. From this collection of biochemical, dynamical and

structural data we thus conclude that the mechanism by which DcpS recognises RNA as an eligible substrate is conserved.

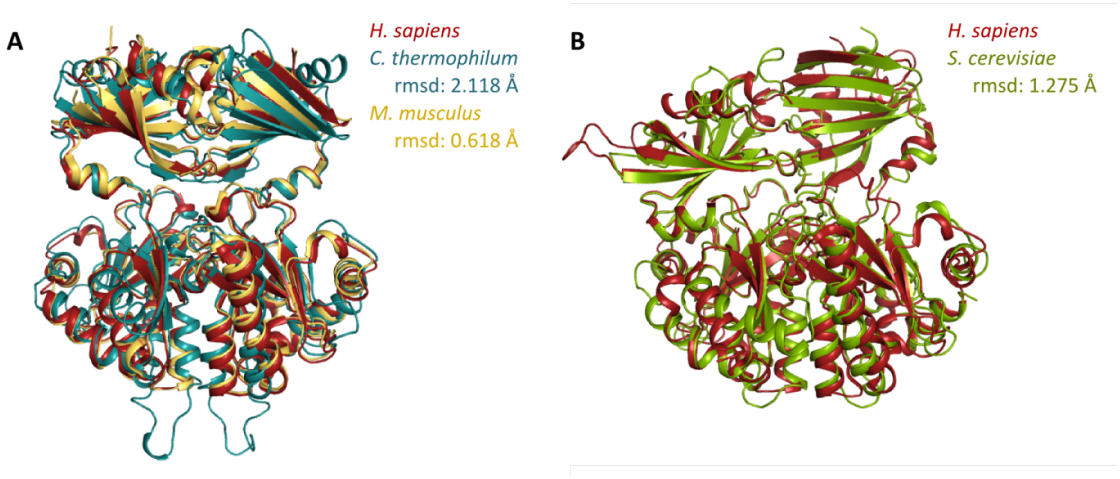


Figure 3.13 | The structure of DcpS is conserved among species. A: Structures of *H.s.* DcpS, *C.t.* DcpS and *M.m.* DcpS in the symmetrical conformation show high structural similarity. **B:** Structures of *H.s.* DcpS and *S.c.* DcpS in the asymmetrical conformation likewise show a high structural conservation.

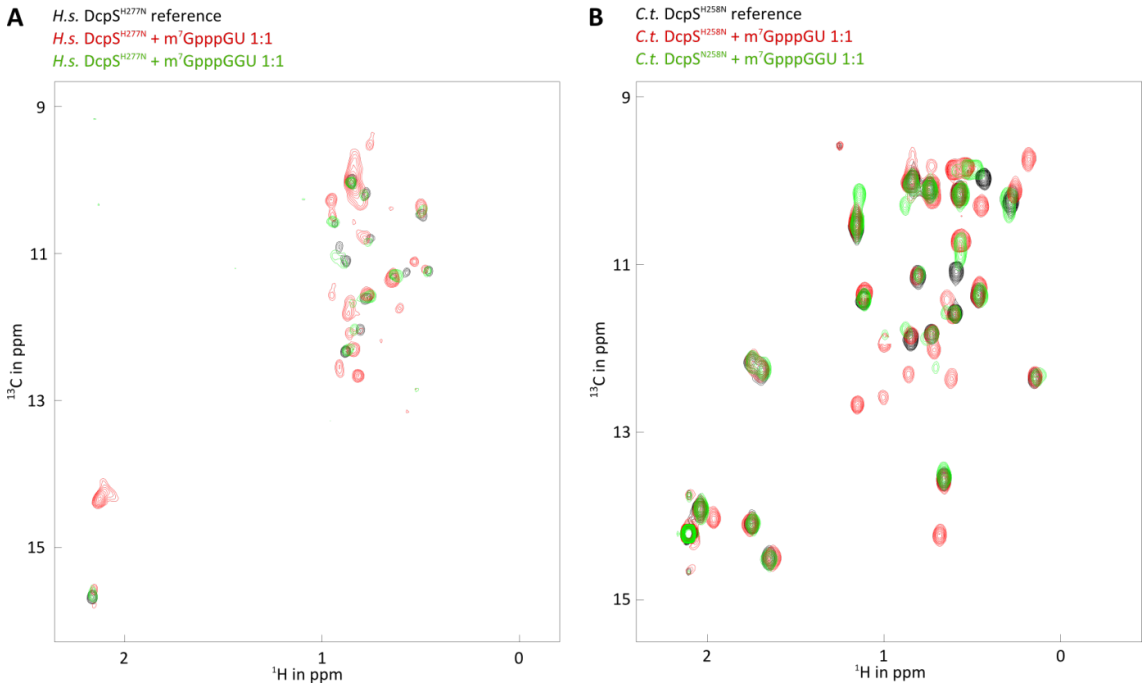


Figure 3.14 | NMR titration experiments of *H.s.* DcpS and *C.t.* DcpS with capped di- and trinucleotide. A: Titration of *H.s.* DcpS with capped dinucleotide shows resonance splitting indicative for the asymmetric conformation (red), whereas capped trinucleotide shows CSP indicating binding of the RNA but no conformational change (green). **B:** Titration of *C.t.* DcpS with capped di- and trinucleotide shows the same binding and conformational change behaviour as *H.s.* DcpS.

If an intrinsic feature of an enzyme is so well conserved, it is likely to serve an important function for the cell. Decapping by DcpS is the last step in the 3'-5' mRNA decay pathway. It is supposed to further degrade the products of RNA hydrolysis by the exosome. The length of those products was approximated to 4 to 5 nucleotides [87,88]. This product length, however, is in contradiction to the substrate preference we found for DcpS, which means that either the products of the exosome are shorter than reported or another enzyme further trims the short capped mRNA before it is handed over to DcpS. To accurately determine the product length of the exosome, we used recombinant Exo10 complex from *S. cerevisiae* and performed activity assays with capped 30 nucleotides long RNA as a substrate. Preliminary results of these experiments point towards a shorter product length than was previously reported. The length could not yet be determined exactly due to incompatibility of reaction and mass spectrometry conditions, but narrowed down to between one and two nucleotides in the RNA body (**Figure 3.15**). This length would allow for decapping by DcpS and thus a direct handover between the exosome complex and DcpS.

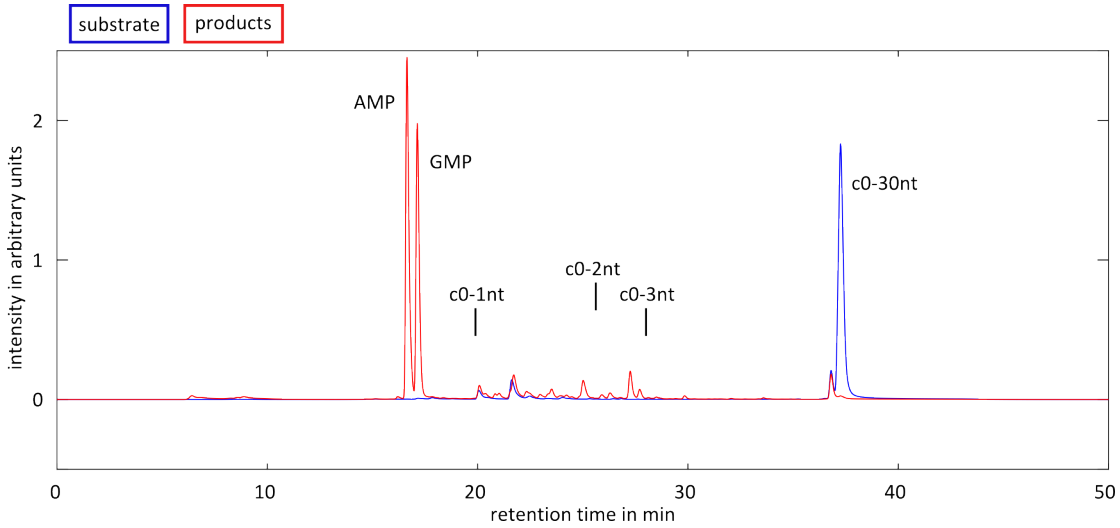


Figure 3.15 | RNA degradation reaction by the exosome complex. Preliminary ion exchange HPLC analysis of capped 30 nucleotide long RNA before (blue) and after (red) reaction with Exo10. Retention times of references are indicated. The length of the main products is probably between one and two nucleotides.

Taken together, we could, for the first time, study the interaction of DcpS with native substrates under label-free conditions. We could show a clear threshold in activity for substrates between two and three nucleotides in the RNA body. This preference in substrate usage can be explained on the structural level as only RNA of up to two nucleotides is able to induce the necessary conformational change to a catalytically competent conformation of DcpS. Longer RNA substrates would compromise the closing of the active site due to steric clashes with the enzyme. As DcpS is responsible for the decapping of exosome products that are most likely one

to two nucleotides long, this threshold in substrate usage is of relevance to prevent DcpS from decapping longer RNA that is not a product of exosomal degradation.

4 Discussion

4.1 DcpS as an Example for the Modulation of Enzymatic Turnover by Protein Dynamics

By a combination of protein crystallography for structure determination and NMR techniques for dynamics measurements, we here showed that the Scavenger Decapping Enzyme (DcpS) relies on domain motions to perform catalysis. The apo form of the homodimeric protein is in a symmetrical conformation. Binding of a substrate molecule induces a conformational change and the enzyme adopts an asymmetric conformation with one closed and one open binding site. Only the asymmetric conformation is catalytically competent. Structure determination by NMR of proteins larger than 30 kDa is very challenging and time consuming, and these larger systems are often only accessible for structural studies by crystallography or cryo electron microscopy (cryo-EM). These latter techniques, however, mostly display static snapshots of a protein in an artificial, solid environment. Liquid state NMR methods, on the other hand, can give exceptional insights into conformational changes of biomolecules on a per-residue level that can be linked to the static snapshots from crystallography or EM. In this thesis, we study the DcpS protein, and rely on crystallographic structures to interpret our dynamics NMR data. For this 80 kDa protein, conventional NMR labelling schemes with ^{15}N and ^{13}C result in low spectral quality. However, due to the favourable relaxation rates of methyl groups, ^{13}C-labelling can be employed to study large proteins. With NMR, there are several methods to observe protein motions on a wide temporal range, including T1 and T2 measurements, longitudinal exchange experiments, CPMG experiments, RDC or PRE measurements, which are suitable to detect and quantify dynamic processes with different timescales [116].

Many enzymes, including the DcpS enzyme that we study here, have been shown to undergo catalytically important domain motions. One example from our laboratory is the enzyme Dcp2, the mRNA decapping enzyme that is involved in 5'-3' mRNA degradation. This enzyme forms a complex with the decapping enhancers Dcp1 and Edc1. Upon recruitment of capped mRNA, the catalytic domain harbouring the active Nudix helix reorients with respect to the regulatory domain and is then in a catalytically active conformation [70]. These findings underscore the importance of studies that go beyond static structures derived from crystallography.

Although domain motions in DcpS are necessary for catalysis, we found that these dynamics accelerate with increasing substrate concentrations such that efficient catalysis is no longer possible. This inhibition of the enzyme most likely results from instabilities in the active site that is under fast domain motions not closed for long enough to allow for efficient substrate turnover. Such substrate inhibition mechanisms are a well-known concept in biology with about 20 % of enzymes showing decreased activity at high substrate concentrations. The function and mechanism of substrate inhibition are of very different nature depending on the particular enzyme. As an example, tyrosine-hydroxylase catalyses the reaction of tyrosine to L-Dopa in the synthesis of dopamine. Here, substrate inhibition has the function of keeping the production of L-Dopa at a constant rate although tyrosine concentrations vary during the day in dependence of meals. Acetylcholinesterase, a structurally unrelated enzyme, degrades acetylcholine after its release into the synaptic cleft. It is a very efficient enzyme and has to clear the synaptic cleft fast after the end of pre-synaptic acetylcholine release. Nevertheless, it must not be too fast as otherwise acetylcholine could not reach the post-synaptic membrane and fulfil its signalling function. To regulate acetylcholine concentrations in the synaptic cleft efficiently, acetylcholinesterase is inhibited by the fast increase in substrate, but is very effectively degrading once the acetylcholine concentration is reduced at the end of acetylcholine release [128–132]. Phosphofructokinase is another example for substrate inhibition. It is inhibited by ATP, which on the one hand is a substrate, but on the other hand also the end-product of the whole metabolic pathway in which the enzyme is involved. Unless the ATP concentration is low and thus substrate inhibition overcompensated, phosphofructokinase can only be stimulated by activators like insulin that can outcompete the substrate inhibition and ensure the clearance of sugar from the blood [128,133]. These three examples show how substrate inhibition is involved in the fine tuning of very different pathways and reactions. The importance of the substrate inhibition of DcpS, however, is not yet fully established.

A conceivable function of the substrate inhibition of DcpS is to limit the amount of m^7GMP produced in the cell. An accumulation of the product m^7GMP could lead to its incorporation into newly synthesised DNA or RNA. Thus the DcpS product should be degraded or demethylated at some point. In human and *D. melanogaster*, nucleotidases have been described that remove the phosphate group of m^7GMP. However, in *S. cerevisiae*, m^7GMP is converted into a yet unknown compound and might rely on an alternative degradation pathway [134]. In case the enzymes processing m^7GMP cannot be upregulated fast enough if high amounts of m^7GMP are produced by DcpS at once, substrate inhibition of DcpS would ensure the regulated degradation of the product, in analogy to the tyrosine-hydroxylase mentioned above. Furthermore, the decapping

of m^7GpppG results not only in m^7GMP, but also GDP. This needs to be recycled which involves phosphorylation steps and thus the hydrolysis of ATP. Hence, substrate inhibition of DcpS avoids the rapid increase of m^7GMP and GDP and thus the need for further processing.

4.2 DcpS Activity is Dependent on Substrate Length

The turnover rates of DcpS strongly depend on the substrate length, as it has an influence on the stability of the conformational state that is required for a catalytically competent active site. This active conformation is induced stably by substrates of up to two nucleotides in length. RNA with three or more nucleotides in the RNA body drastically reduces the activity of DcpS and the required asymmetric conformation is no longer stably induced. Modelling of a capped trinucleotide into the active site revealed several steric clashes for all possible orientations of the third nucleotide. To fit a longer RNA in between the N- and C-terminal domains of the closed conformation, at least side chains, most likely also the loop opposite the exit channel have to make way for the RNA body. The basal activity of DcpS for substrates of more than three nucleotides show that a catalytically active conformation of DcpS can be induced by longer substrates and that the exit channel can thus be transiently enlarged. This structural rearrangement, however, appears to be associated with high energy costs.

A capped dinucleotide is decapped by DcpS, albeit with a reduced turnover rate. The slower rates can be due to two reasons. First, a capped dinucleotide has more degrees of conformational freedom than a capped mononucleotide. Thus the RNA has to be in a conformation that allows for optimal binding and closing of the active site, which is less likely for a capped dinucleotide than for a capped mononucleotide. This entropic effect can slow down the turnover rate.

Another possible reason for the decrease in activity for RNA longer than one nucleotide is the possibly self-intercalating nature of the capped RNA. There the cap-guanine can intercalate in between the first and second base of the RNA body. For this formation of secondary structure at least two nucleotides are needed in the RNA body [135]. In cases of self-intercalated cap structures, the RNA can be present in two different conformations of which only in the linear conformation the cap-base is accessible for binding by DcpS.

The aforementioned reasons could lead to a change in affinity for longer substrates and thus reduced activity. Binding affinities are only known for m^7GMP, m^7GDP and cap analogues. The determination of affinities for short RNAs is complicated by the fact that most methods for affinity measurements rely on the addition of labels to one of the binding partners. In the case of a capped di- or trinucleotide, this is not possible as a label-group would destroy the close fitting into the active site and interfere with the closing mechanism. With methods that use intrinsic label groups of the protein, like tryptophan-quenching or label-free MST, so far we did not yield consistent and reproducible data. Also label-based MST did not give reliable results, probably due to interaction of DcpS with the capillary material. Possible methods that do not rely on labels are NMR and ITC. Both methods require titration of the protein with a large amount of substrate. Although our protocol for the capping of *in vitro* transcribed RNA yields amounts sufficient for crystallisation or NMR titrations, the amounts required for affinity determination with NMR or ITC are expensive and time-consuming to produce for small RNA of only two or three nucleotides. The production and purification of these RNAs have such low recovery yields that affinity measurements are very challenging.

4.3 Placement of DcpS and its Products within the 3'-5' Degradation Pathway

DcpS removes the cap structure from short capped RNA fragments that are the product of RNA hydrolysis by the exosome complex. Such a close functional interaction would be promoted by complex formation of DcpS and the exosome. A complex of the two proteins has indeed been reported, albeit only a subset of the exosome complex including Rrp40 and Rrp41 was found in a 300 kDa complex containing DcpS activity [57]. For a direct handover of product from the exosome as substrate to DcpS, the length range of the two reactions has to be the same. For DcpS we found a preferred substrate usage for capped RNA of one and two nucleotides. The length of the exosome product has as of yet not been accurately determined. The shortest fragments that have been reported are 4-5 nucleotides long. The experiments were performed with fluorescently labelled, uncapped RNA in gel chromatography [87,88]. Here, two aspects could influence the length determination. First, these experiments were performed with uncapped RNA, so any influence that the cap structure may have on the product length was neglected. As the cap structure is farther apart of the first nucleotide of the RNA body than in a regular

dinucleotide and also the charge is different, it will interact with the exosome in a manner different to uncapped RNA, which will have an influence on the product. Furthermore, a fluorescent label, even if it does not influence the hydrolysis reaction, does influence the behaviour of RNA in gel electrophoresis and will make the RNA appear longer than it actually is. Thus it is inevitable to use capped, unlabelled RNA for degradation assays to gain reliable information about the product length. From our preliminary results on the exosome degradation products we can propose that the products are indeed mostly capped mono- and dinucleotides. However, further investigations on the product length are necessary to strengthen these initial insights. Our data so far suggests an agreement in product and substrate length of the exosome and DcpS and we propose that a direct handover of RNA is possible.

A strict threshold in substrate usage of DcpS between two and three nucleotides in the RNA body conserved among species is indicative of an essential cellular function. In general, on the one hand, obsolete capped RNA has to be degraded not only to terminate their translation and recycle nucleotides, but also to remove them from the pool of possible targets of cap binding proteins. On the other hand, RNA that is still transcribed must not be randomly targeted for degradation. The restriction of decapping by DcpS to products of exosomal degradation can be achieved by this threshold in substrate usage.

4.4 The Possibility of New Drugs Based on Interference with Protein Dynamics

DcpS has been connected with two diseases, the Al-Raqad syndrome and spinal muscular atrophy (SMA). In the Al-Raqad syndrome, the malfunctioning or non-functioning protein is DcpS itself and would need to be supplemented for treatment, whereas in SMA not DcpS is affected, but the protein SMN1. For SMA, therapeutic inhibitors have been developed on the basis of quinazoline derivatives that bind to DcpS and thereby the promoter of SMN2 is stimulated. However, it still remains unclear how DcpS is involved in that process. [95,110]. As DcpS was also shown to perform regulatory functions in RNA processing independent of its decapping activity (see Chapter 1.5), the modulation of promoter activity might be another example for such a regulatory function. This could either be achieved by direct or mediated interaction with the SMN promoter, or by a stabilising effect on the mRNA of SMN2.

Structural data on DcpS in complex with pharmacological inhibitors are available [95,97–99]. In all structures, DcpS is in the asymmetric conformation and Singh *et al.* proposed that the inhibitory effect is due to the preservation of this conformation. This would be an extreme case of dynamics modulation in which domain motions were decreased to a complete rest. However, it is also conceivable that a modulation of the dynamics in the other direction would be possible by interaction with pharmacological inhibitors. If the domain motions of DcpS would be increased to a level at which the active site is not closed long enough for catalysis, this would have a comparable effect, resembling the state of motions in substrate inhibition.

In general, many drugs target the occupation of the active site of their target protein by mimicking a reaction intermediate, by binding to an allosteric regulatory site or by disturbance of allosteric pathways of enzymes. As protein dynamics are a common theme in enzymes, influence on the dynamics of a protein during catalysis could be a further level for the design of novel drugs targeting malfunctioning proteins.

4.5 Vaccinia Capping Enzyme Enables Studies of Cap Binding Proteins with Native Ligands

The 5’ cap structure is a key feature of mRNA and an interaction platform for many proteins at all stages of mRNA lifetime. However, in-depth studies of cap binding proteins and their interaction patterns with capped RNA have to date mostly been restricted to cap analogues. Most studies were performed with m^7GDP or m^7GpppG, as no method for the large scale production of capped RNA of defined sequence and length was available. We presented a robust method for capping of *in vitro* transcribed RNA that enables experiments that were previously prohibited due to the lack of sufficient amounts of substrate or ligand. Now a vast range of structural and biochemical experiments is possible. However, RNA produced with our capping protocol is not limited to the study of RNA:protein interactions but also useful for the study of RNA itself. Experiments with capped 5’ leader RNA of HIV-1 gave insights into the distinction between RNA molecules that are used for translation and those packaged into virions [136]. Until now, we could provide more than 30 laboratories worldwide with our plasmid coding for the vaccinia capping enzyme.

To study molecular interactions with binding partners, that are as close to the native ones as possible, is of great importance to prevent artefacts, especially when studying interactions of biological relevance. One example for this is the length determination of RNA fragments produced by the hydrolytic reaction of the exosome. The cap structure does not have the characteristics of a nucleotide bond and thus will be recognised by the enzyme in another manner. The other way around, the interaction of cap binding proteins with RNA may in some cases be influenced by the RNA body. Even if an interaction between RNA and protein is solely mediated by the cap structure itself, the affinity can change dependent on the long and flexible RNA body that is attached to the cap.

The studies summarised in this thesis are an example for substrate processing that is dependent on protein dynamics. These dynamics can be influenced by the features of different substrates. They stress the importance of the substrate nature on enzymatic properties in a biological context and offer a method to solve the problematic lack of access to native mRNA 5' ends to study the interaction with binding partners on a functional as well as on a structural level. We anticipate that this work will therefore support studies that are related to capped RNA.

References

1. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15,** 163–75 (2014).

2. Carmody, S. R. & Wente, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122,** 1933–1937 (2009).

3. Stewart, M. Polyadenylation and nuclear export of mRNAs. *J. Biol. Chem.* jbc.REV118.005594 (2019).

4. Tudek, A., Schmid, M. & Jensen, T. H. Escaping nuclear decay: the significance of mRNA export for gene expression. *Curr. Genet.* (2018).

5. Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci.* **25,** 1390–1406 (2016).

6. Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* **13,** 448–448 (2012).

7. Papasaikas, P. & Valcárcel, J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.* **41,** 33–45 (2016).

8. Mouilleron, H., Delcourt, V. & Roucou, X. Death of a dogma: Eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Res.* **44,** 14–23 (2016).

9. Moore, M. J. & Proudfoot, N. J. Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. *Cell* **136,** 688–700 (2009).

10. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19,** 158–174 (2018).

11. Mayr, C. Regulation by 3 `-Untranslated Regions. *Annu. Rev. Genet. Vol 51* **51,** 171–194 (2017).

12. Covelo-Molares, H., Bartosovic, M. & Vanacova, S. RNA methylation in nuclear pre-mRNA processing. *Wiley Interdiscip. Rev. RNA* e1489 (2018).

13. Banerjee, a K. 5′-Terminal Cap Structure in Eucaryotic Messenger Ribonucleic Acids. *Microbiol. Rev.* **44,** 175–205 (1980).

14. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **49,** 18–29 (2013).

15. Wang, X. *et al.* N 6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505,** 117–120 (2014).

16. Walkley, C. R. & Li, J. B. Rewriting the transcriptome: Adenosine-to-inosine RNA editing by ADARs. *Genome Biol.* **18,** 1–13 (2017).

17. Gilbert, W. V, Bell, T. A. & Schaening, C. Messenger RNA modifications: Form,

distribution, and function. *Science* **352,** 1408–12 (2016).

18. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. TL - 63. *Microbiol. Mol. Biol. Rev.* **63 VN-r,** 405–445 (1999).

19. Casañal, A. *et al.* Architecture of eukaryotic mRNA 3'-end processing machinery. *Science (80-.).* 1056–1059 (2017).

20. Schreieck, A. *et al.* RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat. Struct. Mol. Biol.* **21,** 175–179 (2014).

21. Lewis, J. D. & Izaurralde, E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247,** 461–469 (1997).

22. Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. & Darnell, J. E. The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19,** 69–78 (1980).

23. Reddy, R., Ro-Choi, T. S., Henning, D. & Busch, H. Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J. Biol. Chem.* **249,** 6486–6494 (1974).

24. Furuichi, Y. & Miura, K. I. A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253,** 374–375 (1975).

25. Shatkin, A. J. Capping of eucaryotic mRNAs. *Cell* **9,** 645–653 (1976).

26. HsuChen, C.-C. & Dubin, D. T. Di- and trimethylated congeners of 7-methylguanine in Sindbis virus mRNA. *Nature* **264,** 190–191 (1976).

27. Liou, R. F. & Blumenthal, T. trans-spliced Caenorhabditis elegans mRNAs retain trimethylguanosine caps. *Mol. Cell. Biol.* **10,** 1764–1768 (1990).

28. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonné, R. 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 (2002).

29. Hausmann, S. & Shuman, S. Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J. Biol. Chem.* **280,** 4021–4024 (2005).

30. Singh, R. & Reddy, R. Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *PNAS* **86,** 8280–8283 (1989).

31. Izaurralde, E., Lewis, J., Jankowska, M., Darzynkiewicz, E. & Molecular, E. A Nuclear Cap Binding Protein Complex Involved in Pre-mRNA Splicing. *Cell* **78,** 657–668 (1994).

32. Schwer, B. & Shuman, S. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA* 574–583 (1996).

33. Konarska, M. M., Padgett, R. a & Sharp, P. a. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38,** 731–736 (1984).

34. Hamm, J. & Mattaj, I. W. Monomethylated cap structures facilitate RNA export from the

nucleus. *Cell* **63,** 109–118 (1990).

35. Both, G. W., Banerjee, A. K. & Shatkin, A. J. Methylation-dependent translation of viral messenger RNAs in vitro. *PNAS* **72,** 1189–93 (1975).

36. McKendrick, L., Pain, V. M. & Morley, S. J. Translation initiation factor 4E. *Int. J. Biochem. Cell Biol.* **31,** 31–35 (1999).

37. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5'-Terminal structure and mRNA stability. *Nature* **266,** 235–239 (1977).

38. Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.* **10,** 51–65 (2012).

39. Gu, M., Rajashankar, K. R. & Lima, C. D. Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA Capping Apparatus. *Structure* **18,** 216–227 (2010).

40. Ho, C. K. *et al.* The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **273,** 9577–9585 (1998).

41. Benarroch, D., Smith, P. & Shuman, S. Characterization of a Trifunctional Mimivirus mRNA Capping Enzyme and Crystal Structure of the RNA Triphosphatase Domain. *Structure* **16,** 501–512 (2008).

42. Shuman, S. What messenger RNA capping tells us about eukaryotic evolution. *Nat. Rev. Mol. Cell Biol.* **3,** 619–625 (2002).

43. Mao, X. & Shuman, S. Intrinsic RNA (Guanine-7) Methyltransferase Activity of the Vaccinia Virus Capping Enzyme D1 Subunit Is Stimulated by the D12 Subunit. *J. Biol. Chem.* **269,** 24472–24479 (1994).

44. De la Peña, M., Kyrieleis, O. J. P. & Cusack, S. Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J.* **26,** 4913–4925 (2007).

45. De Vlugt, C., Sikora, D. & Pelchat, M. Insight into Influenza: A Virus Cap-Snatching. *Viruses* **10,** (2018).

46. Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc. Natl. Acad. Sci.* **106,** 7768–7773 (2009).

47. Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y. & Liu, D. R. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.* **5,** 879–881 (2009).

48. Cahová, H., Winz, M. L., Höfer, K., Nübel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519,** 374–377 (2015).

49. Bird, J. G. *et al.* The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. *Nature* **535,** 444–447 (2016).

50. Contreras, R., Cheroutre, H., Degrave, W. & Fiers, W. Simple, efficient in vitro synthesis of

capped RNA useful for direct expression of cloned eukaryotic genes. *Nucleic Acids Res.* **10,** 6353–6362 (1982).

51. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* **40,** 515–526 (1985).

52. Nielsen, D. A. & Shapiro, D. J. Preparation of capped RNA transcripts using T7 RNA potymerase. *Nucleic Acids Res.* **14,** 5936 (1986).

53. Pasquinelli, A. E., Dahlberg, J. E. & Lund, E. Reverse 5′ caps in RNAs made in vitro by phage RNA polymerase. *RNA* 957–967 (1995).

54. Grudzien-Nogalska, E. *et al.* Synthesis of Anti-Reverse Cap Analogs (ARCAs) and their Applications in mRNA Translation and Stability. *Methods Enzymol.* **431,** 203–227 (2007).

55. Thillier, Y. *et al.* Synthesis of 5′ cap-0 and cap-1 RNAs using solid-phase chemistry coupled with enzymatic methylation by human (guanine-N7)-methyl transferase. *RNA* **18,** 856–868 (2012).

56. Lewdorowicz, M. *et al.* Solid-Supported Synthesis of 5′-mRNA CAP-4 from Trypanosomatids. *Nucleosides, Nucleotides and Nucleic Acids* **26,** 1329–1333 (2007).

57. Wang, Z. & Kiledjian, M. Functional link between the mammalian exosome and mRNA decapping. *Cell* **107,** 751–762 (2001).

58. Parker, R. RNA degradation in Saccharomyces cerevisae. *Genetics* **191,** 671–702 (2012).

59. Parker, R. & Song, H. The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* **11,** 121–127 (2004).

60. Wilusz, C. J., Wormington, M. & Peltz, S. W. The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* **2,** 237–246 (2001).

61. Mugridge, J. S., Coller, J. & Gross, J. D. Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. *Nat. Struct. Mol. Biol.* **25,** (2018).

62. Deshmukh, M. V. *et al.* mRNA Decapping Is Promoted by an RNA-Binding Channel in Dcp2. *Mol. Cell* **29,** 324–336 (2008).

63. Van Dijk, E. *et al.* Human Dcp2: A catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* **21,** 6915–6924 (2002).

64. Lykke-Andersen, J. Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay. *Mol. Cell. Biol.* **22,** 8114–8121 (2002).

65. Fromm, S. a *et al.* The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J.* **31,** 279–90 (2012).

66. Sharif, H. & Conti, E. Architecture of the Lsm1-7-Pat1 Complex: A Conserved Assembly in Eukaryotic mRNA Turnover. *Cell Rep.* **5,** 283–291 (2013).

67. Tharun, S. & Parker, R. Targeting an mRNA for decapping: Displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol. Cell* **8,** 1075–1083 (2001).

68. Wang, Z., Jiao, X., Carr-Schmid, A. & Kiledjian, M. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12663–12668 (2002).

69. Wurm, J. P., Overbeck, J. & Sprangers, R. The S. Pombe mRNA decapping complex recruits cofactors and an Edc1-like activator through a single dynamic surface. *RNA* **22,** 1360–1372 (2016).

70. Wurm, J. P., Holdermann, I., Overbeck, J. H., Mayer, P. H. O. & Sprangers, R. Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc. Natl. Acad. Sci.* **114,** 6034–6039 (2017).

71. Steiger, M., Carr-schmid, A., Schwartz, D. C., Kiledjian, M. & Parker, R. O. Y. Analysis of recombinant yeast decapping enzyme. 231–238 (2003).

72. Hsu, C. L. & Stevens, A. 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 (1993).

73. Ingelfinger, D., Arndt-Jovin, D. J., Lührmann, R. & Achsel, T. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *RNA (New York, NY)* **8,** 1489–1501 (2002).

74. Januszyk, K. & Lima, C. D. The eukaryotic RNA exosome. *Curr. Opin. Struct. Biol.* **24,** 132–140 (2014).

75. Makino, D. L., Baumgärtner, M. & Conti, E. Crystal structure of an rna-bound 11-subunit eukaryotic exosome complex. *Nature* **495,** 70–75 (2013).

76. Makino, D. L. & Conti, E. Structure determination of an 11-subunit exosome in complex with RNA by molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69,** 2226–2235 (2013).

77. Han, J. & van Hoof, A. The RNA Exosome Channeling and Direct Access Conformations Have Distinct In Vivo Functions. *Cell Rep.* **16,** 3348–3358 (2016).

78. Schneider, C. & Tollervey, D. Threading the barrel of the RNA exosome. *Trends Biochem. Sci.* **38,** 485–493 (2013).

79. Mitchell, P., Petfalski, E., Shevchenko, a, Mann, M. & Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* **91,** 457–466 (1997).

80. Nuss, D. & Furuichi, Y. Characterization of the m7G(5')pppN-pyrophosphatase Activity from HeLa Cells. *J. Biol. Chem.* **252,** 2815–2821 (1977).

81. Karousis, E. D., Nasif, S. & Mühlemann, O. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA* **7,** 661–682 (2016).

82. Ghosh, S. & Jacobson, A. RNA decay modulates gene expression and controls its fidelity.

Wiley Interdiscip. Rev. RNA **1,** 351–361 (2010).

83. Bicknell, A. A. & Ricci, E. P. When mRNA translation meets decay. *Biochem. Soc. Trans.* **45,** 339–351 (2017).

84. Meister, G. Argonaute proteins: Functional insights and emerging roles. *Nat. Rev. Genet.* **14,** 447–459 (2013).

85. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12,** 99–110 (2011).

86. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A. & Conti, E. Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the RNase II Nuclease Family. *Mol. Cell* **29,** 717–728 (2008).

87. Wasmuth, E. V., Zinder, J. C., Zattas, D., Das, M. & Lima, C. D. Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *Elife* **6,** 1–24 (2017).

88. Zinder, J. C., Wasmuth, E. V. & Lima, C. D. Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. *Mol. Cell* **64,** 734–745 (2016).

89. Liu, H., Rodgers, N. D., Jiao, X. & Kiledjian, M. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J.* **21,** 4699–4708 (2002).

90. Chen, N., Walsh, M. a., Liu, Y., Parker, R. & Song, H. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J. Mol. Biol.* **347,** 707–718 (2005).

91. Gu, M. *et al.* Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol. Cell* **14,** 67–80 (2004).

92. Séraphin, B. The HIT protein family: a new family of proteins present in prokaryotes, yeast and mammals. *DNA Seq.* **3,** 177–179 (1992).

93. Lima, C. D., Klein, M. G. & Hendrickson, W. A. Structure-based analysis of catalysis and substrate definition in the HIT protein family. *Science* **278,** 286–290 (1997).

94. Liu, S.-W., Jiao, X. & Liu, H. Functional analysis of mRNA scavenger decapping enzymes. *RNA* **10,** 1412–1422 (2004).

95. Singh, J. *et al.* DcpS as a Therapeutic Target for Spinal Muscular Atrophy. *ACS Chem. Biol.* **3,** 711–722 (2008).

96. Wojtczak, B. A. *et al.* 5′-Phosphorothiolate Dinucleotide Cap Analogues: Reagents for Messenger RNA Modification and Potent Small-Molecular Inhibitors of Decapping Enzymes. *J. Am. Chem. Soc.* **140,** 5987–5999 (2018).

97. Liu, S. Dcps in complex with covalent inhibitors. *TO BE Publ.* doi:10.2210/PDB4QDE/PDB

98. Liu, S. Dcps in complex with covalent ligands targeting Tyrosines. *TO BE Publ.* doi:10.2210/PDB4QDV/PDB

99. Liu, S. Human Dcps in complex with covalent inhibitor. *TO BE Publ.* doi:10.2210/PDB4QEB/PDB

100. Liu, S. W., Rajagopal, V., Patel, S. S. & Kiledjian, M. Mechanistic and kinetic analysis of the dcps scavenger decapping enzyme. *J. Biol. Chem.* **283,** 16427–16436 (2008).

101. Sinturel, F., Bréchemier-baey, D., Kiledjian, M., Condon, C. & Bénard, L. Activation of 5'-3' exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. *PNAS* **109,** 8264–8269 (2012).

102. Liu, H. & Kiledjian, M. Scavenger decapping activity facilitates 5' to 3' mRNA decay. *Mol. Cell. Biol.* **25,** 9764–9772 (2005).

103. Zhou, M. I., Bail, S., Plasterer, H. L., Rusche, J. & Kiledjian, M. DcpS is a transcript-specific modulator of RNA in mammalian cells. *RNA* **21,** 1306–1312 (2015).

104. Meziane, O. *et al.* The human decapping scavenger enzyme DcpS modulates microRNA turnover. *Sci. Rep.* **5,** 16688 (2015).

105. Bossé, G. D. *et al.* The decapping scavenger enzyme DCS-1 controls microRNA levels in Caenorhabditis elegans. *Mol. Cell* **50,** 281–7 (2013).

106. Shen, V., Liu, H., Liu, S., Jiao, X. & Kiledjian, M. DcpS scavenger decapping enzyme can modulate pre-mRNA splicing. *RNA* 1132–1142 (2008).

107. Malys, N., Carroll, K., Miyan, J., Tollervey, D. & McCarthy, J. E. G. The 'scavenger' m7GpppX pyrophosphatase activity of Dcs1 modulates nutrient-induced responses in yeast. *Nucleic Acids Res.* **32,** 3590–3600 (2004).

108. Malys, N. & McCarthy, J. E. G. Dcs2, a Novel Stress-induced Modulator of m7GpppX Pyrophosphatase Activity that Locates to P Bodies. *J. Mol. Biol.* **363,** 370–382 (2006).

109. Ng, C. K. L. *et al.* Loss of the scavenger mRNA decapping enzyme DCPS causes syndromic intellectual disability with neuromuscular defects. *Hum. Mol. Genet.* **24,** 3163–71 (2015).

110. Gogliotti, R. G. *et al.* The dcpS inhibitor RG3039 improves survival, function and motor unit pathologies in two SMA mouse models. *Hum. Mol. Genet.* **22,** 4084–4101 (2013).

111. Ziemniak, M., Strenkowska, M., Kowalska, J. & Jemielity, J. Potential therapeutic applications of RNA cap analogs. *Futur. Med. Chem.* **5,** 1141–1172 (2013).

112. Gopalsamy, A. *et al.* Design of Potent mRNA Decapping Scavenger Enzyme (DcpS) Inhibitors with Improved Physicochemical Properties To Investigate the Mechanism of Therapeutic Benefit in Spinal Muscular Atrophy (SMA). *J. Med. Chem.* **60,** 3094–3108 (2017).

113. Pentikäinen, U., Pentikäinen, O. T. & Mulholland, A. J. Cooperative symmetric to asymmetric conformational transition of the apo-form of scavenger decapping enzyme revealed by simulations. *Proteins* **70,** 498–508 (2007).

114. Pervushin, K. V., Wider, G., Riek, R. & Wüthrich, K. The 3D NOESY- [1 H , 15 N , 1 H] -ZQ-TROSY NMR experiment with diagonal peak suppression. *PNAS* **96,** 9607–9612 (1999).

115. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H-13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125,** 10420–8 (2003).

116. Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* **34,** 601–11 (2009).

117. Coleman, T. M., Wang, G. & Huang, F. Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic Acids Res.* **32,** (2004).

118. Moss, B., Gershowitz, A., Wei, C. M. & Boone, R. Formation of the guanylylated and methylated 5'-terminus of vaccinia virus mRNA. *Virology* **72,** 341–351 (1976).

119. Gong, C. & Shuman, S. Mapping the active site of vaccinia virus RNA triphosphatase. *Virology* **309,** 125–134 (2003).

120. Higman, M. a., Bourgeois, N. & Niles, E. G. The vaccinia virus mRNA (guanine-N7-)-methyltransferase requires both subunits of the mRNA capping enzyme for activity. *J. Biol. Chem.* **267,** 16430–16437 (1992).

121. Cong, P. & Shuman, S. Mutational Analysis of mRNA Capping Enzyme Identifies Amino Acids Involved in GTP Binding, Enzyme-Guanylate Formation, and GMP Transfer to RNA. **15,** 6222–6231 (1995).

122. Yu, L. & Shuman, S. Mutational analysis of the triphosphatase component of vaccinia virus mRNA capping enzyme. *J. Virol.* **70,** 6162–6168 (1996).

123. Zheng, S. & Shuman, S. Structure – function analysis of vaccinia virus mRNA cap methyltransferase. *RNA* **1,** 696–705 (2008).

124. Roth, M. J. & Hurwitz, J. RNA capping by the vaccinia virus guanylyltransferase. Structure of enzyme-guanylate intermediate. *J. Biol. Chem.* **259,** 13488–13494 (1984).

125. Kinkelin, K., Veith, K., Grünwald, M. & Bono, F. Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression. *RNA* **18,** 1624–1634 (2012).

126. Neu, A., Neu, U., Fuchs, A.-L., Schlager, B. & Sprangers, R. An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nat. Chem. Biol.* **11,** (2015).

127. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **62,** 453–471 (2015).

128. Reed, M. C., Lieb, A. & Nijhout, H. F. The biological significance of substrate inhibition: A mechanism with diverse functions. *BioEssays* **32,** 422–429 (2010).

129. Rosenberry, T. L., Mallender, W. D., Thomas, P. J. & Szegletes, T. A steric blockade model for inhibition of acetylcholinesterase by peripheral site ligands and substrate. *Chem. Biol. Interact.* **119–120,** 85–97 (1999).

130. Cohen, S. G. *et al.* General occurrence of binding to acetylcholinesterase-substrate complex in noncompetitive inhibition and in inhibition by substrate. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1076,** 112–122 (1991).

131. Quinsey, N. S., Luong, A. Q. & Dickson, P. W. Mutational analysis of substrate inhibition in tyrosine hydroxylase. *J. Neurochem.* **71,** 2132–8 (1998).

132. Colletier, J. P. *et al.* Structural insights into substrate traffic and inhibition in acetylcholinesterase. *EMBO J.* **25,** 2746–2756 (2006).

133. Cabrera, R., Ambrosio, A. L. B., Garratt, R. C., Guixé, V. & Babul, J. Crystallographic Structure of Phosphofructokinase-2 from Escherichia coli in Complex with Two ATP Molecules. Implications for Substrate Inhibition. *J. Mol. Biol.* **383,** 588–602 (2008).

134. Buschmann, J. *et al.* Identification of drosophila and human 7-methyl GMP-specific nucleotidases. *J. Biol. Chem.* **288,** 2441–2451 (2013).

135. Kim, C. H. & Sarma, R. H. Spatial configuration of mRNA 5'-terminus. *Nature* **270,** (1977).

136. Kharytonchyk, S. *et al.* Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci.* **113,** 13378–13383 (2016).

Appendix

Manuscripts discussed in this thesis are reprinted in the following order hereafter.

1. Neu, A., Neu, U., **Fuchs, A.-L.**, Schlager, B., & Sprangers, R. (2015). An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nature Chemical Biology*

2. **Fuchs, A.-L.***, Neu, A.*, & Sprangers, R. (2016). A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA*

3. **Fuchs, A.-L.**, Neu, A., & Sprangers, R.*;* Molecular basis for mRNA length sensing by the DcpS decapping enzyme

ARTICLE
PUBLISHED ONLINE: 10 AUGUST 2015 | DOI: 10.1038/NCHEMBIO.1866

An excess of catalytically required motions inhibits the scavenger decapping enzyme

Ancilla Neu[1], Ursula Neu[2,3], Anna-Lisa Fuchs[1], Benjamin Schlager[1,3] & Remco Sprangers[1]*

The scavenger decapping enzyme hydrolyzes the protective 5′ cap structure on short mRNA fragments that are generated from the exosomal degradation of mRNAs. From static crystal structures and NMR data, it is apparent that the dimeric enzyme has to undergo large structural changes to bind its substrate in a catalytically competent conformation. Here we studied the yeast enzyme and showed that the associated opening and closing motions can be orders of magnitude faster than the catalytic turnover rate. This excess of motion is induced by the binding of a second ligand to the enzyme, which occurs at high substrate concentrations. We designed a mutant that disrupted the allosteric pathway that links the second binding event to the dynamics and showed that this mutant enzyme is hyperactive. Our data reveal a unique mechanism of substrate inhibition in which motions that are required for catalytic activity also inhibit efficient turnover when they are present in excess.

Enzyme catalysis involves a finely tuned balance between protein structure and dynamics[1–3]. Over the past several decades our knowledge regarding protein structure has grown significantly[4]. At the same time, our understanding of how fluctuations in protein structure modulate function has lagged far behind[5–8]. Although indirect information about structural changes during catalysis can be gained from crystal structures of enzymes at different states along the reaction pathway, these static structures do not provide information about the rates with which an enzyme samples these different states. These rates have important implications for catalytic activity; however, only in very few cases has it been possible to directly link protein dynamics with function[9–14]. Addressing this relationship experimentally is pivotal to advancing our understanding of enzyme function. Notably, conformational fluctuations also have implications for enzyme design[15] and for the development of pharmaceutical compounds that modulate catalytic activity.

In eukaryotes, mRNA is protected against degradation by the presence of a 3′ poly(A) tail and a 5′ cap structure (**Supplementary Results**, **Supplementary Fig. 1a**). The scavenger decapping enzyme (DcpS in humans; Dcs1p in yeast) catalyzes cap hydrolysis of short (<10 nucleotides) mRNA fragments that are produced by 3′-to-5′ exosome-mediated mRNA degradation[16,17]. The products of the reaction are m^7G monophosphate (m^7GMP) and an RNA body that carries a 5′ diphosphate (**Supplementary Fig. 1a**). The activity of the DcpS enzyme must be regulated to prevent premature decapping and unintentional degradation of an mRNA transcript. In yeast the Dcs2p protein can directly interact with and modulate the catalytic activity of the Dcs1p enzyme[18]. Recently there has been growing interest in the scavenger decapping enzyme as a target for the treatment of spinal muscular atrophy (SMA)[19,20].

DcpS is an 80-kDa homodimeric enzyme for which the human complex has been crystallized in its ligand-free form, in the presence of minimal cap substrates (such as m^7GpppG) (**Supplementary Fig. 1**) and in complex with pharmaceutically relevant inhibitors[19,21–23]. DcpS is composed of a large 50-kDa dimeric C-terminal domain that is flexibly connected by a hinge region to a smaller domain-swapped 30-kDa dimeric N-terminal lid-like domain (**Supplementary Fig. 1b**). In the substrate-free form, the enzyme has a symmetric conformation with few contacts between the two domains[21]. Upon interaction with substrate, however, the complex adopts an asymmetric conformation in which one substrate molecule is sandwiched between the N- and C-terminal domains on one side of the dimer (referred to as the closed site)[21,22]. In this binding site, the catalytic triad is assembled to promote hydrolysis of the substrate[17,24]. The other side of the dimeric enzyme adopts an open and catalytically incompetent conformation (referred to as the open site) that can interact with a second substrate molecule.

These static structures[21,22] together with results from biochemical experiments[25] and molecular dynamics simulations[26] have led to the view that DcpS undergoes large domain rearrangements during the catalytic cycle to both interact with substrate and release products. Presumably the N-terminal lid domain flips over during catalysis, which then results in an opening of the closed site and a simultaneous closing of the open site[22]. This mechanism suggests that the dimeric nature of the enzyme and its domain motions enhance catalytic efficiency because a substrate molecule can be positioned in the open binding site as catalysis is taking place at the closed binding site. However, several lines of evidence are in conflict with this model. First, single point mutations that enhance DcpS activity have been identified, showing that the wild-type enzyme does not function at the maximally possible catalytic rate[21,22]. Second, the DcpS enzyme is substrate inhibited, as it is more active under low substrate concentrations than under high substrate concentrations[25]. This phenomenon is suggested to be due to a change in the rate-limiting step of the reaction from substrate binding (under single-turnover conditions) to a conformational change in the enzyme (under multiple-turnover conditions)[25]. Quantitative data to test these hypotheses are not available, and a more elaborate model that accurately explains the behavior of the enzyme is required.

To gain insights into the catalytic cycle of the scavenger decapping enzyme and its regulation, we combined methyl-group NMR spectroscopy, X-ray crystallography and isothermal titration calorimetry (ITC) studies. This enabled us to obtain an unexpected view of how interdomain motions directly regulate catalytic turnover and to provide a unique example for the fact that regulatory mechanisms can be hidden in static structures. We show that the domain-flipping motions in the scavenger decapping enzyme are important for substrate binding and product release. We also show that these motions increase with increasing substrate concentration and can reach frequencies that are two orders of magnitude faster than those of

[1]Max Planck Institute for Developmental Biology, Tübingen, Germany. [2]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany. [3]Present addresses: Max Planck Institute for Colloids and Interfaces, Potsdam, Germany (U.N.); FluidSolids AG, Zürich, Switzerland (B.S.).
*e-mail: remco.sprangers@tuebingen.mpg.de

substrate turnover. Unexpectedly, these excessive flipping motions ultimately inhibited substrate hydrolysis. In agreement with that, we showed that the hyperactivity of the K126A mutant enzyme is because of the reduced motions of the lid domain. In addition, our data establish a unique mode of substrate inhibition for DcpS in which high substrate concentrations induce excessive conformational changes in the enzyme that are too fast to permit catalysis.

RESULTS

The ligand-free scavenger decapping enzyme is symmetric

To gain detailed information on the structure and dynamics of the scavenger decapping enzyme along the catalytic pathway, we first studied the ligand-free complex in solution by NMR spectroscopy. Methyl-TROSY NMR techniques[27] are ideally suited to probe motions in protein complexes and are applicable to large and complex systems[28–31]. In this study, we used the scavenger decapping enzyme Dcs1p from *Saccharomyces cerevisiae*, as the human enzyme for which X-ray structures are available did not provide the long-term stability required for detailed NMR studies. In our NMR experiments, we observed that the ligand-free form of the yeast DcpS enzyme displayed only a single set of resonances (**Supplementary Fig. 1d**). This was consistent with both protein chains of the dimer being in an identical chemical environment, which occurs only in the case of a symmetric homodimeric complex. Furthermore, we observed that the methyl-TROSY spectra of the isolated dimeric N- and C-terminal domains superimposed very well with the full-length protein (**Supplementary Fig. 1d**). This showed that there are no close contacts between the dimeric N-terminal domain and the dimeric C-terminal domain. In summary, we established that the substrate-free enzyme from yeast is symmetric in solution and that the N- and C-terminal domains behave as independent units. These results are in full agreement with X-ray[21] and computational studies[26] of the ligand-free human enzyme that show that although the angle between the N- and C-terminal domains can change, the domains themselves move as rigid bodies.

Dcs1p interacts with substrates in a sequential manner

High-resolution structures are a prerequisite for the analysis of NMR studies that probe dynamics and interactions. The partially low sequence identity between the yeast and human DcpS enzymes precluded the generation of a reliable homology model for the yeast protein. We therefore used X-ray crystallography to determine the structure of the yeast Dcs1p enzyme in complex with m^{7}GDP. To prevent hydrolysis of the substrate during measurements, we used a catalytically inactive variant of the enzyme (Dcs1p^{H268N})[22]. We chose m^{7}GDP as a ligand because it mimics the substrate and is not hydrolyzed efficiently[32]. We solved the structure of the yeast enzyme at a resolution of 2.3 Å (**Supplementary Table 1**) and found that it featured the same architecture as the structure of the human DcpS enzyme bound to an m^{7}GpppG substrate (1.8 Å r.m.s. deviation over all Cα atoms in common secondary structure elements) (**Fig. 1a** and **Supplementary Fig. 1c**). As in the human protein, Dcs1p adopted an asymmetric conformation such that the ligand was placed between the N- and C-terminal domains in the closed binding site, whereas the other side of the dimer adopted an open conformation (**Fig. 1a**,**b**). It is worth noting that the second site is occupied with substrate in the structure of the human enzyme[22] but was empty in our yeast enzyme structure (our enzyme was crystallized at a lower substrate concentration). This finding indicates that both binding sites in the asymmetric form of the enzyme differ substantially in substrate affinity.

To probe the binding mechanism between Dcs1p^{H268N} and its substrate m^{7}GpppG in solution, we used methyl-TROSY NMR spectroscopy to directly follow structural changes in the enzyme upon stepwise addition of the substrate (**Fig. 2**). To map the residues involved in this interaction, we assigned numerous isoleucine and methionine methyl resonances that were close to the active site or that could potentially probe conformational changes during the catalytic cycle[33,34] (see **Supplementary Fig. 2**). Notably, our NMR titration experiments displayed a two-step mechanism for substrate binding to Dcs1p. First, we observed that a large number of methyl resonances split into two signals upon addition of an equimolar amount of m^{7}GpppG (such as those for I12, I36, I150, M153 and I219). This demonstrated that the initially symmetric ligand-free enzyme was converted into an asymmetric dimer upon complex formation with the substrate (**Figs. 1** and **2b**). It is worth noting that the change in chemical environment between the symmetric and asymmetric conformation is much larger for the closed side than for the open side. We therefore attributed methyl resonances that displayed large chemical shift perturbations (CSPs) upon substrate addition to residues in the closed binding site,

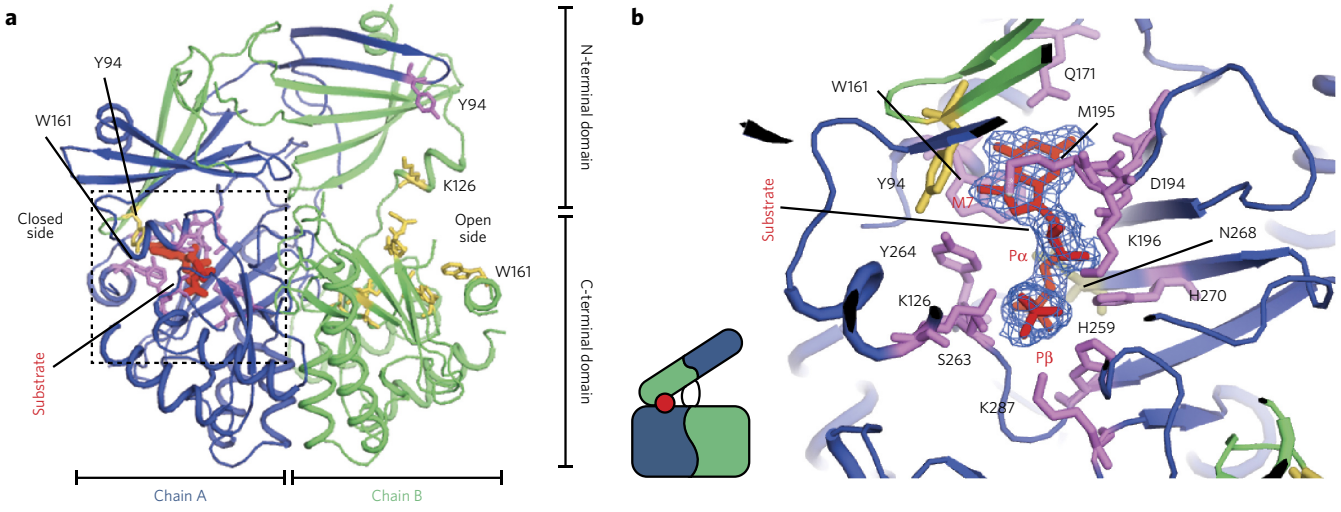


Figure 1 | Structure of the yeast DcpS–substrate complex. (**a**) 2.3 Å resolution crystal structure of the Dcs1p enzyme from *S. cerevisiae* in complex with m^{7}GDP. The dimeric enzyme adopts an asymmetric conformation with one closed and one open binding site. The protein chains are colored blue with magenta side chains and green with yellow side chains, respectively, and the substrate is colored red. The boxed region is shown in panel **b** in a similar orientation. A cartoon representing the conformation of the enzyme is indicated, and the red dot refers to the substrate. (**b**) The substrate is tightly bound in the closed binding site. Residues Y94 and W161 highlight the high asymmetry of the enzyme in the ligand-bound form; these aromatic rings come within 5 Å of each other in the closed binding site but are more than 22 Å apart in the open binding site. Without conformational changes, the substrate or product is not able to dissociate from the enzyme.

whereas residues in the open side remained, in many instances, unaffected (**Fig. 2b**). The fact that the observed CSPs are in slow exchange on the NMR time scale strongly suggested that this first binding event occurs with high affinity. Upon addition of hyperstoichiometric concentrations of m^{7}GpppG, we observed additional CSPs for residues that were exclusively located in the open binding site of the asymmetric enzyme (**Fig. 2c**). Hence we can conclude that this second binding event takes place in the open side of the asymmetric dimer. Taken together, our NMR data are consistent with a sequential binding mechanism in which the first substrate molecule that binds to the enzyme induces an asymmetric conformation of the enzyme and forms a closed binding site. At higher substrate concentrations, the second binding site on the open side of the asymmetric dimer can then be occupied, albeit with a lower affinity.

Dcs1p contains a high- and a low-affinity binding site

To quantify the affinities associated with these two binding events, we performed accurate NMR titration experiments (**Fig. 2d** and **Supplementary Fig. 3a**). We then used the intensities of the resonances that reported on the first binding event (those of I12 and I36; **Fig. 2b**) and the positions of the resonances that reported on the occupancy of the second binding site (those of I42 and M153; **Fig. 2c**) to extract dissociation constants for both binding events. To that end we fitted the spectral properties of these four methyl groups using a sequential binding model (**Supplementary Fig. 3a**). From this analysis we extracted a high-nanomolar-affinity value for the first binding event (68 ± 255 nM) and a three-orders-of-magnitude weaker (micromolar) value for the second binding event (105 ± 20 µM) (**Table 1**). A high-affinity binding event was previously reported for the human enzyme[25,35], whereas the low-affinity binding event has not been observed before. To independently validate the sequential binding of substrate molecules to the enzyme and the large difference in binding affinities for the two binding events, we turned to ITC measurements (**Supplementary Fig. 3a**). In the ITC thermograph, a first binding event was observed in the nanomolar range, whereas the affinity for the second binding site was determined reliably to be 324 ± 67 µM, in agreement with our NMR data. Taken together, our NMR and ITC data demonstrate that the scavenger decapping enzyme interacts with two substrates in a sequential manner, in which the second binding site emerges only after interaction of the enzyme with the first substrate molecule.

The C-terminal domain binds the second substrate molecule

To characterize the second binding site in more detail, we performed ITC and NMR experiments with the isolated C-terminal domain of the yeast Dcs1p enzyme and m^{7}GpppG. Using ITC we showed that the isolated C-terminal domain of the enzyme interacts

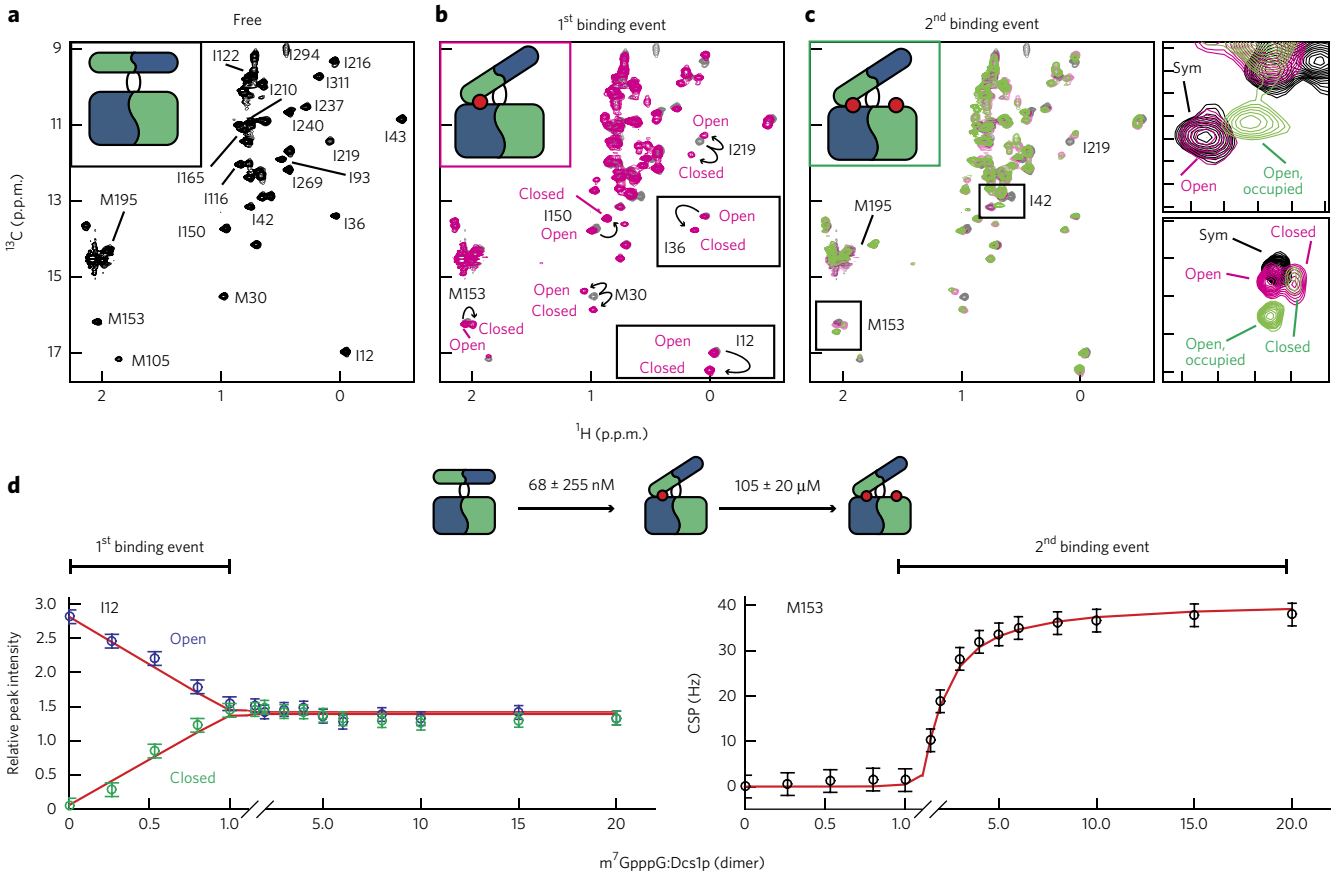


Figure 2 | The scavenger decapping enzyme binds to two substrate molecules in a sequential manner. (**a**–**c**) Methyl-TROSY NMR spectra of the yeast enzyme (0.25 mM dimer concentration) without (**a**) or with (**b**,**c**) substrate. Cartoons indicate the states of the complexes. (**a**) Ligand-free, symmetric Dcs1p. Assignments for a number of isoleucine and methionine residues that are close to the active site are indicated (see also **Supplementary Fig. 2**). (**b**) Dcs1p after addition of one molar equivalent of m^{7}GpppG (0.25 mM) to the dimeric enzyme. A number of resonances that report on the symmetric to asymmetric conversion are indicated with arrows. (**c**) As in (**b**) but after addition of 20 molar equivalents of substrate (5 mM). Resonances that result from residues in the open binding site shift, indicating that a second substrate molecule interacts in the open binding site of the enzyme. The insets on the right are expansions of the boxed regions in **c** and show a detailed view of the sequential binding process. Sym refers to the symmetric ligand-free protein. (**d**) Plot of the changes in the peak intensities for residue I12 (left, open and closed resonances) and the changes in the peak position for residue M153 (right, CSP in carbon) for 15 titrations points (see **a**–**c** and boxed regions in the spectra in **b**,**c**). The errors (s.d.) in the extracted binding constants were based on Monte Carlo simulations for a simultaneous fit of multiple residues (**Supplementary Fig. 3a**). For clarity, the scale of the *x* axis is shown differently below and above the 1:1 m^{7}GpppG:Dcs1p ratio as indicated.

Table 1 | Summary of the measured affinities, domain-flipping rates and turnover rates

	WT	Dcs1p^{K126A}
K_d substrate (m^{7}GpppG)[d]	Closed site (Dcs1p^{H268N}): 68 ± 255 nM[a,b]	Closed site (Dcs1p$^{H268N\ K126A}$): 112 ± 390 nM[a,b]; 240 ± 90 nM[c]
	Open site (Dcs1p^{H268N}): 105 ± 20 μM[a]; 324 ± 67 μM[c]	Open site (Dcs1p$^{H268N\ K126A}$): 196 ± 31 μM[a]; 280 ± 30 μM[c]
Flipping rate[d]	Dcs1p^{H268N}: from 8 Hz (low substrate excess) to 34 Hz (high substrate excess)	Dcs1p$^{H268N\ K126A}$: <1 Hz under high and low substrate excess
Turnover rate	WT: 0.27 s^{-1} under high substrate excess	Dcs1p^{K126A}: 0.48 s^{-1} under high substrate excess

[a]Determined using NMR. [b]The high uncertainty in the determined K_d of the closed binding site results from the limited number of data points that define the transition. [c]Determined using ITC. [d]Inactive mutant (H268N) background.

with m^{7}GpppG with an affinity that is comparable to that of the second binding event in the full-length complex (**Supplementary Fig. 3c**). Together with the structure of the human DcpS enzyme[22] (**Supplementary Figs. 1b** and **3c**), this shows that the second ligand interacts predominantly with the C-terminal domain in the open binding pocket. To structurally characterize substrate binding to the open site of the yeast enzyme, we performed NMR titration experiments using the isolated dimeric C-terminal domain and compared these results to those from the titration experiment performed with the full-length enzyme. For the isolated C-terminal domain, we observed that the same residues experienced CSPs upon addition of substrate as those in the full-length protein that was incubated with hyperstoichiometric amounts of substrate (**Supplementary Fig. 3c**). Notably, despite the fact that the second ligand mainly contacts the C-terminal domain, we observed that occupation of the second binding site in the full-length protein resulted in CSPs in the N-terminal domain (such as those for I42; **Fig. 2c**). Based on the structure of the human protein in complex with m^{7}GpppG substrate, the distance between the second substrate and I42 in the yeast enzyme is expected to be over 12 Å. This long-range effect points toward an allosteric mechanism that can relay the substrate-binding event in the open binding pocket to the N-terminal lid domain and potentially further toward the closed binding pocket.

Activity requires the asymmetric form of the complex

Both X-ray (**Fig. 1**) and NMR (**Fig. 2**) data strongly suggested that the scavenger decapping enzyme needs to adopt an asymmetric form to be catalytically competent. In yeast Dcs1p (**Fig. 1**), this closed conformation is mediated by a network of interactions that involves contacts between the 7-methyl group of the substrate and the aromatic ring of Tyr94. To assess the importance of these interactions for stable domain closure, we recorded methyl NMR spectra of the Dcs1p$^{H268N\ Y94A}$ enzyme in the presence of m^{7}GpppG and spectra of the Dcs1p^{H268N} enzyme in presence of the unmethylated ligand GpppG (**Supplementary Fig. 4a**). In both cases we observed no resonances that corresponded to the asymmetric conformation (**Supplementary Fig. 4a**) like the ones we observed for the Dcs1p^{H268N} enzyme with m^{7}GpppG (**Fig. 2b**). However, we did observe small CSPs in the same region where we had observed CSPs for substrate binding to Dcs1p^{H268N} (**Fig. 2** and **Supplementary Fig. 4a**). This showed that impairing the interaction between Dcs1p Y94 and the 7-methyl group results in a situation where the enzyme-substrate complex is no longer able to form a stable asymmetric conformation.

To directly test the importance of the asymmetric conformation for catalytic activity, we monitored the hydrolysis of m^{7}GpppG and GpppG by the WT yeast enzyme and the Y94A mutant, respectively, using a time series of one-dimensional ^{31}P NMR spectra. The WT enzyme hydrolyzed m^{7}GpppG into m^{7}GMP and GDP (**Supplementary Figs. 1a** and **4b**); however, we did not detect enzymatic turnover for the Y94A enzyme with m^{7}GpppG or for the WT enzyme with GpppG (**Supplementary Fig. 4b**). In summary, we show that the formation of a sufficiently long-lived closed active site is crucial for catalytic activity, which underscores the importance of conformational changes in the catalytic cycle of the scavenger decapping enzyme.

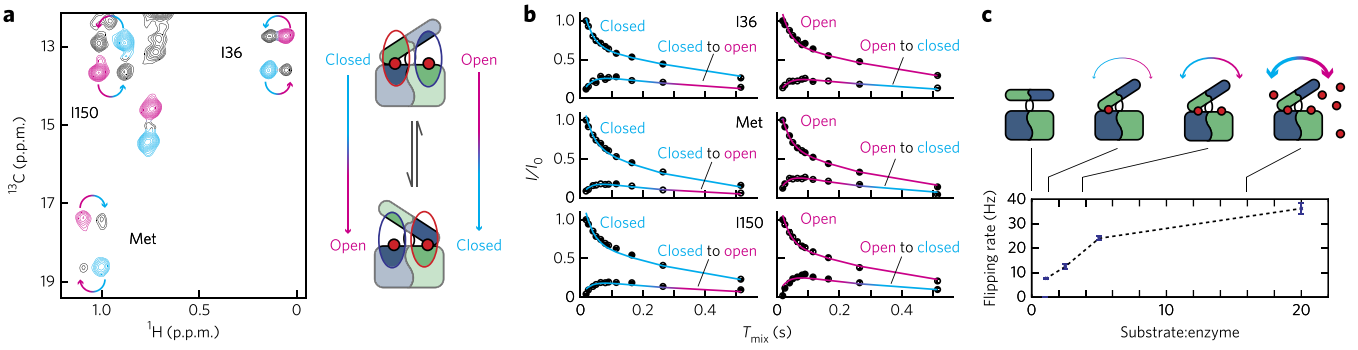


Figure 3 | Quantification of domain-flipping motions in the scavenger decapping enzyme. (**a**) Longitudinal exchange experiment that directly reports on motions (the open binding site closes and the closed binding site opens) in the enzyme. The cyan- and pink-colored resonances result from the closed and open sites, respectively. The arrows point to the resonances that appear as a result of the domain-flipping motions. The spectrum shown here is recorded with a mixing time of 75 ms in the presence of a 20-fold excess of substrate over the dimeric enzyme. The cartoons represent the flipping motion detected in the NMR spectra. (**b**) Quantification of the exchange process in the presence of a 20-fold excess of ligand. The circles indicate resonance intensities of the cross and auto peaks. The error bars (which are within the circles) represent uncertainties in the resonance intensities. The drawn lines are a best global fit to the data and yield an exchange rate of 35.4 (± 4.8) s^{-1}. The error (s.d.) in the extracted parameters is based on 100 Monte Carlo simulations. I and I_0 refer to the signal intensities of the exchange and reference experiment, respectively. T_{mix} indicates the length of the exchange delay in the NMR experiment. (**c**) The flipping motions (**a**,**b**) depend on the m^{7}GpppG ligand excess. Higher excess of ligand results in higher occupation of the open, second binding site and in faster exchange rates. This correlation between exchange rates and binding site occupancy shows that the flipping motions are directly induced by the presence of the second ligand in the open binding site.

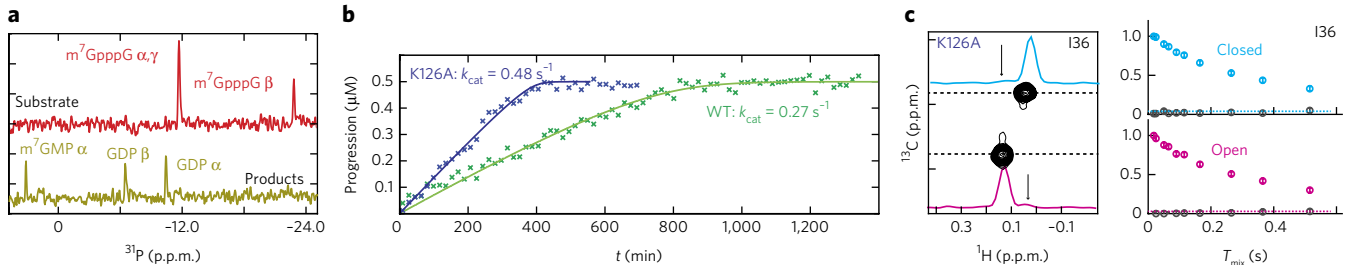


Figure 4 | Turnover rates of the enzyme are much slower than the flipping rates. (**a**) ^{31}P NMR spectra that report on the Dcs1p-mediated degradation (50 nM enzyme) of the (0.5 mM) m^{7}GpppG substrate (top, red) into m^{7}GMP and GDP products (bottom, yellow). NMR spectra (18 min each) were recorded successively until completion of the reaction. (**b**) Progression curves of the reaction. Crosses correspond to the mean concentrations of the substrate and product signals and are derived from the peak intensities of all ^{31}P signals. The drawn line corresponds to the best fit of the data. The extracted turnover rates are indicated. (**c**) Longitudinal exchange experiment of Dcs1p$^{H268N\ K126A}$ in the presence of a 20-fold excess of substrate and using a mixing time of 75 ms (at which the exchange peaks have maximum intensity for the Dcs1p^{H268N} protein). Left, exchange peaks were not observed (arrows in the traces), demonstrating that the K126A enzyme does not undergo unproductive flipping motions. Right, signal intensities of the resonances at different exchange delays (see also **Fig. 3b**).

A conformational change results in product release

The reaction products, m^{7}GMP and GDP, are formed in the closed pocket of the scavenger decapping enzyme after hydrolysis of the m^{7}GpppG substrate. Using methyl NMR spectroscopy, we observed that binding of the m^{7}GMP product to the enzyme preserves the asymmetric form of the complex (**Supplementary Fig. 5a**). During the next step in the catalytic cycle, the enzyme complex needs to open to release the product. To determine the affinity of the product for the closed binding site, we performed ITC measurements and found that the m^{7}GMP product binds to yeast Dcs1p in a single-step process (K_d = 1.1 ± 0.6 μM) (**Supplementary Fig. 5b**). The product thus interacts with the closed binding site four times more weakly than the substrate does. Notably, using ITC and NMR we were not able to detect an interaction between the m^{7}GMP product and the second, open binding site. This suggests a scenario in which hydrolysis of the substrate in the closed binding site and a flipping motion of the enzyme result in a situation where the product occupies the open binding site. In that case, m^{7}GMP would be efficiently released from the enzyme. To test this hypothesis we used methyl-TROSY NMR spectroscopy, in which we added an m^{7}GpppG substrate to the m^{7}GMP–Dcs1p complex. This resulted in the efficient release of product and the formation of an m^{7}GpppG–Dcs1p complex and shows that the enzyme can indeed spontaneously change from a product-bound state into a substrate-bound state upon addition of substrate (**Supplementary Fig. 5c**). In summary, our results indicate that the first and second binding sites can concomitantly open and close during the catalytic cycle.

Dcs1p undergoes spontaneous domain-flipping motions

To observe whether the scavenger decapping enzyme undergoes domain-flipping motions in conditions of excess m^{7}GpppG substrate, we performed longitudinal exchange experiments[29,33,36] on highly deuterated methyl-labeled samples. In these experiments the conversion of an open binding site into a closed one and vice versa can be observed directly because the exchange process results in cross peaks at resonance frequencies where one dimension corresponds to the open (or closed) state and the other dimension corresponds to the closed (or open) state. These experiments are possible because NMR is unique in the fact that it can distinguish between the open and closed states within one enzyme complex (**Fig. 2b,c**). Here we observed exchange peaks for all of the residues that display a different chemical shift in the open as compared to the closed state, indicating that the enzyme undergoes continuous domain-flipping motions (**Fig. 3a**). Measurements using a number of different exchange delays in these experiments allowed for the extraction of the exchange rates and the associated populations. For all of the residues that displayed well-resolved resonances for the auto and cross peaks, we determined very similar exchange parameters, indicating that the process we detect results from a global conformational change of the type associated with the N-terminal domain flipping on the C-terminal domain. In addition, the extracted populations of the open and closed states were, as expected, equal, as the number of open and closed sites in the enzyme is by definition 1:1. To further analyze the data, we thus used a global fit, in which all residues were fitted to one exchange process and in which the open- and closed-state populations (p_{closed} and p_{open}) were fixed to 0.5. In the presence of a 20-fold excess of substrate over enzyme, we extracted a domain-flipping rate of 35.4 ± 4.8 s^{-1} (**Fig. 3b**). Notably, the flipping motions that we detected are not a result of enzymatic turnover and subsequent product release because we used an inactive (H268N) mutant. Rather, the motions report on a situation in which an occupied closed binding site turns into an occupied open binding site and vice versa.

The second substrate induces domain-flipping motions

The second (open) binding site in Dcs1p is saturated with substrate (**Fig. 2d**) under the conditions in which we observed N-terminal domain–flipping motions. To probe whether the substrate in the open binding site induced these motions, we measured the domain-flipping rates at decreasing substrate concentrations. Notably, the domain-flipping rates decreased with reduced occupation of the second binding site (**Fig. 3c**). These results display the presence of an allosteric pathway that links both binding sites in the enzyme. Such an allosteric pathway has previously been suggested on the basis of biochemical experiments[25], but the molecular basis underlying the cross-talk between the active sites was not known. Here we demonstrate that substrate binding in the open binding site of the enzyme directly induces domain rearrangements and causes the closed binding site to open and the open binding site to close.

The catalytic rates are slower than the flipping rates

To assess whether the substrate-induced flipping motions in the enzyme correlated with catalytic turnover, we determined the activity of Dcs1p in conditions of excess substrate. To that end, we continuously recorded phosphorus NMR spectra during enzymatic turnover. The initial spectra (t = 0) displayed the phosphorus resonances of m^{7}GpppG and the final spectra displayed resonances of the products, m^{7}GMP and GDP (**Fig. 4a**). From these data we extracted a turnover rate (k_{cat}) of 0.27 s^{-1} (**Fig. 4b**), which agreed with the turnover rate (under high substrate concentrations) that was previously determined for the human enzyme complex[25].

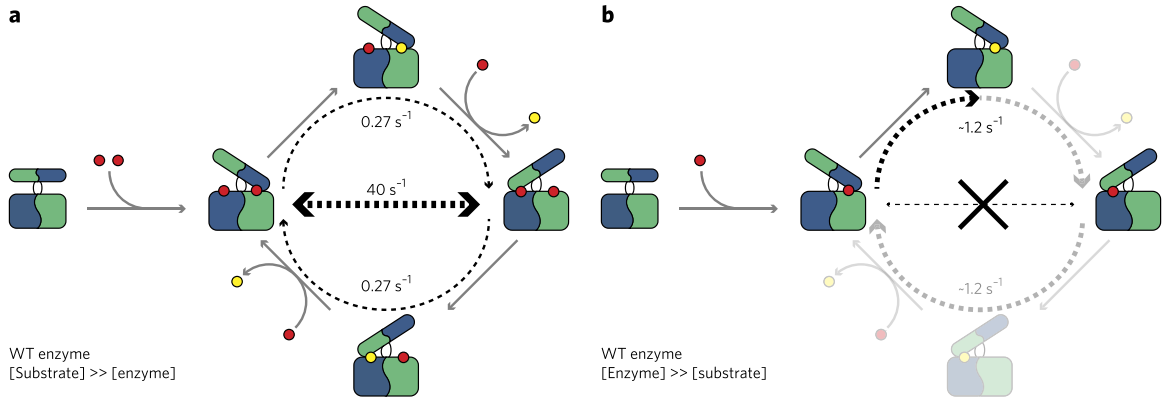


Figure 5 | Cartoon representation of the substrate inhibition mechanism of the scavenger decapping enzyme. (**a**) Under conditions of substrate excess, both binding sites in the enzyme are occupied. These result in fast, nonproductive flipping motions (horizontal dashed arrow) and low turnover rates (curved dashed arrows). Substrate and product are indicated by red and yellow circles, respectively. Determined flipping and turnover rates are indicated. (**b**) Under single-turnover conditions, only one of the two binding sites in the enzyme is occupied. The unproductive motions (crossed out horizontal line) no longer take place and the catalytic turnover is increased (curved dashed arrow). The indicated turnover rate corresponds to that of the human enzyme[25] and may differ for the yeast complex. Note that under these conditions part of the catalytic cycle is not sampled (indicated in a light color for reference only).

However, it is important to note that the turnover rates at high substrate excess are two orders of magnitude slower than those of the domain-flipping motions (35.4 s^{-1}) that we have measured (**Fig. 3**).

A point mutation interferes with fast flipping motions

As detailed above, we have determined that the N-terminal domain senses the presence of substrate in the open binding pocket (**Fig. 2c**). The structure of the enzyme (**Fig. 1**) features positively charged amino acids (such as K126) in the hinge region. We reasoned that these residues could be part of the allosteric pathway that links the occupation of the open binding site to a conformational change. To assess this hypothesis, we mutated K126 to an alanine and used longitudinal exchange experiments to probe whether the domain-flipping motions in this mutant enzyme were altered. Indeed, under conditions of substrate excess, the domain motions that we detected in the $Dcs1p^{H268N}$ enzyme were no longer detectable in the $Dcs1p^{H268N\ K126A}$ mutant enzyme (**Fig. 4c**). Taking into account the detection limits of the longitudinal exchange experiment, the domain-flipping motions were reduced from 35.4 s^{-1} in the $Dcs1p^{H268N}$ protein to <1 s^{-1} in the $Dcs1p^{H268N\ K126A}$ mutant. It is important to note that $Dcs1p^{H268N\ K126A}$ behaved like $Dcs1p^{H268N}$ with respect to the formation of an asymmetric conformation and interaction with substrate (**Supplementary Fig. 3c**). Indeed, the affinity for the substrate did not change substantially when the K126A mutation was introduced into the enzyme (**Table 1**). We thus found that a single point mutation in Dcs1p is able to uncouple the binding of substrate in the second (open) binding pocket from fast domain rearrangements.

Reduced domain-flipping motions lead to hyperactivity

To probe whether reduced conformational exchange influences enzymatic activity, we measured the activity of the K126A enzyme and found that it displayed two-fold higher activity (**Fig. 4b**) compared to the WT enzyme. It is important to note that the increased catalytic efficiency is not due to changes in substrate binding (**Supplementary Fig. 3a**,**b** and **Table 1**). In addition, potential changes in product release are likely not the cause of the observed hyperactivity because for the human enzyme it has been established that product release is not the rate-limiting step in the catalytic cycle[25].

Hyperactive mutants of DcpS have been described for the human enzyme[22]; however, the molecular basis for this hyperactivity has remained elusive. Here we show that unproductive domain-flipping motions decrease the catalytic efficiency of the enzyme. Interference with the underlying allosteric pathway through mutations reduces the unproductive motions and results in hyperactivity. Hence, domain-flipping motions in the enzyme that are important for the catalytic cycle (**Supplementary Figs. 4** and **5**) can reduce activity when they are too fast (**Figs. 3** and **4**).

DISCUSSION

The correlation between enzyme motions and catalytic activity remains poorly understood despite the fact that this relationship was proposed decades ago[37].

Here we studied the dynamic changes that occur during the catalytic cycle of the yeast scavenger decapping enzyme. Static structures of the enzyme[21,22] (**Fig. 1** and **Supplementary Fig. 1**) suggest that domain rearrangements are important for substrate binding and product release. We showed that the formation of a stable asymmetric conformation is indispensable for activity (**Fig. 2** and **Supplementary Figs. 4** and **5**). In addition, we established that the interaction of substrate with the second (open) binding pocket of the dimeric enzyme induces unproductive motions that impair catalytic efficiency (**Figs. 3** and **4**).

At high substrate concentrations the hinge motions interfered with efficient catalysis (**Fig. 5a**) because the time the substrate spends in the closed active site (life time) is on average too short to allow for efficient hydrolysis: before catalysis can take place, the enzyme rearranges into its mirror form (**Fig. 5**). This implies that the rate-limiting step in the catalytic cycle is after the formation of the closed binding site (such as structural rearrangements in the substrate). Previous studies indicate that, for the human enzyme, such a conformational change is relevant for the activation of cap hydrolysis[22].

Using a designed mutant enzyme (K126A), we uncoupled the second substrate-binding event from the induction of unproductive motions, which resulted in hyperactivity (**Fig. 4** and **Supplementary Fig. 6**). Mechanistically, we propose that electrostatic interactions between the positively charged residues in the hinge region (K126 and K125) and the negatively charged phosphate groups of the substrate in the open binding site induce motions of the N-terminal domain. The conservation of positively charged residues in the hinge region points toward a preserved electrostatic mechanism that regulates the activity of the scavenger decapping enzyme family. In agreement with that, mutations of positively charged hinge residues in the human enzyme (such as K138) also lead to hyperactivity[22].

It has previously been shown that the human DcpS enzyme is more active at low substrate concentrations than it is at high

substrate concentrations[25]. The physiological relevance of substrate inhibition has been demonstrated for a variety of enzymes, including phosphofructokinase, tyrosine hydroxylase and acetylcholinesterase[38]. DcpS is part of a complex enzymatic network that eliminates mRNA cap structures[39], and substrate inhibition could play a role in a regulatory feedback loop. In addition, substrate inhibition of DcpS might be required to limit cellular levels of the product m^7GMP, which can be misincorporated into RNA[18]. The model of the catalytic cycle of DcpS that we introduced here (**Fig. 5**) provides the molecular explanation for the observed substrate-mediated inhibition of DcpS activity (**Fig. 5b**). At low substrate concentrations, the second DcpS binding site is not occupied (**Fig. 2**) and the domain-flipping motions are significantly reduced or absent (**Figs. 3** and **5b**). This results in an increase in the catalytic turnover rate that can reach 1.2 s^{-1} in the human enzyme[25]. To our knowledge, this is the first time that a mechanism for substrate inhibition through an excess of protein motions has been described.

The substrate-induced molecular motions provide an additional layer of regulation for the activity of the scavenger decapping enzyme. In that regard it is worth mentioning that the enzyme has been reported to interact with a number of partners, including the exosome complex in humans[16] and Xrn1 (ref. 40) and Dcs2p (ref. 18) in *S. cerevisiae*. Dcs2p has been reported to inhibit Dcs1p activity by heterodimerization, which potentially influences the dynamics in the complex. Currently, efforts to pharmaceutically inhibit DcpS are under way[19,20]. We demonstrate here that DcpS operates in a narrow range of dynamics. Thus both small compounds that block interdomain motions and those that accelerate them can impair activity. Finally, DcpS displays different activities toward substrates of various RNA lengths. The mechanism by which the enzyme discriminates short substrates from longer ones[35] may be through the modulation of domain motions by the RNA body. Our results provide the conceptual framework to address these questions in detail.

In summary, we present here an example of an enzyme in which conformational changes that are required for activity can be detrimental for catalysis in the event they are too fast. This insight could not have been obtained from knowledge of static structures alone and underscores the need for studies that localize and quantify motions in large biomolecular assemblies. We expect these studies to elucidate more instances of complex inter-relationships between dynamics and activity and thus greatly advance our understanding of biomolecular function.

Received 19 December 2014; accepted 1 June 2015;
published online 10 August 2015

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. PDB: the coordinates for Dcs1p in complex with m^7GDP are deposited under accession code 5BV3.

References

1. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, 1999).
2. Hammes-Schiffer, S. & Benkovic, S.J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541 (2006).
3. Boehr, D.D., Dyson, H.J. & Wright, P.E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–3079 (2006).
4. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
5. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
6. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
7. McGeagh, J.D., Ranaghan, K.E. & Mulholland, A.J. Protein dynamics and enzyme catalysis: insights from simulations. *Biochim. Biophys. Acta* **1814**, 1077–1092 (2011).
8. van den Bedem, H. & Fraser, J.S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat. Methods* **12**, 307–318 (2015).
9. Eisenmesser, E.Z., Bosco, D.A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520–1523 (2002).
10. Boehr, D.D., McElheny, D., Dyson, H.J. & Wright, P.E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
11. Henzler-Wildman, K.A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
12. Wolf-Watz, M. *et al.* Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **11**, 945–949 (2004).
13. Závodszky, P., Kardos, J., Svingor, Á. & Petsko, G.A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. USA* **95**, 7406–7411 (1998).
14. Agarwal, P.K., Billeter, S.R., Rajagopalan, P.T., Benkovic, S.J. & Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **99**, 2794–2799 (2002).
15. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
16. Wang, Z. & Kiledjian, M. Functional link between the mammalian exosome and mRNA decapping. *Cell* **107**, 751–762 (2001).
17. Liu, H., Rodgers, N.D., Jiao, X. & Kiledjian, M. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J.* **21**, 4699–4708 (2002).
18. Malys, N. & McCarthy, J.E. Dcs2, a novel stress-induced modulator of m7GpppX pyrophosphatase activity that locates to P bodies. *J. Mol. Biol.* **363**, 370–382 (2006).
19. Singh, J. *et al.* DcpS as a therapeutic target for spinal muscular atrophy. *ACS Chem. Biol.* **3**, 711–722 (2008).
20. Gogliotti, R.G. *et al.* The DcpS inhibitor RG3039 improves survival, function and motor unit pathologies in two SMA mouse models. *Hum. Mol. Genet.* **22**, 4084–4101 (2013).
21. Chen, N., Walsh, M.A., Liu, Y., Parker, R. & Song, H. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J. Mol. Biol.* **347**, 707–718 (2005).
22. Gu, M. *et al.* Insights into the structure, mechanism and regulation of scavenger mRNA decapping activity. *Mol. Cell* **14**, 67–80 (2004).
23. Han, G.W. *et al.* Crystal structure of an apo mRNA decapping enzyme (DcpS) from mouse at 1.83 Å resolution. *Proteins* **60**, 797–802 (2005).
24. Séraphin, B. The HIT protein family: a new family of proteins present in prokaryotes, yeast and mammals. *DNA Seq.* **3**, 177–179 (1992).
25. Liu, S.W., Rajagopal, V., Patel, S.S. & Kiledjian, M. Mechanistic and kinetic analysis of the DcpS scavenger decapping enzyme. *J. Biol. Chem.* **283**, 16427–16436 (2008).
26. Pentikäinen, U., Pentikäinen, O.T. & Mulholland, A.J. Cooperative symmetric to asymmetric conformational transition of the apo form of scavenger decapping enzyme revealed by simulations. *Proteins* **70**, 498–508 (2008).
27. Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E. & Kay, L.E. Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high-molecular-weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).
28. Sprangers, R. & Kay, L.E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).
29. Audin, M.J. *et al.* The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding. *Angew. Chem. Int. Edn Engl.* **52**, 8312–8316 (2013).
30. Gelis, I. *et al.* Structural basis for signal sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131**, 756–769 (2007).
31. Rosenzweig, R. & Kay, L.E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* **83**, 291–315 (2014).
32. Wypijewska, A. *et al.* 7-methylguanosine diphosphate (m^7GDP) is not hydrolyzed but strongly bound by decapping scavenger (DcpS) enzymes and potently inhibits their activity. *Biochemistry* **51**, 8003–8013 (2012).
33. Sprangers, R., Gribun, A., Hwang, P.M., Houry, W.A. & Kay, L.E. Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl. Acad. Sci. USA* **102**, 16678–16683 (2005).
34. Amero, C. *et al.* A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR* **50**, 229–236 (2011).
35. Liu, S.W. *et al.* Functional analysis of mRNA scavenger decapping enzymes. *RNA* **10**, 1412–1422 (2004).

36. Farrow, N.A., Zhang, O., Forman-Kay, J.D. & Kay, L.E. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**, 727–734 (1994).
37. Hammes, G.G. Mechanism of enzyme catalysis. *Nature* **204**, 342–343 (1964).
38. Reed, M.C., Lieb, A. & Nijhout, H.F. The biological significance of substrate inhibition: a mechanism with diverse functions. *Bioessays* **32**, 422–429 (2010).
39. Taverniti, V. & Seraphin, B. Elimination of cap structures generated by mRNA decay involves the new scavenger mRNA decapping enzyme Aph1/FHIT together with DcpS. *Nucleic Acids Res.* **43**, 482–492 (2015).
40. Sinturel, F., Brechemier-Baey, D., Kiledjian, M., Condon, C. & Benard, L. Activation of 5′-3′ exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. *Proc. Natl. Acad. Sci. USA* **109**, 8264–8269 (2012).

Acknowledgments

We acknowledge J. Peters for excellent technical support, S. Wiesner for discussions, G. Zocher for assistance in recording diffraction data, T. Stehle for support and L. Kay for valuable comments on the manuscript. This work has received funding from the Max Planck Society and the European Research Council under the EU's Seventh Framework Programme (FP7/2007–2013), ERC grant agreement no. 616052.

Author contributions

R.S. conceived the project. A.N. and R.S. designed the experiments. All authors performed experiments. R.S., A.N., U.N. and A.-L.F. analyzed and interpreted the data. R.S. wrote the manuscript and all authors commented on the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to R.S.

ONLINE METHODS

Protein preparation. The *DCS1* gene from *S. cerevisiae* was amplified from genomic DNA and cloned into a modified pET vector that contained an N-terminal His_6-NusA-His_6 tag that could be removed using TEV protease. Mutations were introduced using standard quick-change methods. BL21 (DE3) codon plus cells (Stratagene) were transformed with the respective plasmid and grown in rich or minimal medium in the presence of kanamycin and chloramphenicol. Minimal medium was made using 100% D_2O and contained [$^2H^{12}C$]glucose as a carbon source. Isoleucine and methionine labeling was achieved by addition of 50 mg/liter α-ketobutyric acid and 250 mg/liter of [methyl-$^{13}CH_3$]methionine, respectively, 1 h before induction with 1.0 mM IPTG at a cell density (OD_{600}) of 0.8 and a temperature of 25 °C. 12 h after induction the cells were lysed in buffer A (50 mM $NaPO_4$ pH 8.0, 150 mM NaCl and 5 mM imidazole) supplemented with 0.1% Triton, 1 mM EDTA and 100 µg/ml lysozyme. Subsequently the cell lysate was centrifuged at 50,000*g* for 20 min to remove insoluble debris. The supernatant, supplemented with 2 mM $MgCl_2$, was applied to Ni-NTA resin and the column was washed using 10 column volumes of buffer A. The protein was eluted from the resin using buffer A supplemented with 200 mM imidazole. The affinity and solubility tag was removed from the protein by TEV cleavage during dialysis into 25 mM Tris pH 8.0, 75 mM NaCl, 1 mM DTT. The cleaved tag was removed from the target protein using a second Ni-NTA step in dialysis buffer. Finally a size exclusion chromatography purification step in buffer B (25 mM HEPES pH 8.0, 25 mM NaCl) using a Superdex 200 column yielded pure protein. For NMR, the final protein samples were exchanged by successive concentration and dilution steps into buffer B that was made using 100% D_2O (NMR buffer).

Crystallography. Dcs1p in complex with m^7GDP crystallized at a concentration of 10 mg/ml in 100 mM HEPES pH 7.5, 100 mM NaCl, 1.6 M $(NH_4)_2SO_4$ after several weeks. Diffraction data were collected at 100 K using a wavelength of 1 Å and a PILATUS 6 M detector at beamline PXII of the Swiss Light Source (PSI, Villigen, Switzerland). Data were processed using XDS[41] and molecular replacement was performed using Phaser[42] with the crystal structure of the human protein in complex with m^7GDP (1XMM)[21] as a search model. The structure was finalized by iterative manual modeling with Coot[43] and refinement with Phenix[44] and Refmac[45]. The crystalized protein contained a cysteine bridge in a loop region of the N-terminal domain (Cys72-Cys78). This crystallization artifact does not interfere with the data analysis, as the activity of the enzyme does not change between oxidizing and reducing conditions (**Supplementary Fig. 7**). All figures displaying protein structures were generated using PyMOL (http://www.pymol.org/).

NMR. NMR spectra were recorded at 27 °C on Bruker AVIII-600 and AVIII-800 spectrometers with room temperature probe heads. All NMR samples contained buffer made in 100% D_2O. For degradation experiments 0.5 mM m^7GpppG substrate was mixed with 50 nM Dcs1p in NMR buffer and successive phosphorus experiments were recorded over a time of several hours. Proton-carbon–based experiments exploited the methyl-TROSY principle and were recorded using a carbon chemical shift evolution of typically 40 ms. Longitudinal exchange experiments were recorded on samples containing 0.25 mM Dcs1p (dimer concentration) in the presence of a 1:1, 1:2.5, 1:5 or 1:20 molar ratio of m^7GpppG using mixing times ranging from 3 to 1,000 ms, depending on the exchange rate. NMR data were processed using the NMRPipe-NMRDraw software suite[46] and figures displaying NMR spectra were produced using NMRView (onemoonscientific.com). NMR binding affinities were determined as described in the legend of **Supplementary Figure 3a**. All data points of the titration were simultaneously fitted to one sequential binding event. The extracted K_d values indicate that the first binding event takes place in the high-nanomolar range and that the second binding event has a dissociation constant of 105 ± 20 µM. The error bars represent uncertainties in the peak intensity or peak position. The extracted K_d values represent mean values ± s.d. and are based on Monte Carlo simulations where the data points were randomly varied according to their errors. The large uncertainty in the K_d value for the first binding event results from the limited number of titration points that define the event.

Isothermal titration calorimetry (ITC). Binding constants were determined using a NanoITC instrument (TA Instruments) at 25 °C. m^7GpppG or m^7GMP were injected into a 0.5 mM Dcs1p (all variants carried the H268N mutation) solution (either full-length or C-terminal domain only). Data were processed using the in-house–written scripts (see legend to **Supplementary Fig. 3a**). Errors in the extracted K_d values were derived on the basis of Monte Carlo simulations in which the individual heat emissions were randomly varied depending on the noise level in the experiment.

41. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
42. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
43. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
44. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
45. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
46. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

Supplementary information

An excess of catalytically required motions inhibits the scavenger decapping enzyme

Ancilla Neu[1], Ursula Neu[2,3], Anna-Lisa Fuchs[1], Benjamin Schlager[1,4], Remco Sprangers[1*]

[1] Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

[2] Interfaculty Institute of Biochemistry, University of Tübingen, Hoppe-Seyler-Strasse 4, 72076, Tübingen, Germany

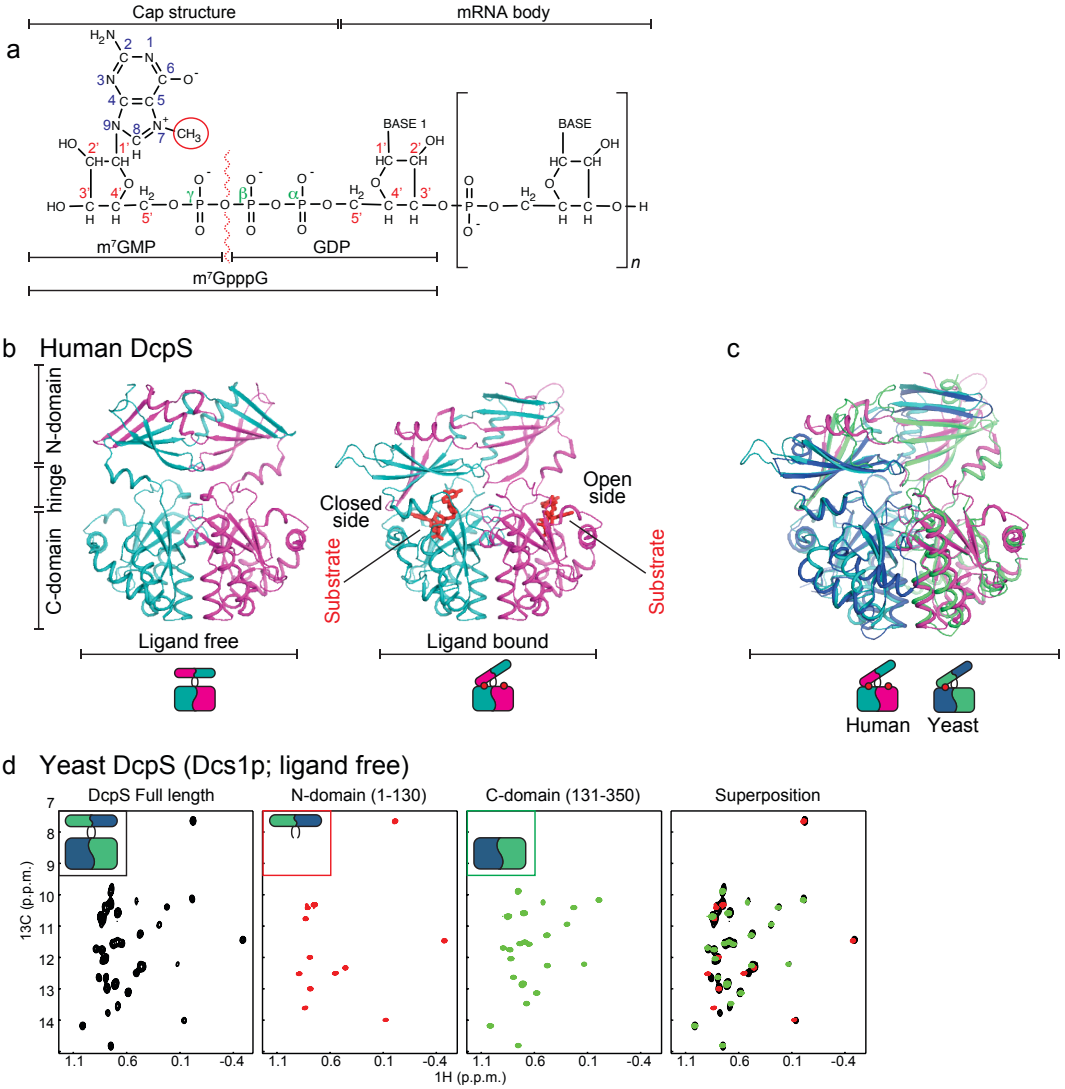
[3] Current address: The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

[4] Current address: FluidSolids AG, Zürich, Switzerland

[*] To whom correspondence should be addressed:
remco.sprangers@tuebingen.mpg.de

Nature Chemical Biology: doi:10.1038/nchembio.1866

Supplementary Results



Supplementary Figure 1 | Structures of the human DcpS and spectra of the isolated yeast enzyme.

(a) Chemical structure of an mRNA with a 5' protecting cap. The nomenclature of the sugars, base and phosphate backbone are indicated in red, blue and green respectively. The methyl group in the cap structure is indicated with a red sphere. The scavenger decapping enzyme hydrolyses the phosphate 5' 5' linkage at the position indicated with a red waved line for mRNA species with n < 10. For the m^7GpppG substrate used in this study n=0 and the first base (base 1) is a guanine. The products of the hydrolysis of m^7GpppG (m^7GMP and GDP) are indicated.

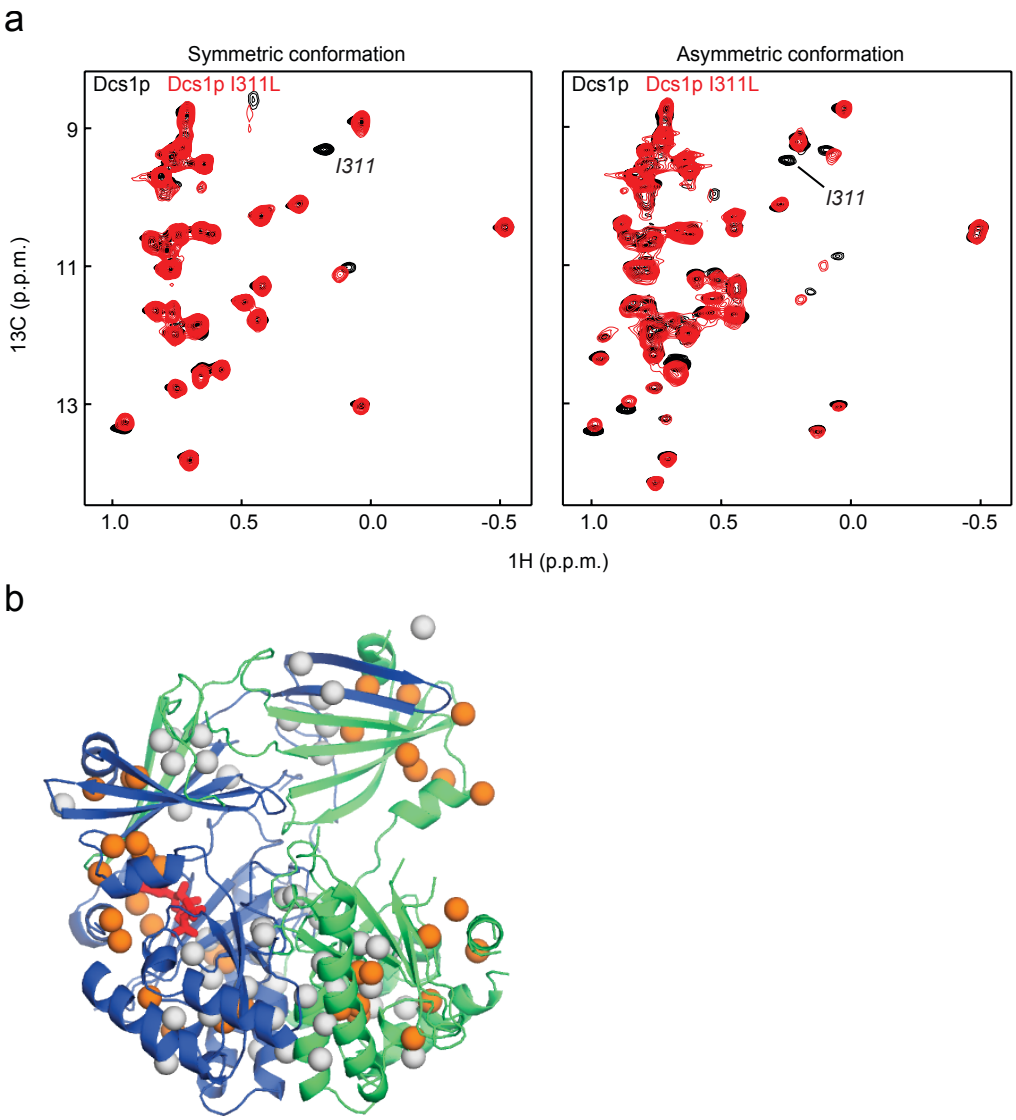
Nature Chemical Biology: doi:10.1038/nchembio.1866

(b) The human DcpS enzyme contains two protein chains that form a dimeric N-terminal lid domain and a dimeric C-terminal domain. The ligand free protein is symmetric (left)[1]. In the substrate bound protein the lid domain undergoes a large conformational change (right)[2]. In the structure of the human protein two substrates are bound to one dimeric enzyme (one in the open binding site and one in the closed binding site).

(c) Superposition of the structures of the human (1ST0)[2] and yeast (this study) scavenger decapping enzymes in complex with ligand. For clarity the ligands are not displayed in the ribbon diagram.

(d) Methyl TROSY spectra of the yeast full length enzyme (left panel) and the isolated N- and C- terminal domains (middle panels). A superposition of the spectra in the first three panels is shown on the right. The spectra of the isolated domains overlay very well with the spectrum of the full-length protein indicating that the two domains behave as independent units in the full-length protein.

Nature Chemical Biology: doi:10.1038/nchembio.1866

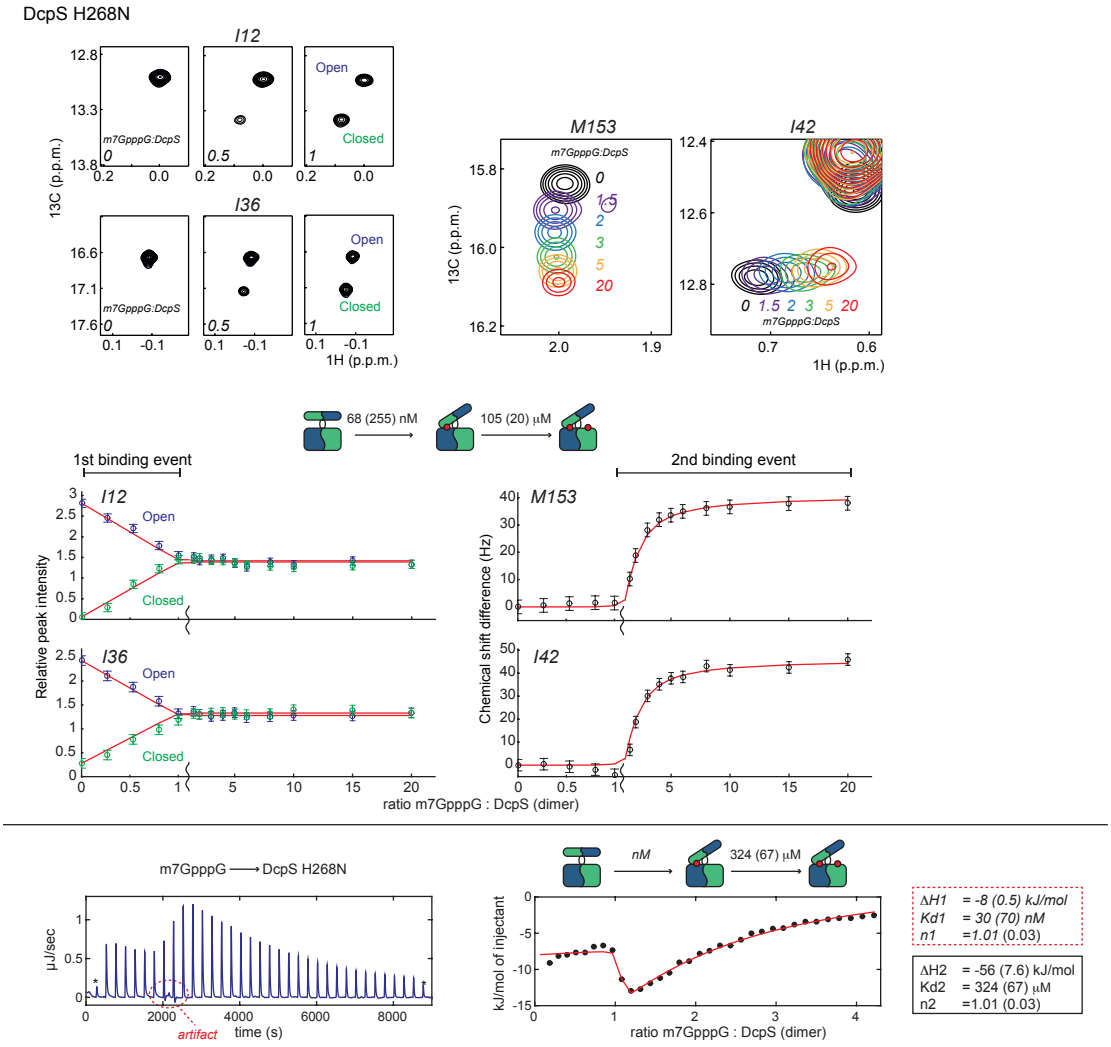


Supplementary Figure 2 | Assignment of methyl groups in the yeast scavenger decapping enyme.

(a) Example of the mutational approach for the methyl group assignments. Single Ile or Met residues were mutated into closely related amino acids. Spectra of the WT and mutant protein where then compared. An overlay of the spectra of the WT (black) and I311 (red) Dcs1p allows for the assignment of I311 in the symmetric (left) and asymmetric (right) conformation of the enzyme.

(b) Structure of the yeast Dcs1p enzyme (see also Figure 1), where all Isoleucine-δ1 and methionine methyl groups are shown as spheres. The assigned methyl groups are coloured orange.

Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 3a| Determination of the affinities of the first and second binding event for the interaction between WT (H268N) DcpS and m^{7}GpppG substrate.

Top: Selected regions of NMR spectra during the NMR titration (regions are indicated with a box in Figure 2b and 2c). The m^{7}GpppG:DcpS (dimer) ratios are indicated in the spectra. The spectra on the left report on the fist binding event (that is in slow exchange on the NMR chemical shift timescale), the spectra on the right report on the second binding event (that is fast on the NMR chemical shift timescale).

Middle: Identical to Figure 2 of the main paper, shown her for completeness. Plot of the changes in the peak intensities for residues I12 and I36 (open and closed resonances) and the changes in the peak position for residues M153 (CSP in

Nature Chemical Biology: doi:10.1038/nchembio.1866

carbon) and I42 (CSP in protons) for 15 titration points (see a-c). All data points of the titration were simultaneously fitted to one sequential binding event (see below). The extracted kD values indicate that the first binding event takes place in the high nano-molar range and that the second binding event has a dissociation constant of 105 (±20) μM. The large uncertainty in the kD value for the first binding event results from the limited number of titration points that define the event. For clarity the scale of the x-axis is shown different below and above the 1:1 m^7GpppG:DcpS ratio, as is indicated by the waveform.

Bottom: The ITC thermograph of the interaction between the yeast scavenger decapping enzyme (H268N) and m^7GpppG substrate displays three events. One of these events could be a result of protein aggregation in the ITC cell (indicated with a red circle); note that no protein aggregation or precipitation has been observed in any of our NMR experiments. The red line in the right panel corresponds to the best fit of the data using a model of sequential binding. This third event prevents the extraction of accurate binding values for the first binding event (indicated with a red dotted box around the indicated extracted values). The kD for the second binding event extracted using NMR is in agreement with the one extracted from the ITC experiment. Note that n=1 corresponds to the stoichiometry of the first and of the second binding event, that both have an occupancy of 1.

In order to extract affinities from the ITC data, we performed baseline correction and peak integration using the NanoAnalyze program (TA instruments) and used in house written scripts to extract the associated binding affinities (the software provided by the manufactures did not include the sequential binding model).

In the case of sequential binding, a mixture of free protein, protein with one ligand bound and protein with two ligands bound can be formed. The associated fractions (F_0, F_1 or F_2) are given by:

$$F_0 = \frac{1}{P} \qquad F_1 = \frac{K_a^1\,[L]}{P} \qquad F_2 = \frac{K_a^1 K_a^2\,[L]^2}{P}$$

Nature Chemical Biology: doi:10.1038/nchembio.1866

in which [L] is the free ligand concentration and

$$P = 1 + K_a^1 [L] + K_a^1 K_a^2 [L]^2$$

where K_a^1 and K_a^2 refer to the binding constants associated with the first and second binding event, respectively, that are defined according to:

$$K_a^1 = \frac{[PL]}{[P][L]} \qquad K_a^2 = \frac{[PL_2]}{[PL][L]}$$

[P] is the concentration of free protein, [PL] is the concentration of singly occupied protein and $[PL_2]$ is the concentration of double occupied protein. The concentration of free ligand [L] is determined numerically and iteratively based on the binding constants and total ligand concentration L_t at a given point during the titration, where

$$L_t = [L] + P_t * (F_1 + 2F_2)$$

and P_t is the total protein concentration.

In an ITC experiment the heat content at a certain time during the titration is given by:

$$Q = P_t V_0 [F_1 \Delta H_1 + F_2 (\Delta H_1 + \Delta H_2)]$$

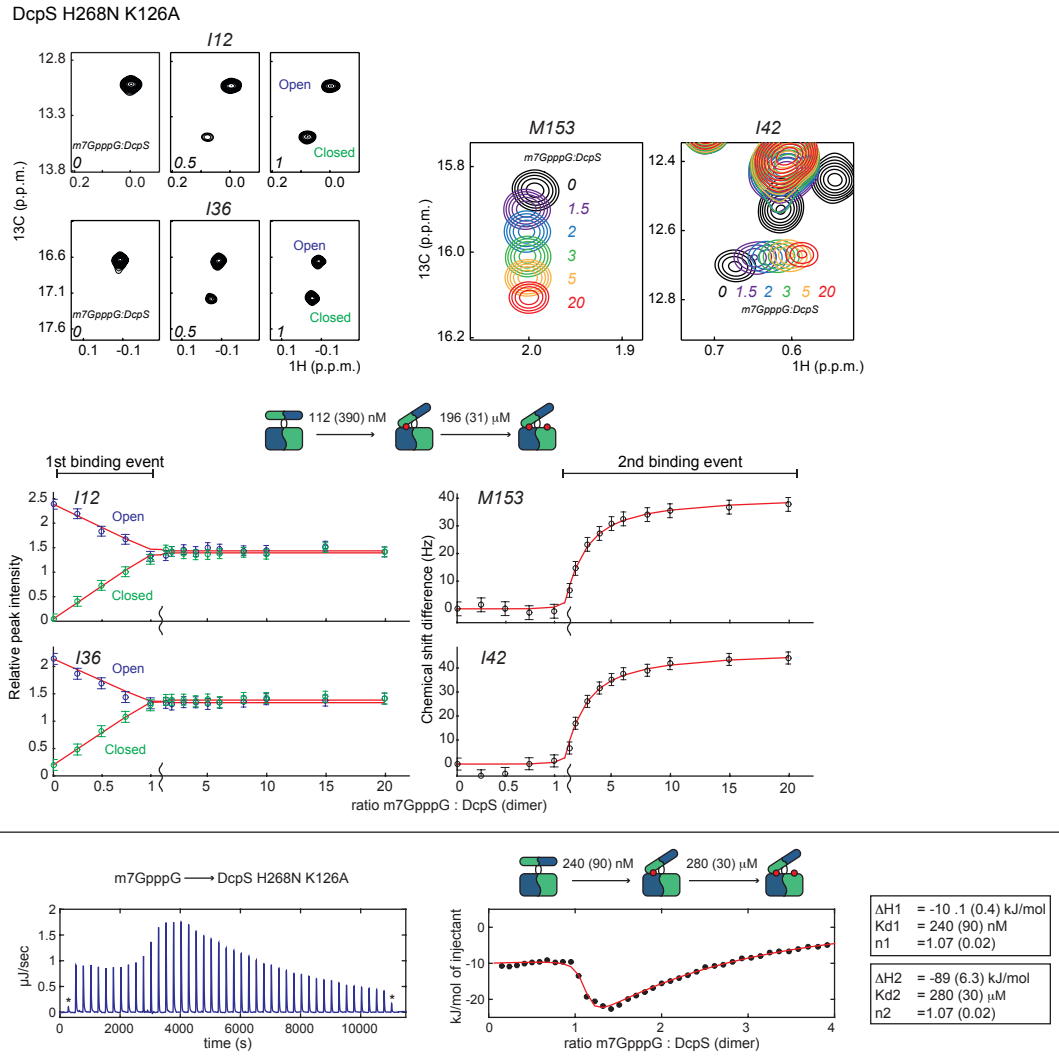
where V_0 is the volume of the ITC cell and ΔH_1 and ΔH_2 are molar heats of ligand binding for the first and second binding steps respectively. In the fitting procedure, the parameters ΔH_1, ΔH_2, K_a^1, K_a^2 and n (the total number of binding events per step) are optimized to minimize the difference between the observed and calculated heat content at all steps of the titration.

Nature Chemical Biology: doi:10.1038/nchembio.1866

In the NMR titration experiments, the first binding event is slow on the chemical shift timescale. Thus, the intensity of the original peaks directly reports on the fraction of free protein (F_0), whereas the newly appearing peaks report on the sum of the singly and double occupied protein (F_1+F_2) (in case the second binding event does not perturb the position of the newly appearing resonance). The second binding event is fast on the NMR timescale and therefore, the peak position reports on the fraction of doubly bound protein F_2. To extract the affinities from the NMR titration experiments the peak intensities that report on the first binding event and the peak positions that report on the second binding event were simultaneously fitted. To that end the parameters K_a^1, K_a^2, n, the resonance position of the fully bound second binding site and a scaling factor for each of the two slow exchanging resonances (to correct for potential differences in line-width of the methyl resonances in the closed and open sites) were optimized to minimize the difference between the observed and calculated spectral properties at all steps of the titration.

Errors in all fits are based on 100 MC (Monte Carlo) simulations where the data points used for the fitting were randomly varied based on the uncertainties in the experimental data.

Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 3b| Determination of the affinities of the first and second binding event for the interaction between K126A (H268N) DcpS and m^{7}GpppG substrate.

Top: Regions of the NMR spectra during the NMR titration experiment. See legend to figure S3a.

Middle: Fit of the peak intensities and positions result in a kD of the first binding event that is in the high nano-molar range and a kD for the second binding event that is 196 (±31) µM.

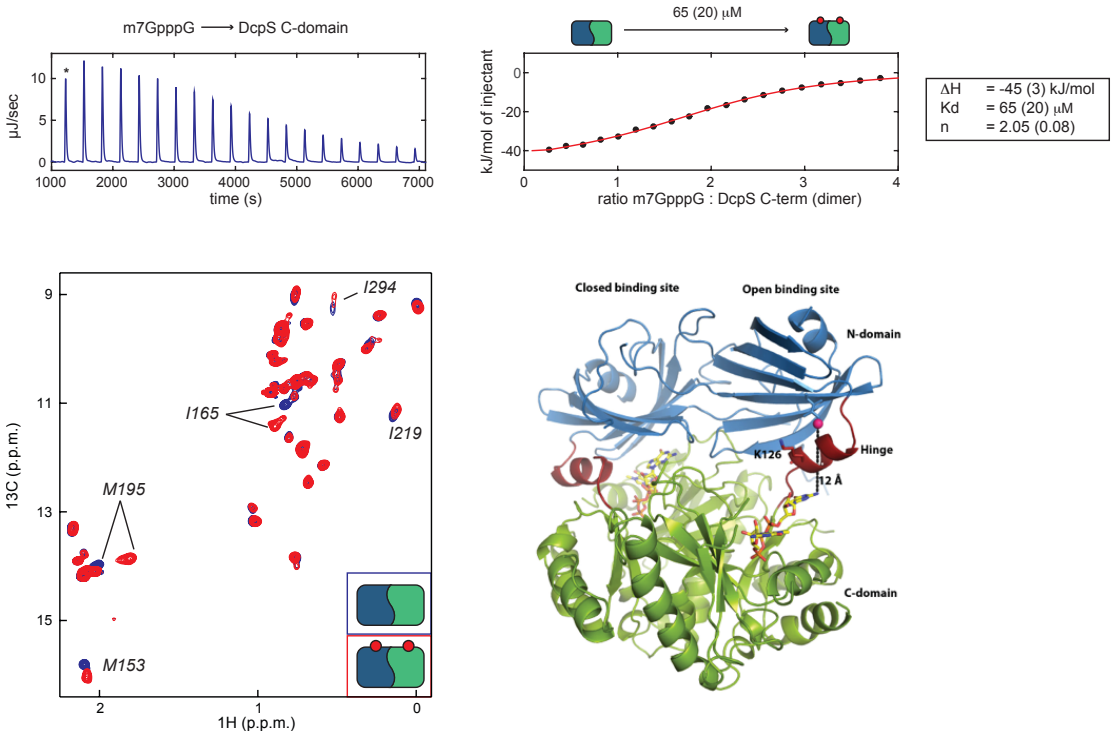
Bottom: ITC thermogram for the interaction of Dcs1p (H268N, K126A) with m^{7}GpppG that displays a sequential interaction mode. The red line corresponds to the fit of the data; the extracted binding constants are indicated next to the graph. Please note that the third event that we observed in the ITC experiments

Nature Chemical Biology: doi:10.1038/nchembio.1866

with the WT protein is not observed in the ITC experiment using the K126A mutant. As a result, the kD for the first binding event can be extracted from the ITC data for the K126A mutant.

It is important to note that the affinities of m^7GpppG for the WT (H268N) and the K126A (H268N) enzymes are very similar. In addition, the NMR spectra during the titration of the substrate to both forms of the enzyme show exactly the same behaviour (compare Figures S3a and S3b). This establishes that both proteins use the same sequential binding mechanism and the K126A mutation does not change the mechanism of the enzyme.

Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 3c | Analysis of the interaction between the C-terminal domain of Dcs1p and m^{7}GpppG.

Top: ITC thermogram for the interaction of the C-terminal domain of Dcs1p with m^{7}GpppG substrate. The affinity of the substrate to the C-terminal domain is comparable to the affinity of the substrate to the open (second) binding site in the full-length enzyme. This confirms that the interaction of the substrate in the open binding site is mainly mediated through the C-terminal domain.

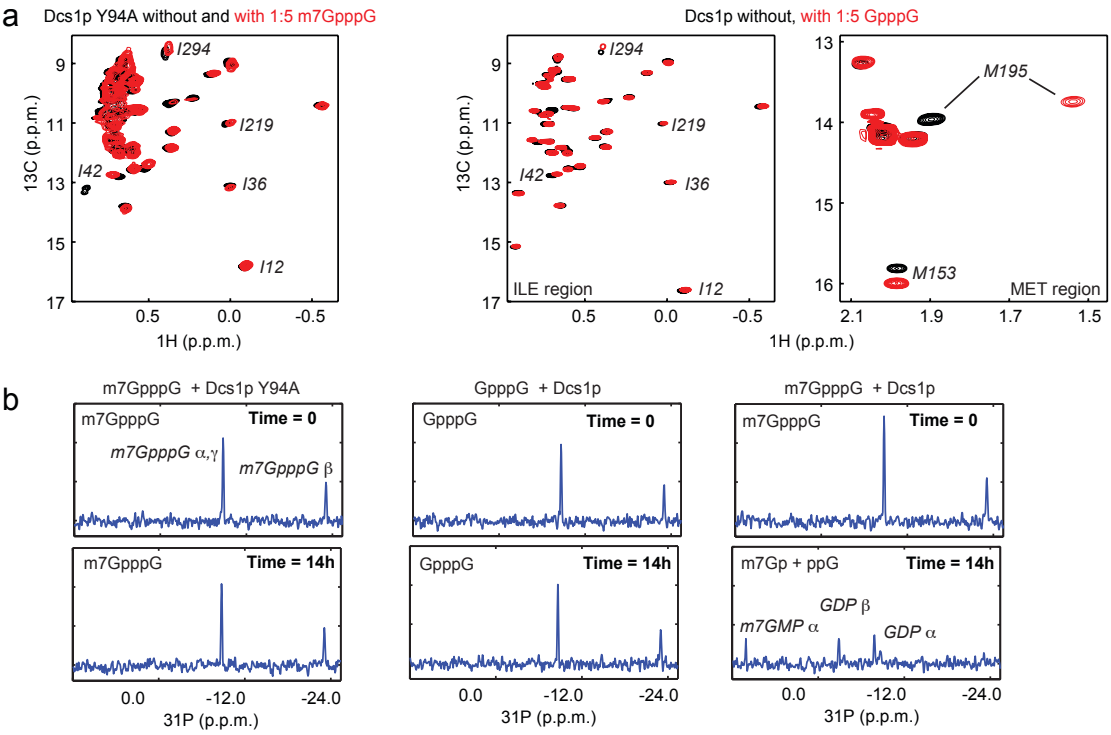
Bottom left: Methyl TROSY NMR spectrum of the C-terminal domain of Dcs1p in the absence (red) and presence (blue) of a 5-fold excess of m^{7}GpppG. The same residues experience chemical shift changes in the isolated domain and in the full-length protein (Figure 2c), indicating that the m^{7}GpppG substrate interacts in the same manner with the isolated C-terminal domain and the full length protein.

Bottom right: To judge how the substrate structurally interacts with the open binding pocket of the yeast protein, we modeled the m^{7}GpppG substrate into the empty open binding pocket of our yeast structure. The model is based the human Dcs1p enzyme (1ST0) that contains substrate in the open and closed binding sites. The N-terminal domain is colored in blue, the hinge region in red and the C-

Nature Chemical Biology: doi:10.1038/nchembio.1866

terminal domain is shown in green, the ligand m^7GpppG in yellow. In our model, the distance between the Cδ1 methyl group of I42 (pink sphere) and the substrate (dashed line) is between 12 and 17 Å (depending on how the substrate is exactly modeled). The shift of I42 in our NMR titration experiments can thus not be due to direct contacts between the substrate and the methyl group in the N-terminal domain. Residue K126 that alters the dynamics of the protein is located in the hinge region, in-between the bound second substrate and I42 in the N-terminal domain.

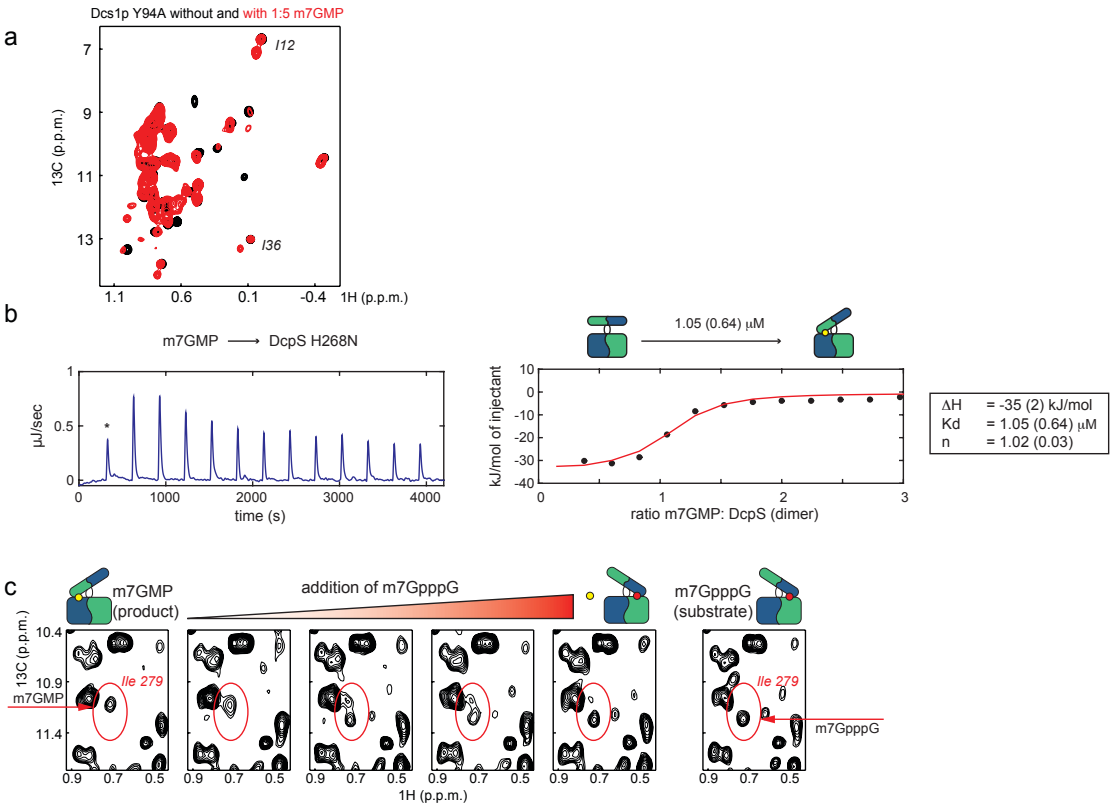
Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 4 | Asymmetry of the enzyme is required for activity. (a) The Y94A Dcs1p enzyme interacts with m^{7}GpppG (left spectrum). The WT Dcs1p enzyme interacts with (non-methylated) GpppG (two right spectra: Ile, respectively Met region). In both cases, the enzyme fails to adopt an asymmetric conformation. This can be clearly seen from e.g. residues I12 and I36, that don't experience peak splitting as we observed for the WT protein in the presence of m^{7}GpppG (Figure 2B). It should be noted that the affinity of the protein for the ligand is unaffected by the mutations as the extent of the chemical shift changes observed here is comparable to the chemical shift changes that are observed for the open binding site in WT protein with m^{7}GpppG (compare e.g. M153 in Figure S3a).

(b) ^{31}P spectra of the m^{7}GpppG or GpppG in the presence of the decapping enzyme. Upon hydrolysis, the phosphate spectrum of the ligand changes. m^{7}GpppG is not hydrolysed by the Y94A enzyme (left) and GpppG is not hydrolysed by WT Dcs1p (middle). As a reference the spectra of m^{7}GpppG in the presence of WT enzyme are shown on the right, where the substrate is hydrolysed completely within 14 hours.

Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 5 | Substrate competes the product out of the enzyme.

(a) DcpS without and with the product m^{7}GMP. In the presence of m^{7}GMP the enzyme adopt an assymmetric conformtion as can be judge from the splitting of the resonances of I12 and I36. Note that I12 appears at lower carbon chemical shifts compared the other spectra shown in this paper due to a difference in the aliasing of the spectrum.

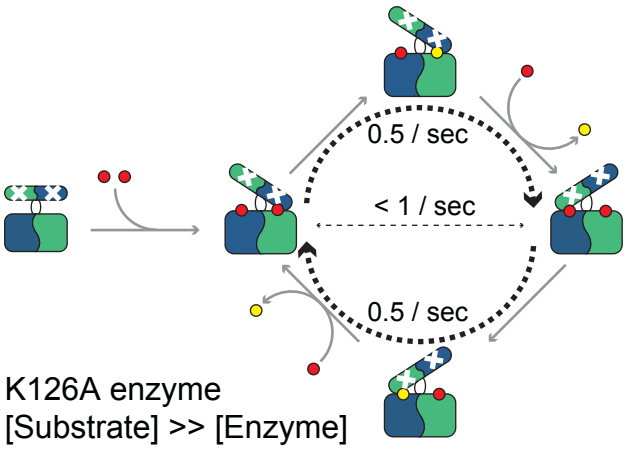
(b) DcpS binds the m^{7}GMP product with an affinity of 1.05 ± 0.64 µM in the first binding event. A second binding event is not observed, indicating that the product has no affinity for the open binding site. Using NMR spectroscopy and ITC we have not been able to detect any interaction of the GDP product with either the first or the second binding site (data not shown), underscoring the importance that the 7-methyl group plays in the recognition process between the nucleotides and the enzyme (see Figure S4).

(c) The substrate is able to compete the product out of the closed binding site. The NMR spectra of the WT protein bound to m^{7}GMP and bound to m^{7}GpppG are very similar and in order to be able to clearly distinguish if substrate or product is bound in the closed binding pocket, we introduced a reporter isoleucine into

Nature Chemical Biology: doi:10.1038/nchembio.1866

the enzyme (Leucine 279 to Isoleucine). The enzyme containing this reporter isoleucine was converted into an asymmetric conformation by the addition of the product m^7GMP. Subsequently, we added m^7GpppG substrate in a stepwise manner. During the addition of the substrate the reporter isoleucine shifts from the m^7GMP bound position to the m^7GpppG bound position. This indicates that the asymmetric product bound form of the enzyme is transferred into an asymmetric substrate bound form. In other words, interaction of substrate in the second binding site results in an opening of the first, product filled binding site. As a result the product can rapidly leave the enzyme, as this exhibits no detectable interaction with the open binding pocket.

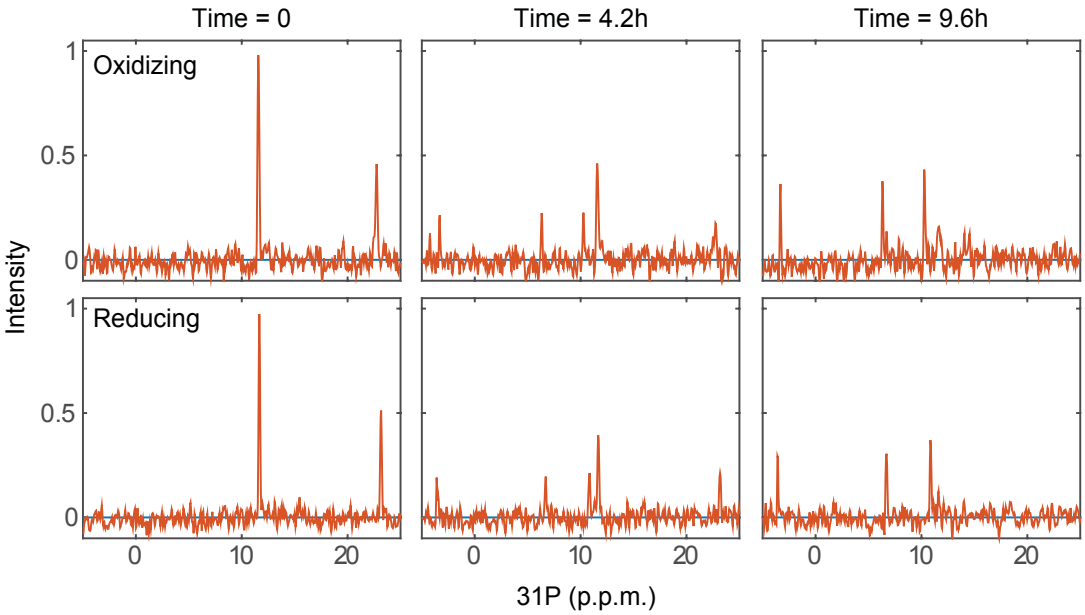
Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 6 | Cartoon representation of the K126A mutant enzyme.

The K126A mutation decouples the occupation of the second binding site from domain flipping motions. Unproductive motions are reduced or absent (horizontal dashed line) and catalytic turnover increases (curved dashed arrows). The K126A mutation is indicated with a cross.

Nature Chemical Biology: doi:10.1038/nchembio.1866



Supplementary Figure 7 | Activity assays under reducing and oxidizing condition.

The activity of Dcs1p is not influenced by the formation of a disulphate bond in the N-terminal domain (that we observe in the crystal structure). Shown are three different time points from a degradation series in the absence (top) or presence (bottom) of 1 mM DTT (in the bottom spectra 1 mM DTT has been present during the complete purification). The spectra are identical within the noise, proving that the activity is independent of the formed disulphate bond. In both spectra 200 nM WT Dcs1p was mixed with 0.5 mM m^7GpppG substrate at 20°C.

Nature Chemical Biology: doi:10.1038/nchembio.1866

Supplementary Table 1 | Data collection and refinement statistics

Data collection	
Space group	19
Cell dimensions	
a, *b*, *c* (Å)	87.99, 104.52, 189.96
a, b, g (°)	90.0, 90.0, 90.0
Resolution (Å)	20 (2.25)
R_{sym} or R_{merge}	6.7 (71.6)
I / sI	12.04 (2.35)
Completeness (%)	99.5 (98.9)
Redundancy	3.3 (3.3)
Refinement	
Resolution (Å)	2.25
No. reflections	83812
R_{work} / R_{free}	0.219/ 0.255
No. atoms	
Protein	10338
Ligand/ion	98
Water	129
B-factors	
Protein	54.63
Ligand/ion	53.01
Water	43.02
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.36

Values in parentheses are for highest-resolution shell.

Nature Chemical Biology: doi:10.1038/nchembio.1866

Supplementary References

1. Chen, N., Walsh, M.A., Liu, Y., Parker, R. & Song, H. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J Mol Biol* **347**, 707-18 (2005).
2. Gu, M. et al. Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol Cell* **14**, 67-80 (2004).

Nature Chemical Biology: doi:10.1038/nchembio.1866

A general method for rapid and cost-efficient large-scale production of 5′ capped RNA

ANNA-LISA FUCHS,[1] ANCILLA NEU,[1] and REMCO SPRANGERS
Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

ABSTRACT

The eukaryotic mRNA 5′ cap structure is indispensible for pre-mRNA processing, mRNA export, translation initiation, and mRNA stability. Despite this importance, structural and biophysical studies that involve capped RNA are challenging and rare due to the lack of a general method to prepare mRNA in sufficient quantities. Here, we show that the vaccinia capping enzyme can be used to produce capped RNA in the amounts that are required for large-scale structural studies. We have therefore designed an efficient expression and purification protocol for the vaccinia capping enzyme. Using this approach, the reaction scale can be increased in a cost-efficient manner, where the yields of the capped RNA solely depend on the amount of available uncapped RNA target. Using a large number of RNA substrates, we show that the efficiency of the capping reaction is largely independent of the sequence, length, and secondary structure of the RNA, which makes our approach generally applicable. We demonstrate that the capped RNA can be directly used for quantitative biophysical studies, including fluorescence anisotropy and high-resolution NMR spectroscopy. In combination with ^{13}C-methyl-labeled *S*-adenosyl methionine, the methyl groups in the RNA can be labeled for methyl TROSY NMR spectroscopy. Finally, we show that our approach can produce both cap-0 and cap-1 RNA in high amounts. In summary, we here introduce a general and straightforward method that opens new means for structural and functional studies of proteins and enzymes in complex with capped RNA.

Keywords: mRNA capping; 5′ cap; mRNA decapping; vaccinia virus capping enzyme; 7-methylguanylate cap structure

INTRODUCTION

The 5′ guanine-*N*7-methyl cap structure is a central architectural feature of eukaryotic mRNA (Reddy et al. 1974; Furuichi and Miura 1975; Shatkin 1976). The presence of the cap structure is essential for almost all subsequent steps in the life cycle of an mRNA, including pre-mRNA splicing (Izaurralde et al. 1994; Fresco and Buratowski 1996; Schwer and Shuman 1996), mRNA export (Hamm and Mattaj 1990), and the initiation of translation (Both et al. 1975; Filipowicz 1978). In addition, the mRNA cap structure stabilizes mRNA by protecting it against 5′-3′ exonucleolytic degradation (Furuichi et al. 1977), and removal of the cap structure irreversibly targets the transcript for decay (Parker and Song 2004). In addition, acquisition of a cap structure is a common mechanism by which viruses ensure efficient translation of their RNA and evasion of the host immune system (Daffis et al. 2010; Decroly et al. 2012).

In the basic eukaryotic cap structure (m^7G or cap-0), a 5′ *N*7-methylguanosine is attached to the mRNA body by an unusual 5′-5′ triphosphate linkage (Fig. 1A). The capping enzymes are recruited to the RNA polymerase II through the carboxy-terminal domain (CTD) (Cho et al. 1997) and act cotranscriptionally when the transcript has reached a length of about 20 to 30 nucleotides (nt) (Salditt-Georgieff et al. 1980). Three successive enzymatic steps are required for formation of cap-0 and involve an RNA triphosphatase, an RNA guanylyltransferase, and an RNA guanine-*N*7-methyltransferase activity. The RNA triphosphatase removes the 5′ γ-phosphate from the nascent pre-mRNA to form an mRNA that contains a 5′ diphosphate. Subsequently the RNA guanylyltransferase transfers GMP from a GTP donor to the 5′ end of the RNA, which results in the formation of the unmethylated cap structure. Finally, the guanine-*N*7-methyltransferase uses *S*-adenosyl methionine (SAM) as a methyl donor to methylate the *N*7 position of the guanine base (Fig. 1A). Although the order of these catalytic steps is conserved in eukaryotes, the enzymes that perform these functions differ between species. In yeast, all three activities reside in separate proteins (Gu et al. 2010), whereas in higher eukaryotes several activities reside on a single protein chain (Takagi et al. 1997).

In yeast, the most predominant RNA cap is the cap-0 structure (Sripati et al. 1976), whereas RNA in higher eukaryotes

[1]These authors contibuted equally to this work.
Corresponding author: remco.sprangers@tuebingen.mpg.de
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.056614.116. Freely available online through the *RNA* Open Access option.

can be additionally methylated at the nucleotides that follow this cap (Adams and Cory 1975; Furuichi and Shatkin 2000). The most common modification is methylation of the 2′-O position of the ribose of the first nucleotide that follows the cap structure (cap-1), which takes place in the nucleus. The mRNA can be further methylated in the cytoplasm at the 2′-O position of the second ribose to yield a cap-2 structure. In different eukaryotes the levels of cap-0, cap-1, and cap-2 in mRNA vary. The amount of methylation has increased during evolution (Banerjee 1980) and is implicated in immunogenicity (Daffis et al. 2010; Devarkar et al. 2016).

Viruses that replicate in the eukaryotic cytoplasm rely on their own genes that code for the enzymes that provide a cap structure for the viral transcript. To that end, viruses either carry genes that code for a viral RNA capping machinery or that code for a cap-snatching machinery that transfers a cap from cellular mRNA (Reguera et al. 2016). Hence, cap structures have been observed in the majority of such viral RNAs (Shatkin 1976), although with different levels of methylation. As an example, cap-0 structures have been found on mRNA from the tobacco mosaic virus (Zimmern 1975) and cap-1 structures on mRNA from the vaccinia virus (Wei and Moss 1975). The capping enzyme of the vaccinia virus is a complex of the two viral proteins D1 and D12 (Fig. 1B; Ensinger et al. 1975; Martin et al. 1975; Shuman et al. 1980). The three activities are combined in the 97-kDa D1 protein, where the RNA 5′-triphosphatase and guanylyltransferase activities are located in the N-terminal half (Myette and Niles 1996) and the methyltransferase activity localizes to the C-terminal half of the protein (Mao and Shuman 1994). The 31-kDa D12 protein has no catalytic activity, but stimulates the methyltransferase activity of the D1 protein (Mao and Shuman 1994; De la Pena et al. 2007). A structure of the complete D1:D12 vaccinia virus complex has recently been determined and provides insights into the modulation of the catalytic activity by extensive interdomain contacts (Kyrieleis et al. 2014). The vaccinia virus also contains a 2′-*O*-methyltransferase enzyme (VP39) that converts the cap-0 structure into a cap-1 structure (Barbosa and Moss 1978; Schnierle et al. 1992; Hodel et al. 1998).

Despite the essential role the cap structure plays, detailed knowledge about the interactions of the 5′ end of mRNA with adapter proteins and decapping enzymes is limited. This is, to a large degree, due to challenges related to the in vitro production of pure and homogeneous capped RNA in large quantities (>500 nmol, or 15 mg for an RNA body of 100 nt) that are required for structural studies. Currently, multiple strategies have been proposed to produce capped RNA using enzymatic or chemical methods or a combination of both.

The vaccinia virus capping system has been used to enzymatically prepare limited amounts of capped RNA based on an in vitro transcribed mRNA body (Paterson and Rosenberg 1979; Green et al. 1983; Ray et al. 2006). These capped RNAs are mainly used for small-scale applications including in vitro translation experiments and assays that rely on detection of fluorescent (Gunawardana et al. 2015) or radiolabeled mRNA. Initially, enzymatic capping using the vaccinia virus capping enzyme was reported to be inefficient (Paterson and Rosenberg 1979; Contreras et al. 1982; Pelletier and Sonenberg 1985). Nowadays, the functional enzyme can be obtained commercially; however, the amount of enzyme required to produce high milligram (μmol) amounts of RNA prevents general applicability, and some of the protocols are standardized only for long (>60–100 nt) RNAs. In an alternative approach to obtain capped RNA, a cap analog (m^7G-ppp-X) is added in high concentrations directly to the in vitro transcription reaction, which results in the incorporation of the cap structure at the 5′ end of the RNA (Contreras et al. 1982; Konarska et al. 1984; Pelletier and Sonenberg 1985; Nielsen and Shapiro 1986). However, due to the almost symmetric nature of the cap analog, the cap is incorporated in the reverse order in the RNA in at least a third of the transcripts (Pasquinelli et al. 1995). To prevent this reverse cap orientation, "anti-reverse" cap analogs can be used in the transcription reaction, where the 3′ OH group of the m^7G nucleotide is modified (Grudzien-Nogalska et al. 2007), which results in the incorporation of a non-natural cap-like structure in the RNA. For very short RNA sequences (<9 nt), the incorporation of a reverse cap can also be prevented by the use of gene 4 primase (Matsuo et al. 2000) or a fragment thereof (Peyrane et al. 2007) that incorporates cap analogs solely in the correct orientation.

Chemically synthesized RNA is not directly suitable for enzymatic capping reactions, as the 5′ end of the RNA is not phosphorylated and thus not a substrate for guanylyltransferase enzymes. Chemically synthesized RNAs can, however, be di- or triphosphorylated at the 5′ end, although often in small scale and with yields in the range of 50% (Brownlee et al. 1995). Recently, optimizations in the solid-phase synthesis of 5′ triphosphate RNA allowed for the production of longer RNA bodies that are up to 130 nt in length (Nagata et al. 2010; Goldeck et al. 2014). The full chemical synthesis of capped RNA sequences is very challenging and inefficient due to the instability of *N*7-methylguanosine under acidic and basic conditions. This problem has been bypassed by synthesizing nonmethylated capped short RNAs that were *N*7-methylated in a subsequent enzymatic step (Thillier et al. 2012). Full chemical synthesis of small (1–3 bases) RNAs is feasible, even with very complex methylation patterns (Lewdorowicz et al. 2007); however, longer capped RNAs seem too challenging for chemical synthesis. Finally, it has been reported that 7-methylguanosine 5′-diphosphate imidazolide can be chemically coupled with 5′-monophosphorylated oligoribonucleotides to form capped RNAs, albeit with yields in the order of 40% (Sawai et al. 1999).

In summary, established in vitro methods for the production of capped RNA have severe limitations, including low yields and the limitation to short oligonucleotides. This is especially reflected in the low number of structures of proteins

in complex with capped RNA (>1 nt) that have been determined to date. Currently, these structures include the human 2′-*O*-ribose methyltransferase CMTr1 with a capped 4-mer that was produced by chemical coupling (Smietanski et al. 2014), the 2′-*O*-ribose methyltransferase of vaccinia virus with a capped 6-mer that was produced by in vitro transcription in the presence of a cap analog (Hodel et al. 1998), the 2′-*O*-ribose methyltransferase in the NS5 protein from dengue virus with an 8-mer cap-0 viral RNA that was produced using a cap analog (Zhao et al. 2015), the dengue virus methyltransferase in complex with a 5′-capped RNA 8-mer that was chemically synthesized (Yap et al. 2010), and the innate immune receptor RIG-I that contains a chemically synthesized 24-nt-long capped hairpin RNA (Devarkar et al. 2016). The high-resolution structural data available is thus confined to a small subset of the numerous enzymes and proteins that directly interact with the mRNA cap structure. This limits our understanding of how these cap-binding factors are able to modulate mRNA metabolism and prompted us to establish a general, easy, rapid, and cost-efficient method for large-scale production of homogeneously capped RNA.

Here, we designed an optimized expression system for the vaccinia virus capping complex that allows for the straightforward purification of large quantities of the highly active enzyme. We then show that the enzyme we produce is able to cap a very wide variety of RNA species, where capping efficiency is virtually independent of the length, sequence, and structure of the RNA body. We establish that milligram amounts of capped RNA can be obtained in biological

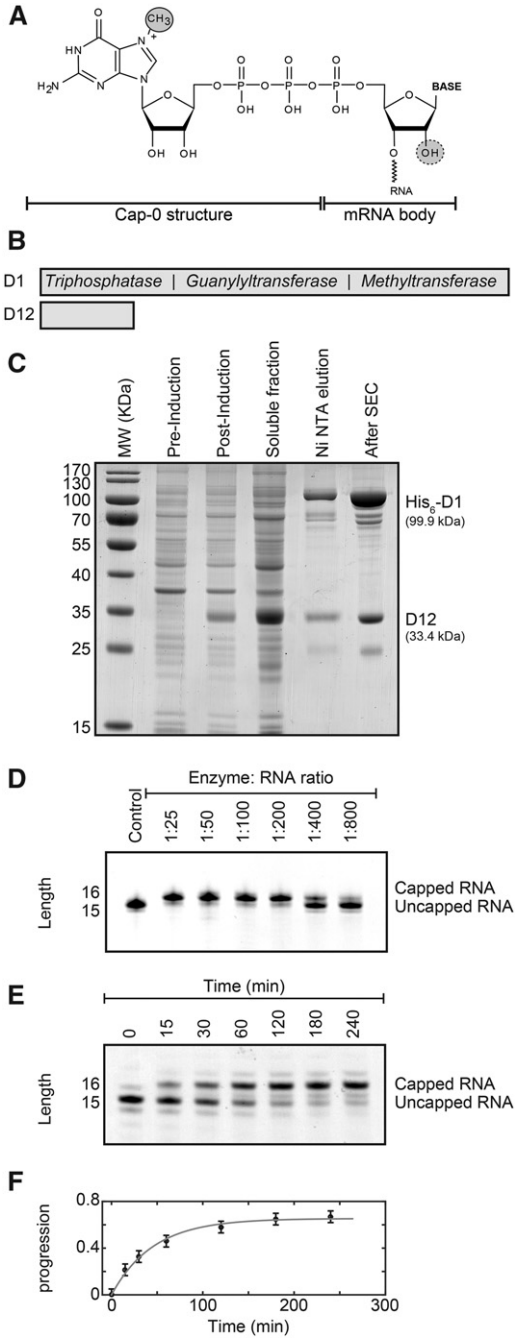


FIGURE 1. Purification and activity of the vaccinia virus capping enzyme. (*A*) Structure of the mRNA cap-0. The *N*7-methyl group in the cap structure is circled. The ribose 2′-OH group that is methylated to yield the cap-1 structure is indicated with a dashed circle. (*B*) Schematic representation of the vaccinia virus capping enzyme complex that contains the D1 protein with the catalytically active sites and the D12 protein that stabilizes the D1 protein. The D1 protein contains three activities: First, the triphosphatase activity hydrolyzes the RNA 5′ triphosphate into a 5′ diphosphate and inorganic phosphate. Second, the guanylyltransferase activity transfers GMP from GTP onto the 5′ end of the diphosphate RNA. In this step, the 5′-5′ triphosphate bond is formed and pyrophosphate is released. Finally, the methyltransferase activity uses SAM (*S*-adenosyl methionine) as a methyl donor to methylate the guanine *N*7 position. (See also Supplemental Fig. 1). (*C*) Overexpression and purification of the vaccinia virus capping enzyme. The designed plasmid contains the DNA of the D1 and D12 proteins that are codon optimized for overexpression in *Escherichia coli* (Supplemental Fig. 2). Shown are the *E. coli* cells before induction of protein expression (pre-induction) and after overexpression of the enzyme complex for 12 h (post-induction), the soluble fraction and the protein after Ni-NTA affinity purification (Ni NTA elution), and the final enzyme complex after size exclusion chromatography (After SEC). Note that the expression of the stabilizing D12 protein is higher than that of the D1 enzyme. The excess of the D12 protein is removed during Ni-NTA purification as only the D1 protein contains an affinity tag. The yield of the purification is ~12 mg of pure enzyme complex from 5 L of *E. coli* culture. (*D*) The purified complex possesses a high mRNA capping activity. Shown are capping reactions with different enzyme:RNA ratios after 1 h incubation time. The capping reaction can be easily followed using Urea-PAGE analysis as the capping reaction adds one base to the RNA substrate. The optimal ratio varies slightly depending on the substrate RNA and should be determined using small-scale capping reactions. (*E*) To obtain insights into the kinetics of the capping reaction, an uncapped RNA was incubated with a low amount of capping enzyme (1:750). The RNA is capped to a larger degree after longer (>1 h) incubation times. (*F*) Quantification of the capping reaction shown in *E*. The solid gray line is a fit of the progression of the reaction [where the progression is defined as capped/(capped + uncapped) RNA] to the experimental data. From the initial slope of the progression curve we extract that the enzyme can cap 16 RNA substrates per minute. The rate of the capping reaction drops over time due to the loss of activity of the enzyme after incubation times that are longer than 1 h.

laboratories in very short times and that the capped RNAs are amenable for a wide range of biophysical and structural studies.

RESULTS

Cloning and purification of the vaccinia virus capping complex

To be able to overexpress the D1:D12 vaccinia virus capping complex (Fig. 1B, bottom) in *E. coli*, we designed an expression vector (Supplemental Fig. S1, see Materials and Methods). We wished to have a system at hand that ensures high levels of overexpression and that allows for a straightforward and reproducible purification of the complex. To that end, we combined coexpression of both proteins for optimal stability of the complex and with a single purification tag on the D1 enzyme for convenient purification (De la Pena et al. 2007; Kyrieleis et al. 2014) with codon optimization of the *D1* and *D12* genes for *E. coli* expression. Using this expression system, we could overexpress the viral capping complex and isolate it using two standard purification steps (Ni affinity chromatography followed by size exclusion chromatography; Fig. 1C; see Materials and Methods). The purification of the enzyme takes 1.5 d and typically yields 12 mg pure enzyme from 5 L of *E. coli* culture, which is sufficient to prepare 5–10 µmol capped RNA.

RNA body preparation

The RNA body that we use for the capping reaction was prepared using established in vitro transcription using the T7 polymerase. In combination with the standard T7 φ6.5 promoter, this enzyme prefers a guanine as the first base, whereas in combination with the T7 φ2.5 promoter, transcription starts with an adenine (Coleman et al. 2004; Ray et al. 2006). As templates for the in vitro transcription reaction, we used plasmid DNA that contains an HDV-ribozyme at the 3′ end and DNA primers in a nonmethylated or methylated form (Helmling et al. 2015).

Capping reaction

To test the activity of the vaccinia virus capping complex on the RNA body, we performed capping reactions using different enzyme:RNA ratios. The capping reaction was then analyzed using Urea-PAGE, where the RNA was visualized using methylene blue (Fig. 1D). During the capping reaction a methylated guanidine is added to the RNA body, which results in a shift of the mobility of the transcript. Interestingly, we found that we could cap an RNA body with close to 100% efficiency within 1 h in all cases. The efficiency depends, however, on the enzyme:RNA ratio and in all subsequent experiments we used small-scale tests to optimize this ratio before large-scale capping reactions were performed. We found that optimal enzyme:substrate ratios varied between 1:10 and 1:500, depending on the RNA body. Longer incubation times can increase the yield of the capping reaction (Fig. 1E); however, the enzyme loses activity over the course of hours (Fig. 1F). Complete capping is thus most reliably assured in reaction times that do not exceed 1–2 h.

After the capping reaction, the capped RNA was separated from the enzyme using a heat precipitation step, which quantitatively removed all protein without loss of the target RNA (Supplemental Fig. S2). Finally, the RNA could be separated from nucleotides and salts by isopropanol precipitation and desalting steps. The yield of the final capped RNA product depends mainly on the efficiency of the transcription reaction, as the RNA body was capped close to 100% in all cases. Routinely, we obtained 500 nmol (15 mg for a 100-mer RNA body) from a 10-mL transcription reaction. From the transcribed RNA to the pure capped product the total experimental time is ∼3.5 h (1 h capping reaction, 2 h precipitation, and 30 min desalting).

To confirm that the capping reaction indeed added a cap-0 structure to the RNA body, we performed native mass spectrometry on an RNA body of 31 nt before and after the capping reaction (Supplemental Fig. S3). As starting material, we used a mixture of 5′ mono -, di-, and triphosphorylated RNA and the identity of all species could be unambiguously identified in the native mass spectrum. After the capping reaction, we detected a mixture of 5′ monophosphorylated RNA and of the cap-0 RNA. These data confirm that the RNA body requires a di- or triphosphate at the 5′ end in order to be a substrate for the vaccinia capping enzyme and, more importantly, this shows that the capping enzyme properly adds a cap-0 structure to the RNA.

Dependence of the capping on the RNA body

To test the dependence of the efficiency of the vaccinia capping enzyme on length, sequence, and secondary structure of the RNA body, we performed a large set of capping reactions, where we varied single parameters. To test how the substrate length modulates capping efficiency, we performed capping reactions using substrates with RNA bodies between 2 and 100 nt. If the RNA body is over 40 nt, it is not possible to directly follow the capping reaction using a standard Urea-PAGE analysis, as a difference of one base cannot be clearly resolved anymore. To follow the capping reaction we therefore designed a substrate that contains a unique RNase A cleavage site that is located 15 bases downstream from the 5′ end of the RNA body (Fig. 2A). In that manner, we can perform the capping reaction on the full-length intact RNA and observe the progress of the capping reaction after cleaving with RNase A. Importantly, from this set of experiments we can conclude that RNAs between 40 and 100 nt are fully and efficiently capped. To test whether very short RNAs are also a substrate for the vaccinia virus capping enzyme, we used RNAs between 2 and 5 nt. Interestingly, we observe that the capping enzyme efficiently and fully caps

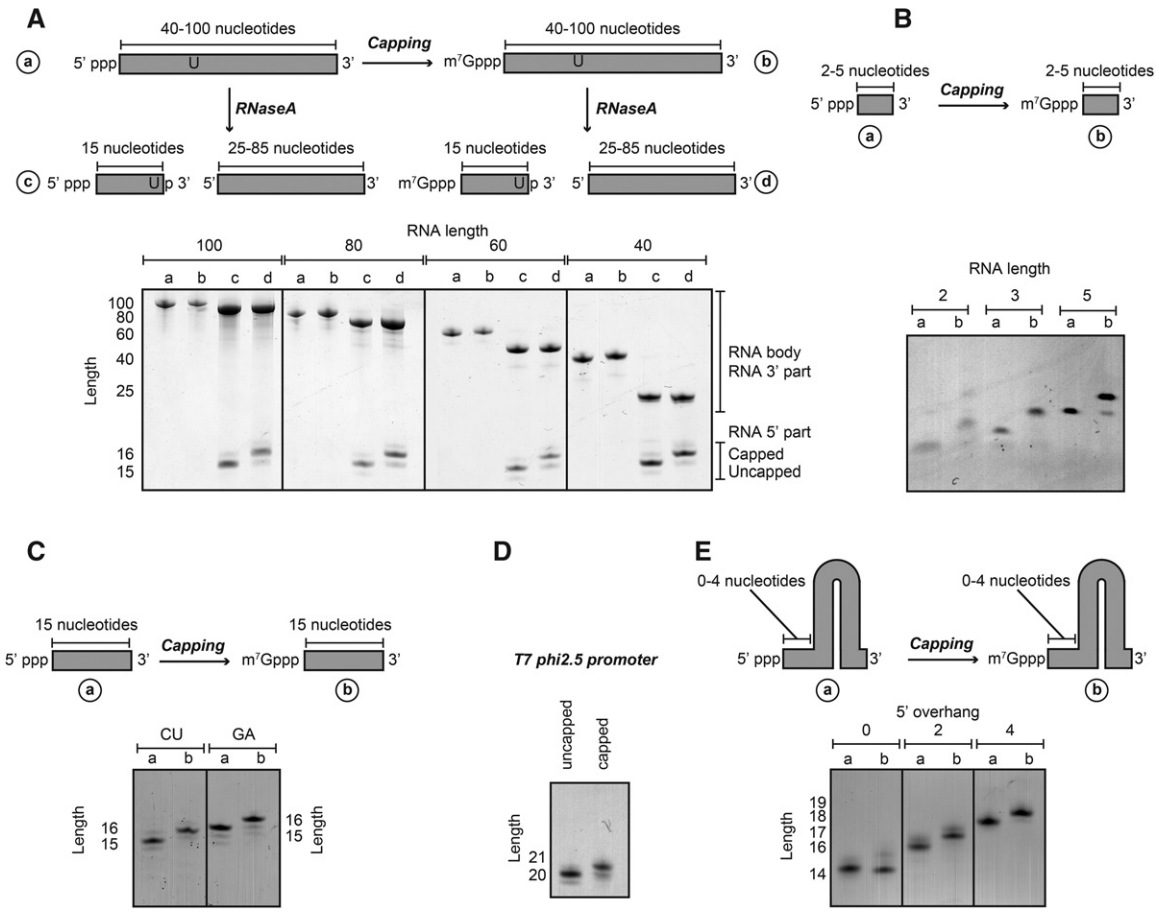


FIGURE 2. The RNA capping reaction is robust and independent of substrate RNA length, sequence, or secondary structure. (*A*) Capping reactions using RNA substrates of different lengths (between 100 and 40 nt). To visualize the capping efficiency a single U was introduced at position 15. RNAse A digestion of the RNA (before and after capping) released the 5′ end of the substrate that only contained 15 nt. The mass shift due to the capping reaction can be readily observed using Urea-PAGE analysis. The capping efficiency of all RNAs was close to 100%. (*B*) Shorter RNA sequences (2–5 bases) are also efficiently capped using the purified capping complex. (*C*) The capping reaction is independent of sequence because both GA and CU rich sequences can be capped with efficiencies close to 100%. Note that the GA and CU RNA migrate slightly differently due to the base composition. (*D*) The capping reaction is efficient when the first base of the RNA body is an adenine. (*E*) RNA secondary structure only interferes with the capping reaction when there is no 5′ overhang. A stable RNA hairpin that contains no 5′ overhang is not capped, whereas a stable hairpin with a 5′ overhang of 2 or 4 bases is efficiently capped.

even dinucleotide sequences (Fig. 2B). Taken together, the vaccinia capping enzyme is able to efficiently cap RNA fully independent of substrate length.

To evaluate whether the sequence of the RNA plays a role in the capping process, we used RNA bodies that are either GA rich or CU rich. Both substrates are fully and efficiently capped, indicating that the primary sequence of the RNA body does not influence the capping efficiency (Fig. 2C). In addition, we established that the RNA capping reaction also functions when the first base of the transcript is an adenine, as transcripts produced with the T7 φ2.5 promoter are also fully capped (Fig. 2D).

Finally, to test whether RNA secondary structure has an influence on the capping efficiency, we used a very stable GC hairpin structure with either no, two, or four bases overhang at the 5′ end. Of these RNA bodies only the RNA hairpin that contains a fully base-paired 5′ end is not a substrate for the capping enzyme (Fig. 2E). The other RNA bodies are efficiently and fully capped, indicating that stable secondary structure elements do not interfere with the capping reaction unless the 5′ end is part of a stable double strand.

In summary, we here show that the vaccinia capping enzyme is able to cap RNA in an efficient manner that is practically independent of the length, sequence, and structure of the RNA body.

Preparative preparation of short capped RNAs

For a number of structural and functional studies it might be advantageous to prepare capped RNA with a minimal possible length. Above, we showed that RNA as short as 2 nt can be efficiently capped. In vitro transcription reactions of very short RNAs (<10 nt) are, however, often very inefficient. We therefore established an alternative approach to prepare

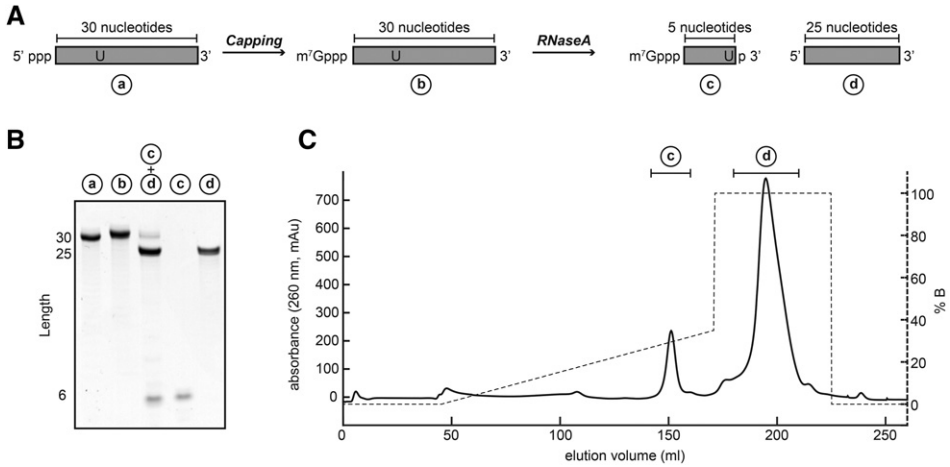


FIGURE 3. Large-scale preparation of short, capped RNAs. (*A*) A longer RNA (30 nt) is prepared using in vitro transcription and capped as described above. The RNA contains a unique cleavage site for an RNase, and we here used RNase A to cleave the RNA after a single uracil. (*B*) Urea-PAGE analysis of the capping reaction and RNase digestion (lanes *1–3*) and of the subsequent purification of the digested RNA (lanes *4*,*5*). (*C*) The two RNA fragments are separated using a Dionex DNAPac column. The drawn line corresponds to the absorbance at 260 nm, whereas the dashed line corresponds to the salt gradient that is used to elute the RNA.

large quantities of short and capped RNA based on the vaccinia capping and the T7 polymerase enzymes. In brief, we demonstrate that it is convenient to transcribe and cap a longer RNA body, followed by the preparative digestion of this product using a specific RNase (Fig. 3A). Indeed, we show that a 30-nt-long RNA is efficiently capped and quantitatively cleaved using RNase A (Fig. 3B). In an alternative approach the longer RNA can be digested using site-specific RNase H cleavage (Duss et al. 2010). The cleaved fragments can be conveniently separated using anion-exchange chromatography (Fig. 3C), which yield highly pure and capped short RNA fragments.

Separation of capped and uncapped RNA

In all our experiments we observed a capping efficiency of close to 100%. In case the capping reaction turns out less efficient, it might, however, be necessary to separate uncapped substrate from capped product. To that end we made use of the eIF4E binding protein that specifically interacts with capped RNA and not with RNA that has a 5′ end devoid of m^7 methylated guanine. We incubated a mixture of 50% capped and 50% uncapped RNA with purified His_6-tagged eIF4E protein and Ni-NTA resin (Fig. 4A). The uncapped RNA did not interact with the eIF4E protein on the resin and could thus be removed in a single wash step. The capped RNA, on the other hand, formed a tight complex with the eIF4E on the resin. This eIF4E:capped RNA complex was then eluted from the Ni-NTA resin using higher imidazole concentrations. Finally, the capped RNA could be separated from eIF4E protein using a phenol–chloroform extraction step. In summary, we show that it is possible and straightforward to quantitatively separate capped and uncapped RNA.

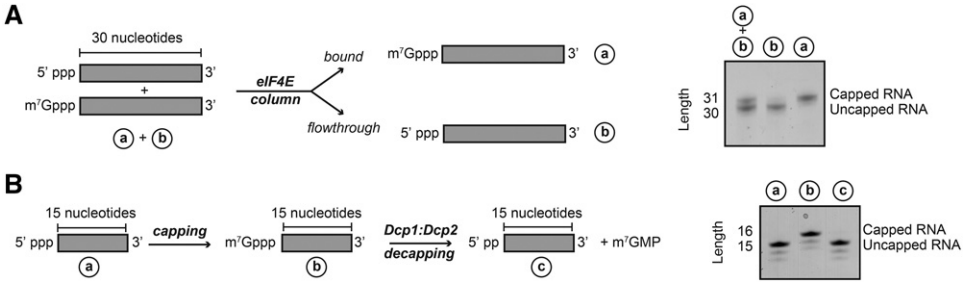


FIGURE 4. (*A*) In case the capping reaction is incomplete, the uncapped RNA can be separated from the capped RNA using an eIF4E pulldown. N-terminally His_6-tagged eIF4E was attached to Ni-NTA resin and a mixture of 50:50 capped and uncapped RNA was added (a + b). The uncapped RNA did not interact with the eIF4E and could be washed away. Capped RNA interacted tightly with the eIF4E protein and was thus retained on the beads. The protein:RNA complex was eluted using imidazole. The capped RNA was separated from the eIF4E protein using a phenol–chloroform extraction. (*B*) The capped RNA is a substrate for the Dcp1:Dcp2 decapping enzyme. An RNA body of 15 nucleotides was capped using the capping enzyme complex and subsequently decapped using the Dcp1:Dcp2 decapping complex.

Translation and decapping assay using the capped RNA

To demonstrate the functionality of the capped RNA, we prepared a capped transcript that codes for the GFP protein and used this in in vitro translation assays (Supplemental Fig. S4). As expected, the translation efficiency is significantly increased after the capping reaction, proving that the modification occurred correctly.

In addition, capped RNA is a substrate for a number of decapping enzymes, including DcpS and the Dcp1:Dcp2 complex (Parker and Song 2004). To confirm that the capped RNA that we prepared here is indeed a substrate for the decapping enzymes we incubated a capped RNA of 15 nt with the purified Dcp1:Dcp2 complex. Urea-PAGE analysis of the RNA shows that the cap structure is, as expected, fully removed (Fig. 4B). We hereby demonstrate the suitability of the capped RNA for activity assays that circumvent radioactive labeling.

Binding studies using the capped RNA

To demonstrate general applicability in biophysical studies, we investigated the complex between capped RNA and eIF4E. To that end, we used fluorescence anisotropy and high-resolution NMR methods as examples. In a first set of experiments, we quantified the interaction strength between the eIF4E protein and the capped RNA using fluorescence anisotropy experiments with a fluorescently labeled capped RNA of 31 nt. To probe the strength of the capped-RNA:eIF4E interaction, we added the *Drosophila melanogaster* (*Dm.*) protein in increasing amounts to the RNA and observed changes in the fluorescence anisotropy (Fig. 5A). From the binding curve, we extracted a kD for the RNA:protein interaction of 294 (42) nM. Previously, the affinity between the *Dm.* eIF4E and m^7GDP (an RNA cap analog that lacks the complete RNA body and that contains two instead of three phosphate groups) was determined to be ~700 nM, using ITC (Kinkelin et al. 2012). The slightly better affinity of the protein for capped RNA in comparison to that for m^7GDP suggests that there are a small number of contacts between the cap binding protein and the mRNA that involve parts of the RNA that are outside the methylated guanine and first two phosphates in the cap linker. This is in agreement with biophysical and structural studies on the murine eIF4E protein that shows contributions of the γ-phosphate in the eIF4E:cap interaction (Niedzwiecka et al. 2002). In summary, our and previous studies show that the interaction between the mRNA cap structure and eIF4E is mainly mediated through the cap structure and the 5′-5′ triphosphate linker and that there are no extensive contacts between the RNA body and the cap binding protein. The fact that the mRNA body does not interact with the eIF4E protein is also confirmed in structural studies with cap analogs, as there the first base of the RNA body is not oriented in a fixed position on the eIF4E structure (Fig. 5B).

NMR studies of the capped RNA

To show the potential of our RNA capping method for structural biology, we performed high-resolution NMR experiments. In those experiments, the interaction between the capped RNA and the eIF4E protein was probed using two complementary approaches. First, we prepared ^{15}N-labeled eIF4E and recorded NMR spectra in the absence and presence of a capped RNA with an RNA body of 5 nt (Fig. 6A). Upon addition of the RNA, we observed clear chemical shift perturbations (CSPs) in the eIF4E resonances that directly report on the specific interaction between the cap binding protein and the capped RNA. The large number of residues that undergo CSPs is in agreement with the enclosure of the cap structure into the core of the eIF4E protein (Matsuo et al. 1997; Kinkelin et al. 2012). In agreement with the limited

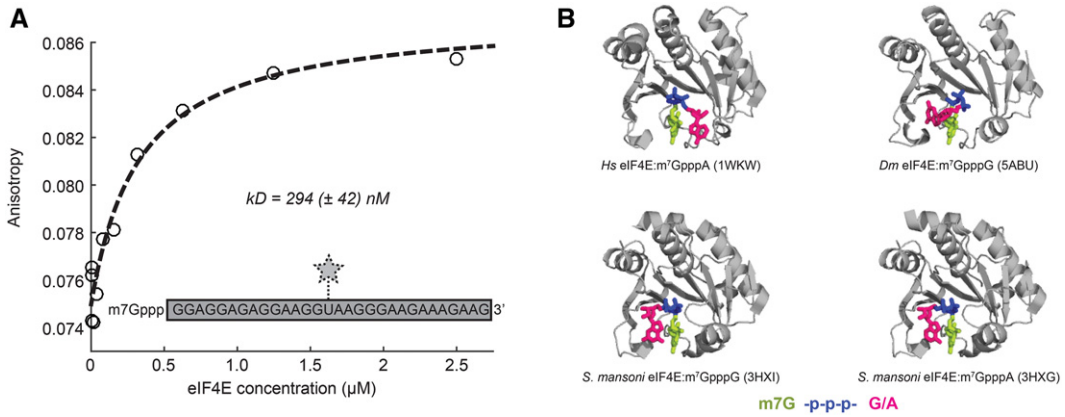


FIGURE 5. Affinity measurements using capped RNA. (*A*) Fluorescence anisotropy binding studies to probe the interaction strength between eIF4E and a capped RNA (kD = 294 ± 42 nM). (*B*) Known structures of eIF4E in complex with the m^7GpppG or m^7GpppA cap analog (Tomoo et al. 2005; Liu et al. 2009; Peter et al. 2015) show that the position of the first nucleotide of the RNA body (pink) is not well defined. This indicates that the interaction between eIF4e and the mRNA is mediated almost exclusively through the m^7GTP part of the cap-0 structure.

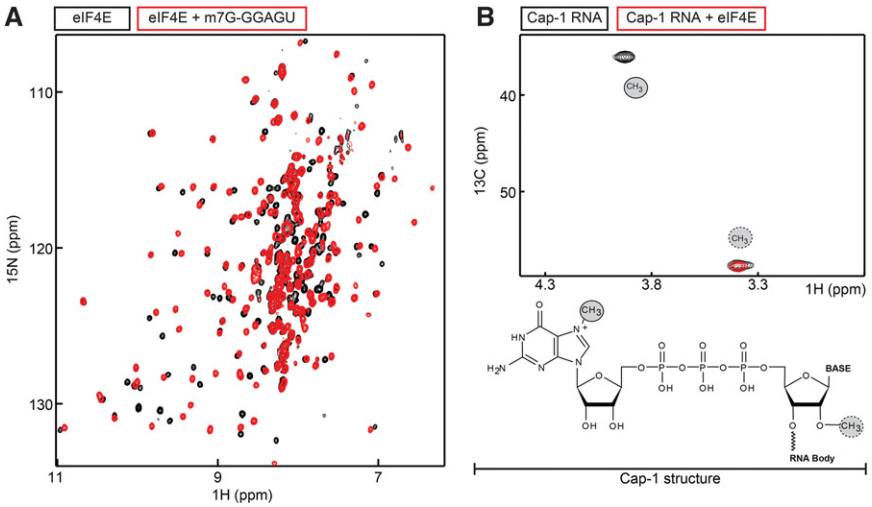


FIGURE 6. NMR studies using capped RNA. (*A*) 2D ^{1}H-^{15}N TROSY spectrum of the eIF4E cap binding protein in the absence (black) and presence (red) of a twofold excess of a 5-mer RNA that contains a 5′ cap-0 structure. (*B*) ^{1}H-^{13}C methyl TROSY spectrum of a 5-mer RNA that contains a m^{7}-[^{13}CH$_3$]/2′-O-[^{13}CH$_3$] cap-1 structure in the absence (black) and presence (red) of eIF4E.

role the RNA body plays in the cap recognition process, the NMR spectrum of eIF4E in complex with the capped RNA and that of eIF4E in complex with m^{7}-GDP are very similar (Supplemental Fig. S5).

In a second set of NMR experiments, we labeled the capped RNA with NMR active methyl groups. To achieve that, we used ^{13}C-labeled SAM (see Materials and Methods) during the RNA capping reaction. In a second step, we modified this ^{13}C-labeled cap-0 RNA further using the vaccinia virus VP39 enzyme that modifies a cap-0 RNA into a cap-1 RNA. This reaction was also performed in the presence of ^{13}C-labeled SAM. The resulting cap-1 RNA thus contained two NMR active methyl groups, at the *N*7 position of the guanine in the cap structure and at the 2′-*O* position of the first nucleotide of the RNA body. The resulting NMR spectrum clearly displays two resonances (Fig. 6B). The two resonances were assigned based on NMR spectra of a methyl labeled cap-0 RNA (data not shown) that only displayed a single resonance. Upon addition of eIF4E to the cap-1 RNA, the guanine *N*7-methyl group was broadened beyond detection, whereas the ribose 2′-*O* methyl group was unaffected (Fig. 6B). These data are also in agreement with the lack of significant interactions between the eIF4E protein and the first nucleotide of the RNA body. In addition, these experiments show that the cap-0 RNA can be entirely converted into a cap-1 RNA with the use of the VP39 enzyme from the vaccinia virus.

DISCUSSION

Functional and structural studies of proteins and enzymes that specifically recognize the 5′ end of mRNA are technically very challenging due to the lack of cost-effective and efficient ways to prepare larger quantities of capped RNA. Here, we show that the vaccinia capping enzyme is capable of capping RNA in quantities that by far exceed those that can be obtained using alternative (chemical or commercial enzymatic) methods. Importantly, we used a number of systematic tests (Figs. 2, 3) to show that the capping efficiency is almost independent of the length, sequence, and structure of the substrate RNA. The only RNA that we were not able to cap using the vaccinia enzyme is a substrate where the most 5′ nucleotide is part of a very stable hairpin. Due to the efficiency of the capping reaction, the obtained quantities of capped RNA are solely limited by the amounts of substrate. In our hands, we were able to obtain up to 500 nmol (15 mg for a 100-mer RNA body) of capped RNA from a 10-mL T7 transcription reaction. It is important to note that these quantities are more than sufficient for NMR applications and crystallization screening.

To have easy access to the vaccinia capping enzyme complex, we designed a highly optimized expression and purification setup (Fig. 1C; Supplemental Fig. S1). Importantly, the methods that are required for the enzyme preparation are generally available in biological laboratories, which makes the introduced system easily accessible. In our laboratory we produce large quantities of the enzyme and store this at −20° C without loss of activity for at least several months. Based on an appropriate mRNA body, capped RNA for biophysical studies (Figs. 5, 6) can then be produced within 3.5 h. The expression vector for the enzyme complex can be obtained from the authors upon request to facilitate future studies.

The capping method can be adapted for the production of various modified RNAs that serve specific purposes. First, we show that thio-uridine containing RNA can be capped efficiently. This enables, e.g., the addition of a fluorescent or spin label into the capped RNA, which we use here for fluorescence anisotropy measurements (Fig. 5). Secondly, NMR active nuclei can be introduced at specific positions in the capped RNA. We show that the use of ^{13}C-SAM results in the site-directed labeling of the methyl groups in the cap structure. Equivalently, the use of NMR active GTP would result in the specific labeling of the cap base. Thirdly, we show that the cap-0 structure that the vaccinia virus produces can be used as a substrate for enzymes that modify the cap-0 structure into a cap-1 structure (Fig. 6). Finally, the capped RNA can potentially be used as a building block to produce complex mRNAs in a strategy where capped RNA fragments and uncapped RNA fragments can be enzymatically ligated (Duss et al. 2010). In brief, we show that the modular nature of the capping reaction system can be exploited to produce various modified capped RNAs in large scales.

In summary, we here introduce an easy, fast, and cost-efficient protocol to prepare large quantities of capped RNA. We envision that this will spark future structural and biophysical studies of mRNA and its associated proteins and enzymes.

MATERIALS AND METHODS

Cloning of vaccinia capping enzyme

The sequences of the vaccinia capping enzyme subunits (UniProt identifiers P04298 and P04318, respectively) were optimized for expression in *E. coli* and synthesized by GenScript. Synthetic *D1* and *D12* sequences were inserted into the first and second multiple cloning sites of a modified pRSF-duet vector (Novagen), respectively (Supplemental Fig. S1). This vector carries an N-terminal TEV protease cleavable His-tag on D1 in the first multiple cloning site and an untagged D12 expressed from the second multiple cloning site, where both proteins have their own T7 promoter. In order to obtain a plasmid with reduced copy number, both multiple cloning sites encompassing both promoters and ORFs were then transferred to a pET-based (Novagen) vector backbone (Supplemental Fig. S1).

Purification of vaccinia capping enzyme

The vaccinia capping enzyme complex is coexpressed in *E. coli* BL21 Gold (DE3) pLysS cells. Cells were grown at 37°C to a cell density of OD600 = 0.6. Temperature was decreased to 20°C and cells were induced by addition of 0.2 mM IPTG. Twelve hours after induction, pelleted cells were resuspended in buffer A (50 mM Tris pH 8, 100 mM NaCl, 5 mM imidazole, 10% glycerol, 1 mM DTT) supplemented with 0.1% Triton X-100 and 1 mM PMSF and incubated on ice for 15 min with lysozyme. After lysis by sonification, the cell lysate was centrifuged at 50,000*g* for 30 min and the supernatant was applied to Ni-NTA resin (QIAGEN) equilibrated in buffer A. To remove unbound proteins, the column was washed with 10 column volumes buffer A. The protein complex was eluted from the resin by buffer A containing 200 mM imidazole. The elution fraction was dialyzed against buffer A without imidazole overnight. Size exclusion chromatography was performed in buffer A without imidazole using a Superdex S200 column. The pure capping enzyme complex was concentrated, aliquoted, and stored in buffer that contains 50% glycerol at −20°C.

In vitro transcription of RNA

RNA was produced by in vitro transcription (IVT) using DNA primers as templates for transcription by T7 RNA polymerase, which requires a duplex template in the promoter (Supplemental Table S1). Therefore, the reaction contained an equimolar amount of antisense strand primers bearing the T7 promoter sequence and sense strand primers encompassing 5′-target RNA sequence (in reverse complement)-CC–T7 promoter (in reverse compliment)-3′. Some of the antisense primers were methylated at the 2′-*O* position of the last 2 nt (Helmling et al. 2015) in order to reduce 3′ inhomogeneity. The use of methylated primers is indicated in the RNA construct list below.

IVT reactions were set up with a final concentration for both primers of 1 µM, 40 mM Tris pH 8, 5 mM DTT, 1 mM spermidine, 0.01% Triton X-100, 4 mM each of NTP in the RNA sequence, 0.2 µM T7 RNA polymerase (purified in the laboratory as described previously [Audin et al. 2013]), and 20–60 mM $MgCl_2$. The optimal Mg^{2+} concentration was determined in small-scale reactions. Transcription reaction mixtures were incubated at 37°C for at least 4 h. Precipitated pyrophosphate was cleared by addition of EDTA and the product RNA was precipitated as described below.

RNA precipitation

RNA was precipitated with 0.3 M NaOAc pH 5.2 and 0.7 volumes of isopropanol at −20°C for at least 2 h. The reaction was then centrifuged at 9000*g* and −5°C for 1 h and the pellet was washed with cold 75% ethanol and centrifuged again for 30 min. The resulting pellet was air dried and resuspended in the desired buffer for further applications. For short RNAs (2–10 nt), precipitation was carried out with 0.2 M NaCl and 3.5 volumes of EtOH (replacing the NaOAc and isopropanol) and a precipitation time of 12 h.

RNA purification using anion-exchange chromatography

RNA from IVT was purified using anion exchange chromatography on a 22 × 250 mm DNAPac PA100 column (Dionex) operated at 80°C. The column was run in buffers containing 20 mM Tris pH 8.0 and 5 M Urea (pH adjusted to 8.0 at room temperature). RNA of different lengths was eluted with a NaCl gradient. Salt gradients were designed according to the respective RNA lengths to be separated. Short RNAs (2–10 nt) typically elute between 50 and 200 mM NaCl while longer constructs (60–80 nt) elute at 300–400 mM NaCl. RNA containing fractions were analyzed using urea-PAGE, pooled and precipitated as described above. After precipitation, RNAs were desalted using a prepacked PD-10 Desalting column (GE Healthcare) and subsequently concentrated using a speedvac device.

Urea-PAGE

Analysis of RNA length was performed using Urea-PAGE. Polyacrylamide gels containing 8 M urea in TBE buffer were prepared with an acrylamide content suitable for the desired resolution (8%–21%). Gels were run with 10 mA until the sample had penetrated the gel completely and were further run with 15–20 mA. Gels were then rinsed with water and stained with a solution of 2g/L methylene blue in 0.4 M NaAc, pH 4.7 and destained in water.

RNA capping protocol

Capping reactions were performed in buffer containing 50 mM Tris, pH 8, 5 mM KCl, 1 mM $MgCl_2$, and 1 mM DTT with 20 µM RNA. 0.5 mM GTP, 0.1 mM SAM, and capping enzyme according to the previously determined substrate to enzyme ratio were added. The reaction was incubated at 37°C and prepared for analysis by Urea-PAGE by adding an equal volume of sample buffer containing 8 M Urea, 20 mM EDTA, 2 mM Tris pH 8, and Bromophenol blue and Xylene Cyanol.

As capping efficiency varies between different RNA constructs, the most efficient ratio of capping enzyme to RNA molecules was

determined in small-scale test reactions of 20 µL. Enzyme to RNA ratios typically varied between 1:10 and 1:100. Preparative scale reactions were scaled up accordingly. After capping, the reaction was heated to 70°C for 10 min and centrifuged for 10 min at 4500*g* to remove the capping enzyme from the reaction. Capped RNA was then precipitated and resuspended in an appropriate buffer and stored at −20°C.

For generation of cap-1 RNA, Vaccinia MTase (NEB) was added to the capping reaction according to the manufacturer's recommendations.

RNase A cleavage

RNase A is an endoribonuclease that cleaves specifically the phosphodiester bond between a pyrimidine nucleotide and the following nucleotide, resulting in a 3′ pyrimidine nucleoside phosphate. RNAse A was used for analytic and preparative specific cleavage of GA containing RNA sequences at a single uracil nucleotide. After capping of the RNA, it was cleaved with RNase A by adding 50 ng RNase A per nmol of RNA and incubating the reaction for 10 min at 37°C. Capping efficiency was then analyzed by Urea-PAGE. After preparative RNase A cleavage, the reaction was phenol–chloroform extracted once to remove all RNase A activity from the mixture before proceeding with RNA purification using anion-exchange chromatography.

Decapping with Dcp1:2

For decapping of capped RNA, a construct of the Dcp1:Dcp2 decapping complex from *S. pombe* was used, which encompasses the regulatory and catalytic domain of Dcp2 (1-266) and full-length Dcp1. The enzyme was purified as described previously (Fromm et al. 2012).

Capped RNA was incubated in capping buffer supplemented with an additional 5 mM $MgCl_2$ and Dcp1:Dcp2 in a molar ratio of enzyme to RNA of 1:20 for 30 min at 30°C. Sample buffer was added to stop the reaction and the samples were analyzed using Urea-PAGE.

Purification of eIF4E

The pEK-vH vector expressing D.m. eIF4E isoform c (UniProt identifier P48598-2) as an N-terminal His-tag fusion was a kind gift of Dr. Fulvia Bono (MPI Tübingen). His-eIF4E was expressed in *E. coli* DE3 BL21 cells carrying a plasmid for rare RIL codons. Cells were cultured at 37°C to an OD600 of 0.6 in LB or M9 medium supplemented with $^{15}N\ NH_4Cl$. Expression was induced by addition of 0.5 mM IPTG and cells were shifted to 25°C overnight. Pelleted cells were resuspended and lysed in buffer E (25 mM Tris pH 8.0, 150 mM NaCl, 5 mM imidazole, 1 mM DTT) supplemented with 0.1% Triton X-100 and 100 µg/mL lysozyme. Cellular debris was removed by centrifugation for 30 min at 50,000*g*. The supernatant was applied to a Ni-NTA column, washed with buffer E, and protein was eluted with buffer E supplemented with an additional 200 mM imidazole. Untagged protein, used for NMR and fluorescence polarization experiments, was cleaved using TEV protease and dialyzed into buffer E before applying it to a second Ni-NTA column. Hexa-His-tagged protein was diluted into 25 mM Tris pH 8.0, 10 mM NaCl and applied to a HiTrap Q-column, washed with 25 mM Tris pH 8.0, 100 mM NaCl, and eluted with buffer containing 250 mM NaCl. Both tagged and untagged protein were concentrated and purified on a Superdex 200 column. Fractions containing monomeric protein were pooled and used for separation and titration experiments.

Separation of capped and uncapped RNAs

The mixture of capped and uncapped RNA was prepared in buffer E (25 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT, 5 mM imidazole). A twofold molar excess of RNase-free His-eIF4E over total RNA concentration was added. Also, 0.2 mL Ni-NTA beads per used mg of His-eIF4E were equilibrated in buffer E and added to the reaction. After incubation for 10 min on ice, the mixture was applied to a spin column and centrifuged for 30 sec at 100*g*. The beads were washed three times with 5 volumes buffer E with the uncapped RNA residing in the flow-through. The purified complex of His-eIF4E and capped RNA was released from the Ni-NTA beads with buffer E supplemented with 200 mM imidazole. To remove the eIF4E protein from the capped RNA, the elution fraction was phenol–chloroform extracted.

Fluorescein-labeling of RNA

Fluorescein-labeled RNA was designed to contain only one uracil nucleotide and was transcribed substituting UTP with thio-UTP and purified and capped as described above. RNA was then fluorescein labeled by incubating the RNA at a concentration between 30 and 100 µM in the presence of 10 mM fluorescein in 100 mM sodium phosphate buffer (pH 8.0) for 24 h at room temperature in the dark (Ramos and Varani 1998; Audin et al. 2016). Afterward, the RNA was purified by three subsequent rounds of RNA precipitation.

Fluorescence anisotropy experiments

Fluorescence polarization experiments were carried out in buffer containing 25 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT, and 0.005% Triton X-100 at room temperature in a Tecan infinite f200 plate reader. Excitation and emission wavelength were 485 and 535 nm, respectively. Concentration of fluorescein-labeled 31-mer RNA was 20 nM, and eIF4E concentrations ranged between 0 and 2.5 µM. Samples were prepared in triplicate and blanked against buffer only and anisotropy was averaged over ten measurements. To obtain a binding constant, the binding curves were fitted to a one-site binding model using a least-squares routine. The binding process did not influence the fluorescence intensity, which is consistent with the 15-nt distance between the label and the eIF4E binding site.

Generation of ^{13}C-methyl-labeled *S*-adenosyl methionine

The purification of MetK and the synthesis of *S*-adenosyl methionine were carried out as described previously (Ottink et al. 2010). The plasmid carrying an N-terminally His-tagged MetK was a kind gift of Professor Jens Wöhnert (University of Frankfurt am Main).

In brief, ^{13}C-methyl-labeled *S*-adenosyl methionine is synthesized from methionine and ATP by the *E. coli* MetK enzyme. The synthesis was carried out in 50 mM Tris, pH 8.2, 100 mM KCl, 10 mM $MgCl_2$, 1 mM L-methionine methyl-^{13}C, 1 mM ATP in the presence of 0.6 mg/mL MetK. The reaction was incubated for 3 h at 30°C. An SP sepharose column (HiTrap SP-FF; GE healthcare) of 30% of the reaction volume was washed with two column volumes of 500 mM HCl and then equilibrated with 20 column volumes of 50 mM HCl. The reaction mixture was adjusted to 50 mM HCl with concentrated HCl and applied to the column. After a wash step with 6 column volumes of 50 mM HCl, pure SAM was eluted with 500 mM HCl. Before storage, SAM was concentrated and the pH was adjusted to 3 with NaOH before storage at −20°C.

NMR spectroscopy

NMR spectra were recorded at 25°C on a Bruker AVIII-600 spectrometer with a room temperature probe head. Samples were prepared in 25 mM Tris, pH 8.0 and 150 mM NaCl, which was based on 100% D_2O or 90% H_2O/10% D_2O for ^{13}C-labeled RNA samples and ^{15}N-labeled protein samples, respectively. The concentration of both the eIF4E sample and the cap-1 RNA sample was 100 µM and for the titration steps a molar excess of 1.5 of interacting RNA or protein was added. NMR data were processed using the NMRPipe-NMRDraw software suite and figures displaying NMR spectra were produced using NMRView (www.onemoonscientific.com).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank all laboratory members for discussions and support. We especially thank Philip Wurm for providing the fluorescently labeled capped RNA and for help in recording the fluorescence anisotropy data. We are grateful to Jens Wöhnert (University of Frankfurt am Main) for sharing the *S*-adenosyl methionine synthetase plasmid and to Fulvia Bono (MPI Tübingen) for sharing the eIF4E plasmid. This work was supported by the Max Planck Society and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013), ERC grant agreement no. 616052.

Received March 21, 2016; accepted May 25, 2016.

REFERENCES

Adams JM, Cory S. 1975. Modified nucleosides and bizarre 5′-termini in mouse myeloma mRNA. *Nature* **255:** 28–33.

Audin MJ, Dorn G, Fromm SA, Reiss K, Schutz S, Vorlander MK, Sprangers R. 2013. The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding. *Angew Chem* **52:** 8312–8316.

Audin MJ, Wurm JP, Cvetkovic MA, Sprangers R. 2016. The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation. *Nucleic Acids Res* **44:** 2962–2973.

Banerjee AK. 1980. 5′-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiol Rev* **44:** 175–205.

Barbosa E, Moss B. 1978. mRNA(nucleoside-2′-)-methyltransferase from vaccinia virus. Purification and physical properties. *J Biol Chem* **253:** 7692–7697.

Both GW, Banerjee AK, Shatkin AJ. 1975. Methylation-dependent translation of viral messenger RNAs in vitro. *Proc Natl Acad Sci* **72:** 1189–1193.

Brownlee GG, Fodor E, Pritlove DC, Gould KG, Dalluge JJ. 1995. Solid phase synthesis of 5′-diphosphorylated oligoribonucleotides and their conversion to capped m^{7}Gppp-oligoribonucleotides for use as primers for influenza A virus RNA polymerase in vitro. *Nucleic Acids Res* **23:** 2641–2647.

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11:** 3319–3326.

Coleman TM, Wang G, Huang F. 2004. Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 φ2.5 promoter. *Nucleic Acids Res* **32:** e14.

Contreras R, Cheroutre H, Degrave W, Fiers W. 1982. Simple, efficient in vitro synthesis of capped RNA useful for direct expression of cloned eukaryotic genes. *Nucleic Acids Res* **10:** 6353–6362.

Daffis S, Szretter KJ, Schriewer J, Li J, Youn S, Errett J, Lin TY, Schneller S, Zust R, Dong H, et al. 2010. 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468:** 452–456.

De la Pena M, Kyrieleis OJ, Cusack S. 2007. Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J* **26:** 4913–4925.

Decroly E, Ferron F, Lescar J, Canard B. 2012. Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Microbiol* **10:** 51–65.

Devarkar SC, Wang C, Miller MT, Ramanathan A, Jiang F, Khan AG, Patel SS, Marcotrigiano J. 2016. Structural basis for m^{7}G recognition and 2′-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc Natl Acad Sci* **113:** 596–601.

Duss O, Maris C, von Schroetter C, Allain FH. 2010. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res* **38:** e188.

Ensinger MJ, Martin SA, Paoletti E, Moss B. 1975. Modification of the 5′-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. *Proc Natl Acad Sci* **72:** 2525–2529.

Filipowicz W. 1978. Functions of the 5,-terminal m^{7}G cap in eukaryotic mRNA. *FEBS Lett* **96:** 1–11.

Fresco LD, Buratowski S. 1996. Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA* **2:** 584–596.

Fromm SA, Truffault V, Kamenz J, Braun JE, Hoffmann NA, Izaurralde E, Sprangers R. 2012. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J* **31:** 279–290.

Furuichi Y, Miura K. 1975. A blocked structure at the 5′ terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253:** 374–375.

Furuichi Y, Shatkin AJ. 2000. Viral and cellular mRNA capping: past and prospects. *Adv Virus Res* **55:** 135–184.

Furuichi Y, LaFiandra A, Shatkin AJ. 1977. 5′-Terminal structure and mRNA stability. *Nature* **266:** 235–239.

Goldeck M, Tuschl T, Hartmann G, Ludwig J. 2014. Efficient solid-phase synthesis of pppRNA by using product-specific labeling. *Angew Chem* **53:** 4694–4698.

Green MR, Maniatis T, Melton DA. 1983. Human β-globin pre-mRNA synthesized in vitro is accurately spliced in *Xenopus* oocyte nuclei. *Cell* **32:** 681–694.

Grudzien-Nogalska E, Stepinski J, Jemielity J, Zuberek J, Stolarski R, Rhoads RE, Darzynkiewicz E. 2007. Synthesis of anti-reverse cap analogs (ARCAs) and their applications in mRNA translation and stability. *Methods Enzymol* **431:** 203–227.

Gu M, Rajashankar KR, Lima CD. 2010. Structure of the *Saccharomyces cerevisiae* Cet1-Ceg1 mRNA capping apparatus. *Structure* **18:** 216–227.

Gunawardana D, Domashevskiy AV, Gayler KR, Goss DJ. 2015. Efficient preparation and properties of mRNAs containing a fluorescent cap analog: Anthraniloyl-m^{7}GpppG. *Translation (Austin)* **3:** e988538.

Hamm J, Mattaj IW. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* **63:** 109–118.

Helmling C, Keyhani S, Sochor F, Furtig B, Hengesbach M, Schwalbe H. 2015. Rapid NMR screening of RNA secondary structure and binding. *J Biomol NMR* **63:** 67–76.

Hodel AE, Gershon PD, Quiocho FA. 1998. Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol Cell* **1:** 443–447.

Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78:** 657–668.

Kinkelin K, Veith K, Grunwald M, Bono F. 2012. Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression. *RNA* **18:** 1624–1634.

Konarska MM, Padgett RA, Sharp PA. 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* **38:** 731–736.

Kyrieleis OJ, Chang J, de la Pena M, Shuman S, Cusack S. 2014. Crystal structure of vaccinia virus mRNA capping enzyme provides insights into the mechanism and evolution of the capping apparatus. *Structure* **22:** 452–465.

Lewdorowicz M, Jemielity J, Kierzek R, Shapira M, Stepinski J, Darzynkiewicz E. 2007. Solid-supported synthesis of 5′-mRNA CAP-4 from Trypanosomatids. *Nucleosides Nucleotides Nucleic Acids* **26:** 1329–1333.

Liu W, Zhao R, McFarland C, Kieft J, Niedzwiecka A, Jankowska-Anyszka M, Stepinski J, Darzynkiewicz E, Jones DN, Davis RE. 2009. Structural insights into parasite eIF4E binding specificity for m^{7}G and m^{2,2,7}G mRNA caps. *J Biol Chem* **284:** 31336–31349.

Mao X, Shuman S. 1994. Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit. Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. *J Biol Chem* **269:** 24472–24479.

Martin SA, Paoletti E, Moss B. 1975. Purification of mRNA guanylyltransferase and mRNA (guanine-7-) methyltransferase from vaccinia virions. *J Biol Chem* **250:** 9322–9329.

Matsuo H, Li H, McGuire AM, Fletcher CM, Gingras AC, Sonenberg N, Wagner G. 1997. Structure of translation factor eIF4E bound to m^{7}GDP and interaction with 4E-binding protein. *Nat Struct Biol* **4:** 717–724.

Matsuo H, Moriguchi T, Takagi T, Kusakabe T, Buratowski S, Sekine M, Kyogoku Y, Wagner G. 2000. Efficient synthesis of ^{13}C, ^{15}N-labeled RNA containing the cap structure m^{7}GpppA. *J Am Chem Soc* **122:** 2417–2421.

Myette JR, Niles EG. 1996. Domain structure of the vaccinia virus mRNA capping enzyme. Expression in *Escherichia coli* of a subdomain possessing the RNA 5′-triphosphatase and guanylyltransferase activities and a kinetic comparison to the full-size enzyme. *J Biol Chem* **271:** 11936–11944.

Nagata S, Hamasaki T, Uetake K, Masuda H, Takagaki K, Oka N, Wada T, Ohgi T, Yano J. 2010. Synthesis and biological activity of artificial mRNA prepared with novel phosphorylating reagents. *Nucleic Acids Res* **38:** 7845–7857.

Niedzwiecka A, Marcotrigiano J, Stepinski J, Jankowska-Anyszka M, Wyslouch-Cieszynska A, Dadlez M, Gingras AC, Mak P, Darzynkiewicz E, Sonenberg N, et al. 2002. Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5′ cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J Mol Biol* **319:** 615–635.

Nielsen DA, Shapiro DJ. 1986. Preparation of capped RNA transcripts using T7 RNA polymerase. *Nucleic Acids Res* **14:** 5936.

Ottink OM, Nelissen FH, Derks Y, Wijmenga SS, Heus HA. 2010. Enzymatic stereospecific preparation of fluorescent S-adenosyl-L-methionine analogs. *Anal Biochem* **396:** 280–283.

Parker R, Song H. 2004. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* **11:** 121–127.

Pasquinelli AE, Dahlberg JE, Lund E. 1995. Reverse 5′ caps in RNAs made in vitro by phage RNA polymerases. *RNA* **1:** 957–967.

Paterson BM, Rosenberg M. 1979. Efficient translation of prokaryotic mRNAs in a eukaryotic cell-free system requires addition of a cap structure. *Nature* **279:** 692–696.

Pelletier J, Sonenberg N. 1985. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* **40:** 515–526.

Peter D, Weber R, Kone C, Chung MY, Ebertsch L, Truffault V, Weichenrieder O, Igreja C, Izaurralde E. 2015. Mextli proteins use both canonical bipartite and novel tripartite binding modes to form eIF4E complexes that display differential sensitivity to 4E-BP regulation. *Genes Dev* **29:** 1835–1849.

Peyrane F, Selisko B, Decroly E, Vasseur JJ, Benarroch D, Canard B, Alvarez K. 2007. High-yield production of short GpppA- and ^{7Me}GpppA-capped RNAs and HPLC-monitoring of methyltransfer reactions at the guanine-N7 and adenosine-2′O positions. *Nucleic Acids Res* **35:** e26.

Ramos A, Varani G. 1998. A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J Am Chem Soc* **120:** 10992–10993.

Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, Dong H, Deas TS, Zhou Y, Li H, Shi PY. 2006. West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. *J Virol* **80:** 8362–8370.

Reddy R, Ro-Choi TS, Henning D, Busch H. 1974. Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J Biol Chem* **249:** 6486–6494.

Reguera J, Gerlach P, Cusack S. 2016. Towards a structural understanding of RNA synthesis by negative strand RNA viral polymerases. *Curr Opin Struct Biol* **36:** 75–84.

Salditt-Georgieff M, Harpold M, Chen-Kiang S, Darnell JE Jr. 1980. The addition of 5′ cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19:** 69–78.

Sawai H, Wakai H, Nakamura-Ozaki A. 1999. Synthesis and reactions of nucleoside 5′-diphosphate imidazolide. A nonenzymatic capping agent for 5′-monophosphorylated oligoribonucleotides in aqueous solution. *J Org Chem* **64:** 5836–5840.

Schnierle BS, Gershon PD, Moss B. 1992. Cap-specific mRNA (nucleoside-O2′-)-methyltransferase and poly(A) polymerase stimulatory activities of vaccinia virus are mediated by a single protein. *Proc Natl Acad Sci* **89:** 2897–2901.

Schwer B, Shuman S. 1996. Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA* **2:** 574–583.

Shatkin AJ. 1976. Capping of eucaryotic mRNAs. *Cell* **9:** 645–653.

Shuman S, Surks M, Furneaux H, Hurwitz J. 1980. Purification and characterization of a GTP-pyrophosphate exchange activity from vaccinia virions. Association of the GTP-pyrophosphate exchange activity with vaccinia mRNA guanylyltransferase. RNA (guanine-7-)methyltransferase complex (capping enzyme). *J Biol Chem* **255:** 11588–11598.

Smietanski M, Werner M, Purta E, Kaminska KH, Stepinski J, Darzynkiewicz E, Nowotny M, Bujnicki JM. 2014. Structural analysis of human 2′-O-ribose methyltransferases involved in mRNA cap structure formation. *Nat Commun* **5:** 3004.

Sripati CE, Groner Y, Warner JR. 1976. Methylated, blocked 5′ termini of yeast mRNA. *J Biol Chem* **251:** 2898–2904.

Takagi T, Moore CR, Diehn F, Buratowski S. 1997. An RNA 5′-triphosphatase related to the protein tyrosine phosphatases. *Cell* **89:** 867–873.

Thillier Y, Decroly E, Morvan F, Canard B, Vasseur JJ, Debart F. 2012. Synthesis of 5′ cap-0 and cap-1 RNAs using solid-phase chemistry

coupled with enzymatic methylation by human (guanine-N^7)-methyl transferase. *RNA* **18:** 856–868.

Tomoo K, Matsushita Y, Fujisaki H, Abiko F, Shen X, Taniguchi T, Miyagawa H, Kitamura K, Miura K, Ishida T. 2005. Structural basis for mRNA Cap-Binding regulation of eukaryotic initiation factor 4E by 4E-binding protein, studied by spectroscopic, X-ray crystal structural, and molecular dynamics simulation methods. *Biochim Biophys Acta* **1753:** 191–208.

Wei CM, Moss B. 1975. Methylated nucleotides block 5′-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci* **72:** 318–322.

Yap LJ, Luo D, Chung KY, Lim SP, Bodenreider C, Noble C, Shi PY, Lescar J. 2010. Crystal structure of the dengue virus methyltransferase bound to a 5′-capped octameric RNA. *PLoS One* **5:** e12836.

Zhao Y, Soh TS, Lim SP, Chung KY, Swaminathan K, Vasudevan SG, Shi PY, Lescar J, Luo D. 2015. Molecular basis for specific viral RNA recognition and 2′-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5). *Proc Natl Acad Sci* **112:** 14834–14839.

Zimmern D. 1975. The 5′ end group of tobacco mosaic virus RNA is $m^7G^{5'}$ $ppp^{5'}$ Gp. *Nucleic Acids Res* **2:** 1189–1201.

A general method for rapid and cost-efficient large-scale production of 5′ capped RNA

Anna-Lisa Fuchs, Ancilla Neu and Remco Sprangers

RNA 2016 22: 1454-1466 originally published online July 1, 2016
Access the most recent version at doi:10.1261/rna.056614.116

SUPPLEMENTARY INFORMATION

A general method for rapid and cost efficient large-scale production of 5' capped RNA.

Anna-Lisa Fuchs[1], Ancilla Neu[1] and Remco Sprangers*

Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

[1] Equal contribution.

* To whom correspondence should be addressed.

Email: remco.sprangers@tuebingen.mpg.de

```
gaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaat
      T7 promotor #1                                    XbaI

aattttgatttaactttaagaaggagatatacc
                    RBS
```

START OF THE D1 PROTEIN

```
atgaaacatcaccatcaccatcaccccatgagcgattacgacatccccactactgagaat
 M  K  H  H  H  H  H  H  P  M  S  D  Y  D  I  P  T  T  E  N

ctttattttcagggcgccatggacgctaatgtcgtgtcttcttctaccatcgcaacctat
 L  Y  F  Q  G  A  M  D  A  N  V  V  S  S  S  T  I  A  T  Y
    TEV site      NcoI
attgacgctctggcaaaaaacgcctcggaactggaacaacgctcaaccgcgtatgaaatc
 I  D  A  L  A  K  N  A  S  E  L  E  Q  R  S  T  A  Y  E  I

aacaatgaactggaactggtgtttatcaaaccgccgctgattacgctgaccaacgtggtt
 N  N  E  L  E  L  V  F  I  K  P  P  L  I  T  L  T  N  V  V

aatatcagcaccattcaggaatcttttattcgtttcacggttaccaacaaagaaggcgtc
 N  I  S  T  I  Q  E  S  F  I  R  F  T  V  T  N  K  E  G  V

aaaatccgcacgaaaattccgctgagcaaagttcatggtctggatgtgaaaaacgttcaa
 K  I  R  T  K  I  P  L  S  K  V  H  G  L  D  V  K  N  V  Q

ctggtcgacgcaatcgataatattgtgtgggaaaagaaaagcctggttaccgaaaatcgt
 L  V  D  A  I  D  N  I  V  W  E  K  K  S  L  V  T  E  N  R

ctgcataaagaatgcctgctgcgtctgagcacggaagaacgccacatctttctggactat
 L  H  K  E  C  L  L  R  L  S  T  E  E  R  H  I  F  L  D  Y

aaaaaatacggcagctctatccgcctggaactggtgaacctgatccaggctaaaaccaaa
 K  K  Y  G  S  S  I  R  L  E  L  V  N  L  I  Q  A  K  T  K

aacttcacgatcgatttcaaactgaaatattttctgggcagtggtgctcaatccaaaagt
 N  F  T  I  D  F  K  L  K  Y  F  L  G  S  G  A  Q  S  K  S

tccctgctgcatgcgatcaaccacccgaaaagtcgtccgaatacctccctggaaattgaa
 S  L  L  H  A  I  N  H  P  K  S  R  P  N  T  S  L  E  I  E

ttcaccccgcgcgacaacgaaacggtgccgtacgatgaactgattaaagaactgaccacg
 F  T  P  R  D  N  E  T  V  P  Y  D  E  L  I  K  E  L  T  T

ctgtcacgtcatatctttatggcgtcgccggaaaacgttattctgagcccgccgatcaat
 L  S  R  H  I  F  M  A  S  P  E  N  V  I  L  S  P  P  I  N

gccccgattaaaaccttcatgctgccgaaacaggacattgttggcctggatctggaaaac
 A  P  I  K  T  F  M  L  P  K  Q  D  I  V  G  L  D  L  E  N

ctgtatgcggtcacgaaaaccgatggtattccgatcaccattcgcgtgacgtcgaatggc
 L  Y  A  V  T  K  T  D  G  I  P  I  T  I  R  V  T  S  N  G

ctgtattgctactttacccacctgggttatattatccgttacccggttaaacgcattatc
 L  Y  C  Y  F  T  H  L  G  Y  I  I  R  Y  P  V  K  R  I  I

gactccgaagtcgtggttttcggcgaagcggtcaaagataaaaattggaccgtgtatctg
 D  S  E  V  V  V  F  G  E  A  V  K  D  K  N  W  T  V  Y  L
```

```
atcaaactgattgaaccggtgaacgccatcaacgatcgtctggaagaatcaaaatacgtg
 I  K  L  I  E  P  V  N  A  I  N  D  R  L  E  E  S  K  Y  V

gaatcgaaactggttgacatctgtgatcgcatcgttttcaaaagcaaaaaatacgaaggt
 E  S  K  L  V  D  I  C  D  R  I  V  F  K  S  K  K  Y  E  G

ccgttcaccacgacctctgaagtcgtggatatgctgagtacctatctgccgaaacagccg
 P  F  T  T  T  S  E  V  V  D  M  L  S  T  Y  L  P  K  Q  P

gaaggcgtgatcctgttttacagcaaaggtccgaaatctaacatcgacttcaaaatcaaa
 E  G  V  I  L  F  Y  S  K  G  P  K  S  N  I  D  F  K  I  K

aaagaaaacaccatcgatcaaacggccaatgttgtctttcgttatatgtcatcggaaccg
 K  E  N  T  I  D  Q  T  A  N  V  V  F  R  Y  M  S  S  E  P

attatctttggcgaaagctctatcttcgtggaatacaaaaaattctcgaacgataaaggc
 I  I  F  G  E  S  S  I  F  V  E  Y  K  K  F  S  N  D  K  G

ttcccgaaagaatacggcagcggtaaaattgtcctgtataacggtgtgaattacctgaac
 F  P  K  E  Y  G  S  G  K  I  V  L  Y  N  G  V  N  Y  L  N

aatatctattgcctggaatacattaacacccataatgaagttggcattaaatctgtggtt
 N  I  Y  C  L  E  Y  I  N  T  H  N  E  V  G  I  K  S  V  V

gtcccgatcaaatttattgcagaattcctggtcaacggtgaaatcctgaaaccgcgtatt
 V  P  I  K  F  I  A  E  F  L  V  N  G  E  I  L  K  P  R  I

gacaaaaccatgaaatacatcaacagtgaagattactacggtaaccagcataacatcatc
 D  K  T  M  K  Y  I  N  S  E  D  Y  Y  G  N  Q  H  N  I  I

gtggaacacctgcgcgaccaatctatcaaaatcggcgatatcttcaacgaagacaaactg
 V  E  H  L  R  D  Q  S  I  K  I  G  D  I  F  N  E  D  K  L

agtgatgtcggtcaccagtatgcgaacaatgataaatttcgtctgaacccggaagtgtcc
 S  D  V  G  H  Q  Y  A  N  N  D  K  F  R  L  N  P  E  V  S

tacttcaccaataaacgtacgcgcggcccgctgggtatcctgtcaaattatgtcaaaacc
 Y  F  T  N  K  R  T  R  G  P  L  G  I  L  S  N  Y  V  K  T

ctgctgatttcaatgtactgttcgaaaacgtttctggatgacagcaacaaacgcaaagtt
 L  L  I  S  M  Y  C  S  K  T  F  L  D  D  S  N  K  R  K  V

ctggccattgactttggcaatggtgcagatctggaaaaatatttctacggcgaaatcgct
 L  A  I  D  F  G  N  G  A  D  L  E  K  Y  F  Y  G  E  I  A

ctgctggttgcgaccgatccggacgcggatgccattgcacgtggcaacgaacgctataac
 L  L  V  A  T  D  P  D  A  D  A  I  A  R  G  N  E  R  Y  N

aaactgaattctggtatcaaaaccaaatactacaaattcgactacatccaggaaaccatt
 K  L  N  S  G  I  K  T  K  Y  Y  K  F  D  Y  I  Q  E  T  I

cgtagtgatacgttcgtgagttccgttcgcgaagtcttttatttcggcaaattcaacatc
 R  S  D  T  F  V  S  S  V  R  E  V  F  Y  F  G  K  F  N  I

atcgattggcaattcgccatccattattctttccatccgcgtcactacgcaaccgtgatg
 I  D  W  Q  F  A  I  H  Y  S  F  H  P  R  H  Y  A  T  V  M

aacaatctgagtgaactgacggcttccggcggtaaagttctgattacgacgatggatggt
 N  N  L  S  E  L  T  A  S  G  G  K  V  L  I  T  T  M  D  G

gataaactgtccaaactgaccgataagaaaaccttcattatccacaaaaacctgccgtca
 D  K  L  S  K  L  T  D  K  K  T  F  I  I  H  K  N  L  P  S
```

```
tcggaaaactacatgtcagtggaaaaaatcgccgatgaccgcattgtggtttataacccg
 S  E  N  Y  M  S  V  E  K  I  A  D  D  R  I  V  V  Y  N  P

agcacgatgtctaccccgatgacggaatacatcattaagaaaaacgatatcgtccgtgtg
 S  T  M  S  T  P  M  T  E  Y  I  I  K  K  N  D  I  V  R  V

tttaatgaatacggtttcgttctggtcgacaacgttgattttgcaaccattatcgaacgc
 F  N  E  Y  G  F  V  L  V  D  N  V  D  F  A  T  I  I  E  R

agcaaaaaattcatcaatggcgcttccacgatggaagatcgtccgtcaacgcgcaacttt
 S  K  K  F  I  N  G  A  S  T  M  E  D  R  P  S  T  R  N  F

ttcgaactgaatcgcggtgcaattaaatgtgaaggtctggatgtggaagatctgctgtcc
 F  E  L  N  R  G  A  I  K  C  E  G  L  D  V  E  D  L  L  S

tattatgtcgtgtatgtgttctctaaacgctaa
 Y  Y  V  V  Y  V  F  S  K  R  -


gctagctaggatccgaattcgagctcggcgcgcctgcaggtcgacaagcttgcggccgca
         BamHI

taatgcttaagtcgaacagaaagtaatcgtattgtacacggccgcataatcgaaattaat
                                    DuetUP2 Primer
                                   DuetDOWN1 Primer
acgactcactatagggggaattgtgagcggataacaattccccatcttagtatattagtta
T7 promotor #2

agtataagaaggagatatacat
         RBS       NdeI
```

START OF THE D12 PROTEIN

```
atggatgaaatcgtcaaaaatatccgcgaaggcacgcacgtcctgctgccgttctatgaa
 M  D  E  I  V  K  N  I  R  E  G  T  H  V  L  L  P  F  Y  E

accctgccggaactgaatctgtcactgggcaaatctccgctgccgagtctggaatatggt
 T  L  P  E  L  N  L  S  L  G  K  S  P  L  P  S  L  E  Y  G

gcaaactactttctgcagatttctcgtgtgaacgatctgaatcgcatgccgaccgacatg
 A  N  Y  F  L  Q  I  S  R  V  N  D  L  N  R  M  P  T  D  M

ctgaaactgttcacgcatgatatcatgctgccggaaagcgatctggacaaagtctacgaa
 L  K  L  F  T  H  D  I  M  L  P  E  S  D  L  D  K  V  Y  E

atcctgaaaatcaactccgttaaatactacggccgttcaaccaaagcggatgccgtggtt
 I  L  K  I  N  S  V  K  Y  Y  G  R  S  T  K  A  D  A  V  V

gcagacctgtccgctcgcaataaactgtttaaacgtgaacgcgatgctattaaatcgaac
 A  D  L  S  A  R  N  K  L  F  K  R  E  R  D  A  I  K  S  N

aatcacctgaccgaaaacaacctgtacatcagcgattacaaaatgctgacgtttgacgtg
 N  H  L  T  E  N  N  L  Y  I  S  D  Y  K  M  L  T  F  D  V

ttccgtccgctgttcgatttcgttaacgaaaaatactgcatcatcaaactgccgaccctg
 F  R  P  L  F  D  F  V  N  E  K  Y  C  I  I  K  L  P  T  L

tttggccgtggtgtgattgatacgatgcgcatctactgcagcctgttcaaaaatgtccgc
 F  G  R  G  V  I  D  T  M  R  I  Y  C  S  L  F  K  N  V  R

ctgctgaaatgtgtgtcggatagctggctgaaagactctgcgattatggtggccagtgac
 L  L  K  C  V  S  D  S  W  L  K  D  S  A  I  M  V  A  S  D
```

```
gtttgtaagaaaaacctggacctgtttatgtcccatgtcaaatcagtgaccaaaagctct
 V  C  K  K  N  L  D  L  F  M  S  H  V  K  S  V  T  K  S  S

agttggaaagacgttaattcggtccaatttagcattctgaacaatccggttgatacggaa
 S  W  K  D  V  N  S  V  Q  F  S  I  L  N  N  P  V  D  T  E

ttcatcaacaaattcctggaattctctaaccgtgtttacgaagcactgtattacgtccac
 F  I  N  K  F  L  E  F  S  N  R  V  Y  E  A  L  Y  Y  V  H

agtctgctgtactcctcaatgacctcggactccaaatccatcgaaaataaacatcaacgc
 S  L  L  Y  S  S  M  T  S  D  S  K  S  I  E  N  K  H  Q  R

cgcctggtgaaactgctgctgtaa
 R  L  V  K  L  L  L  -


ctcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagct
 XhoI

gagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgg
                  T7 Terminator

gtcttgaggggttttttgctgaaaggaggaactatatccggat
```

FIGURE S1. Vaccinia virus Capping Enzyme. Plasmid design, DNA and protein sequences. (Internal ID: 1323).

Figure S2

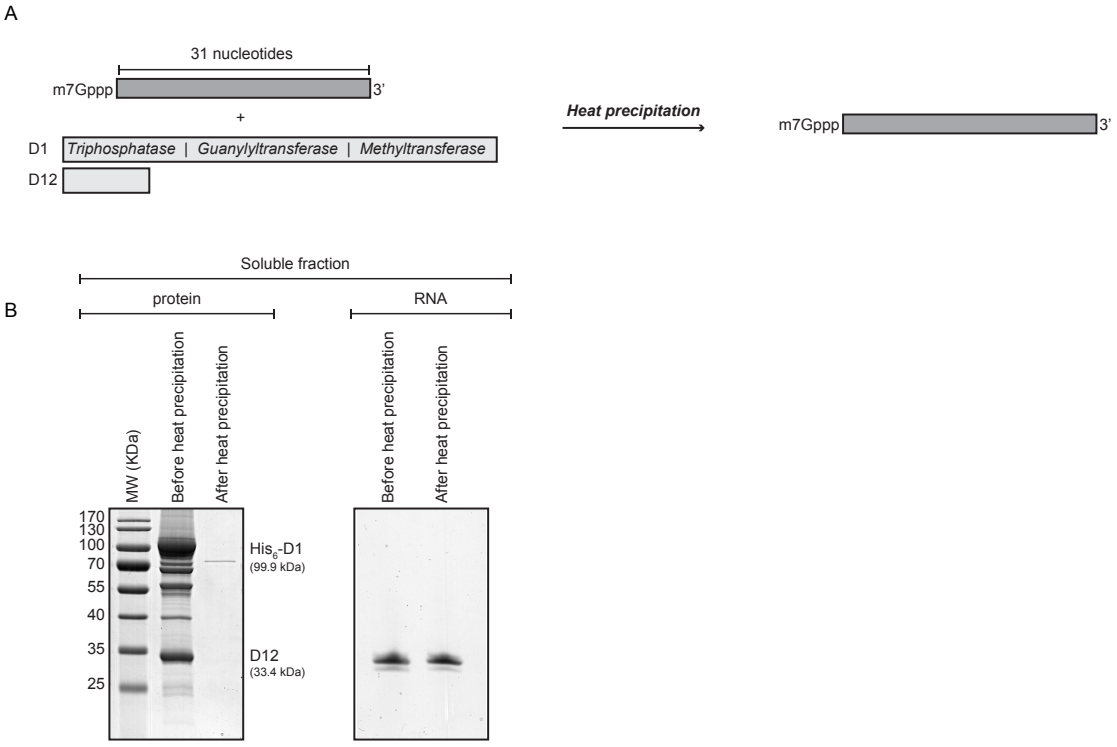


FIGURE S2: To separate the capped RNA from the capping enzyme the reaction mixture was incubated at 75°C for 10 minutes, followed by a 10 minute centrifugation step at 4500g. The capping enzyme was quantitatively removed from the mixture, whereas the RNA remained in solution.

Figure S3

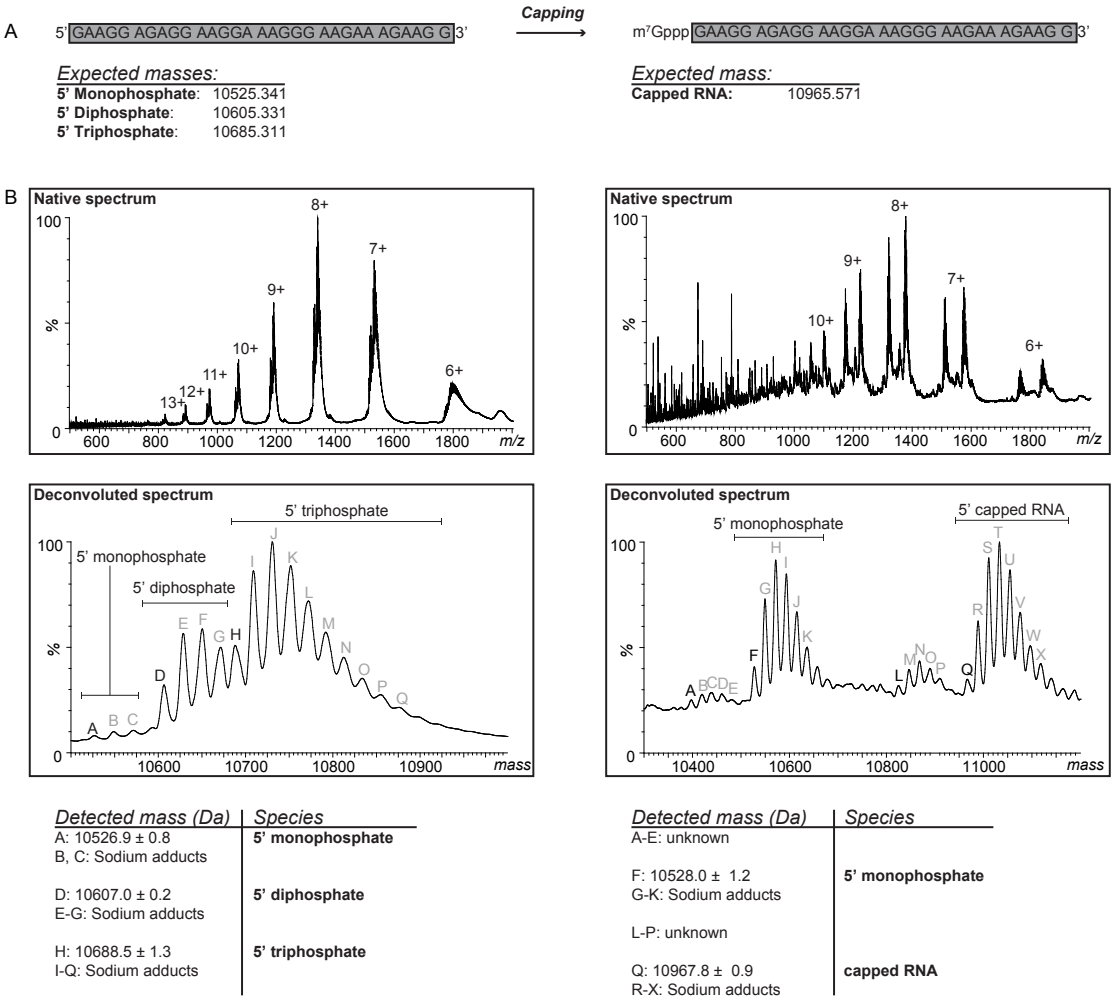


FIGURE S3. Native mass spectrometry analysis of the RNA capping reaction. *(A)* A mixture of 5' mono-, 5' di- and 5' triphosphate RNA was subjected to the vaccinia virus capping enzyme. The 3' end of the RNA contained a cyclic phosphate as a result of ribozyme cleavage. The RNA sequence and the expected masses of the substrates and product are indicated. *(B)* The experimental mass of the substrate confirmed the presence of mono-, di- and triphosphate at the 5' end of the substrate RNA. The experimental mass of the product shows a mixture of mono-phosphate RNA and of capped RNA. The analysis shows that both 5' di- and 5' triphosphate RNA are efficiently and properly capped by the vaccinia virus capping enzyme, whereas 5' monophosphate RNA appears not to be a substrate of the enzyme. The native mass spectrometry analysis was performed by AbLab (http://www.hecklab.com/ablab), a service unit within the Heck lab (http://www.hecklab.com) of the Utrecht University.

Figure S4

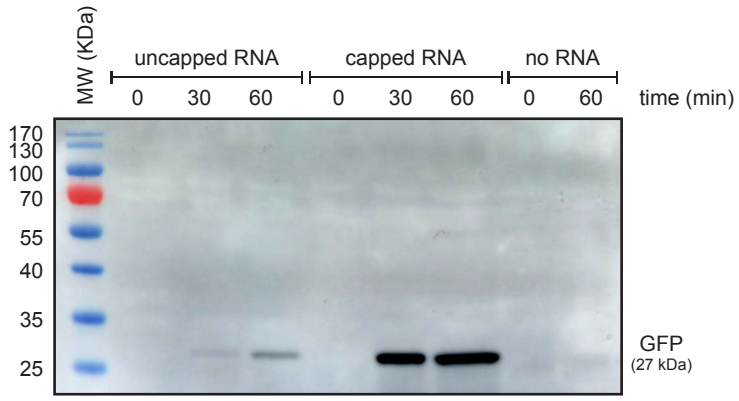


FIGURE S4. *In vitro* translation assays show efficient translation of a capped RNA that codes for the GFP protein.

Fractions of an *in vitro* translation reaction (see below) were loaded on an SDS PAGE gel and the transcribed GFP protein was visualized using a GFP antibody (see below). Samples were taken after 0, 30 or 60 minutes, as indicated on top of the gel. The left three lanes are samples that are taken from a translation reaction in the presence of RNA that has not been capped, in the middle three an RNA that was capped according to the protocols described here was used and in the right two lanes the reaction took place in the absence of RNA.

The DNA template coding for a GFP mRNA was prepared by polymerase chain reaction. PCR was carried out according to manufacturer's recommendations with Q5 polymerase (NEB) on a GFP carrying plasmid. In addition to the sequence specific part, the forward primer and reverse primer encompassed a Kozak sequence and a stop codon followed by 23 adenosine bases, respectively (Forward primer: CGTAATACGACTCACTATAGGCCGCCACCATGGTGAGCAAGGGCGAG, reverse primer: TTTTTTTTTTTTTTTTTTTTTTTTTTATTACTTGTACAGCTCGTCCATGCC). The resulting double stranded DNA was purified (Macherey-Nagel) and used as template in a transcription reaction at a final concentration of 8 ng/µl. The produced RNA was precipitated and dissolved in RNase free water and capped as described at a molar ratio of 1:100 at 37 °C for 1 hour. For uncapped RNA, the capping enzyme was omitted from the reaction. RNA was used without further purification for *in vitro* translation. Commercial wheat germ extract (Promega) was utilized following manufacturer's instruction including capped and uncapped RNA at 20µg/ml and 50 mM potassium acetate. Samples for SDS PAGE were taken at 0 min, 30 min and 60 min reaction time. Translated GFP protein was detected by western blotting using α-GFP (mouse) (Roche) and α-mouse-peroxidase (Sigma) antibodies as primary and secondary antibodies respectively. Chemiluminescence was developed in WesternBright substrate (Advansta) and detected on an Amersham Imager 600 (GE).

Figure S5

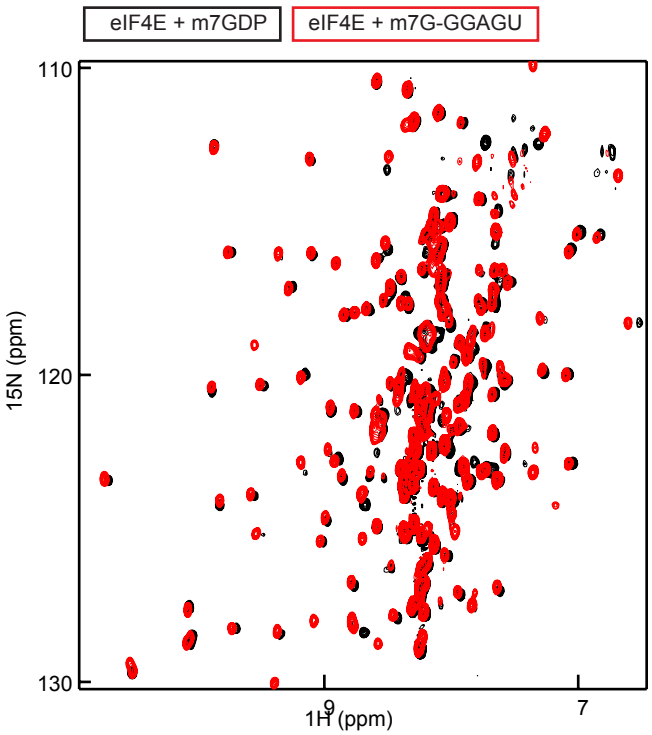


FIGURE S5. NMR spectra of *Dm* eIF4E in the presence of m^7GDP (black) and in the presence of capped RNA. The small chemical shift differences between the two spectra confirm that the RNA body does not interact extensively with the cap binding protein.

Table S1: RNA constructs used in this study.

Construct	RNA sequence	Anti-sense primer sequence
100GA14U	GGGAAGGAAGGGAAUAAAGAAG GGAAGAGGAAGGAGAGGAGGGA AGAAAGAAGAGGAGAGGAAGGA AGGGAAGAAAGAAGAGGGAAGA GGAAGGAGAGGA	TCCTCTCCTTCCTCTTCCCTCT TCTTTCTTCCCTTCCTTCCTCT CCTCTTCTTTCTTCCCTCCTCT CCTTCCTCTTCCCTTCTTTATT CCCTTCCTTCC**C**TATAGTGAGT CGTATTACG
80GA14U	GGGAAGGAAGGGAAUAAAGAAG GGAAGAGGAAGGAGAGGAGGGA AGAAAGAAGAGGAGAGGAAGGA AGGGAAGAAAGAAG	CTTCTTTCTTCCCTTCCTTCCT CTCCTCTTCTTTCTTCCCTCCT CTCCTTCCTCTTCCCTTCTTTA TTCCCTTCCTTCC**C**TATAGTGA GTCGTATTACG
60GA14U	GGGAAGGAAGGGAAUAAAGAAG GGAAGAGGAAGGAGAGGAGGGA AGAAAGAAGAGGAGAG	CTCTCCTCTTCTTTCTTCCCTC CTCTCCTTCCTCTTCCCTTCTT TATTCCCTTCCTTCC**C**TATAGT GAGTCGTATTACG
40GA14U	GGGAAGGAAGGGAAUAAAGAAG GGAAGAGGAAGGAGAGGA	TCCTCTCCTTCCTCTTCCCTTC TTTATTCCCTTCCTTCC**C**TATA GTGAGTCGTATTACG
15GA *	GGAGAAGAGAAGGAG	CTCCTTCTCTTCTC**C**TATAGTG AGTCGTATTA
15CU *	GGCCUCUUCGAAGCG	CGCTTCGAAGAGGC**C**TATAGT GAGTCGTATTA
Hp0 *	GCGGUUCGCCGCA	TGCGGCGAACCG**C**TATAGTGA GTCGTATTA
Hp2 *	GGGCGGUUCGCCGCA	TGCGGCGAACCGCC**C**TATAGT GAGTCGTATTA
Hp4 *	GGAAGCGGUUCGCCGCA	TGCGGCGAACCGCTTC**C**TATA GTGAGTCGTATTA
30GA2U	GTAAGGAGAGGAAGGAAGGGAA GAAAGAAG	CTTCTTTCTTCCCTTCCTTCCT CTCCTTA**C**TATAGTGAGTCGT ATTACG
30GA3U	GGTAGGAGAGGAAGGAAGGGAA GAAAGAAG	CTTCTTTCTTCCCTTCCTTCCT CTCCTAC**C**TATAGTGAGTCGT ATTACG
30GA5U	GGAGTGAGAGGAAGGAAGGGAA GAAAGAAG	CTTCTTTCTTCCCTTCCTTCCT CTCACTC**C**TATAGTGAGTCGT ATTACG
30GA15U	GGAGGAGAGGAAGGTAAGGGAA GAAAGAAG	CTTCTTTCTTCCCTTACCTTCC TCTCCTC**C**TATAGTGAGTCGT ATTACG
20AG	AGGAGGGAGAGGAAGGAAGG	CCTTCCTTCCTCTCCCTCC**T**AA TAGTGAGTCGTATTACG

Primers marked with * were methylated at the 2'-O position of the last two nucleotides. The sequence of the anti-sense primer duplexed with the promoter primer was underlined and the first transcribed nucleotide is marked in bold.

Table S2: T7 promoter sequences used in this study.

RNA promoter sequence	Promoter sequence
T7 promoter Φ6.5 (for start with G)	CGTAATACGACTCACTATAGG
T7 promoter Φ2.5 (for start with A)	CGTAATACGACTCACTATTAG

Molecular basis for mRNA length sensing by the DcpS decapping enzyme

Anna-Lisa Fuchs[1], Ancilla Neu[2], Remco Sprangers[1,*]

[1] Department of Biophysics I, University of Regensburg, 93053 Regensburg, Germany.

[2] Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany.

* Correspondence should be addressed to remco.sprangers@ur.de

Keywords:

Enzyme regulation, mRNA decay, NMR spectroscopy, conformational changes, scavenger decapping enzyme, decapping.

Abstract

mRNA contains a 5’ cap structure that enhances translation and protects the transcript against 5’ to 3’ exonucleolytic degradation. During mRNA turnover, this cap structure can be removed by the scavenger decapping enzyme DcpS. Here, we show that this enzyme only processes mRNA fragments that are shorter than three nucleotides. Based on a combination of methyl TROSY NMR spectroscopy and X-ray crystallography, we unravel the structural basis for this enzymatic molecular ruler. We find that the third nucleotide of a capped mRNA interferes with conformational changes in DcpS that are required for the formation of a catalytically competent active site. The substrate length sensing mechanism in DcpS is conserved from yeast to humans and ensures that the enzyme can act on mRNA fragments that result from 3’ to 5’ exosomal mRNA degradation but that it is incapable of decapping actively translated mRNAs.

Introduction

The accurate regulation of gene expression is essential in order to maintain cellular homeostasis. An irreversible way to terminate gene-expression is the degradation of an mRNA transcript. The sequence of an mRNA transcript is divided into the mRNA body that includes the coding region as well as the 3' and 5' UTRs, the 3' poly(A)-tail and the protecting 5' cap structure. These elements are degraded in a sequential manner, where the removal of one feature triggers the next degradation step. For regular mRNA degradation the decay process is initiated with the gradual shortening of the 3' poly(A)-tail by the Ccr4:Not and Pan2:Pan3 complexes [1–3]. After this rate limiting deadenylation step, the mRNA is rendered instable and will be rapidly degraded in one of two complementary pathways. In the 5' to 3' pathway, the 5' cap structure of the mRNA is first removed by the Dcp1:Dcp2 decapping complex, [4–7] after which the mRNA body is degraded by the exoribonuclease Xrn1 [8]. In the 3' to 5' pathway, the mRNA body is processively hydrolyzed by the 10-component cytoplasmic exosome complex. Structurally, this Exo-10 complex comprises the catalytically inactive Exo-9 core and the Rrp44 enzyme that harbors both endonucleolytic and exonucleolytic activities. The exosome forms a complex with the scavenger decapping enzyme (DcpS; Dcs1p in yeast) enzyme [9] that removes the cap structure of the short mRNA fragment that is released by the exosome [10,11].

The apo DcpS enzyme assembles into a symmetric homodimer that contains dimeric N- and C-terminal domains that are linked by a flexible hinge region (Figure 1A) [12–14]. Due to its dimeric nature, the enzyme possesses two bipartite active sites that are located at the two interfaces between the N- and C-terminal domains. Upon substrate binding to the C-terminal part of one of the active sites, the N-terminal domain flips over to enclose the mRNA substrate in a well-defined pocket. This motion simultaneously opens the other substrate binding site. Within the substrate loaded closed active site, hydrolysis of the bond between the alpha and beta phosphate of the triphosphate linkage is performed by the catalytic HIT motif, which produces m^7GMP and diphosphorylated RNA.

Subsequently, the enzyme undergoes a see-saw motion to simultaneously release the products and to capture a next substrate in the other active site.

The DcpS enzyme should not be active on long mRNA substrates, as this would interfere with the integrity of mRNA transcripts that are still actively translated. In agreement with that, it has been reported that the activity of DcpS is significantly reduced if the length of the mRNA body is increased from 1 to 10 nucleotides. [14]. This suggests that there is a narrow length range for RNAs to be an eligible substrate for DcpS and that the decapping of long RNA is prevented. Here, we determined the properties and the structural basis for the substrate preference in the DcpS enzyme. In summary, we observe that DcpS is only able to efficiently digest capped mRNA substrates that have an RNA body of less than three nucleotides. Based on X-ray crystallography and methyl TROSY NMR spectroscopy [15,16] we find that this molecular ruler results from the inability of the enzyme to form a closed and catalytically competent active site around longer mRNA substrates due to steric clashes between the third base in the mRNA body and parts of the DcpS enzyme.

Results

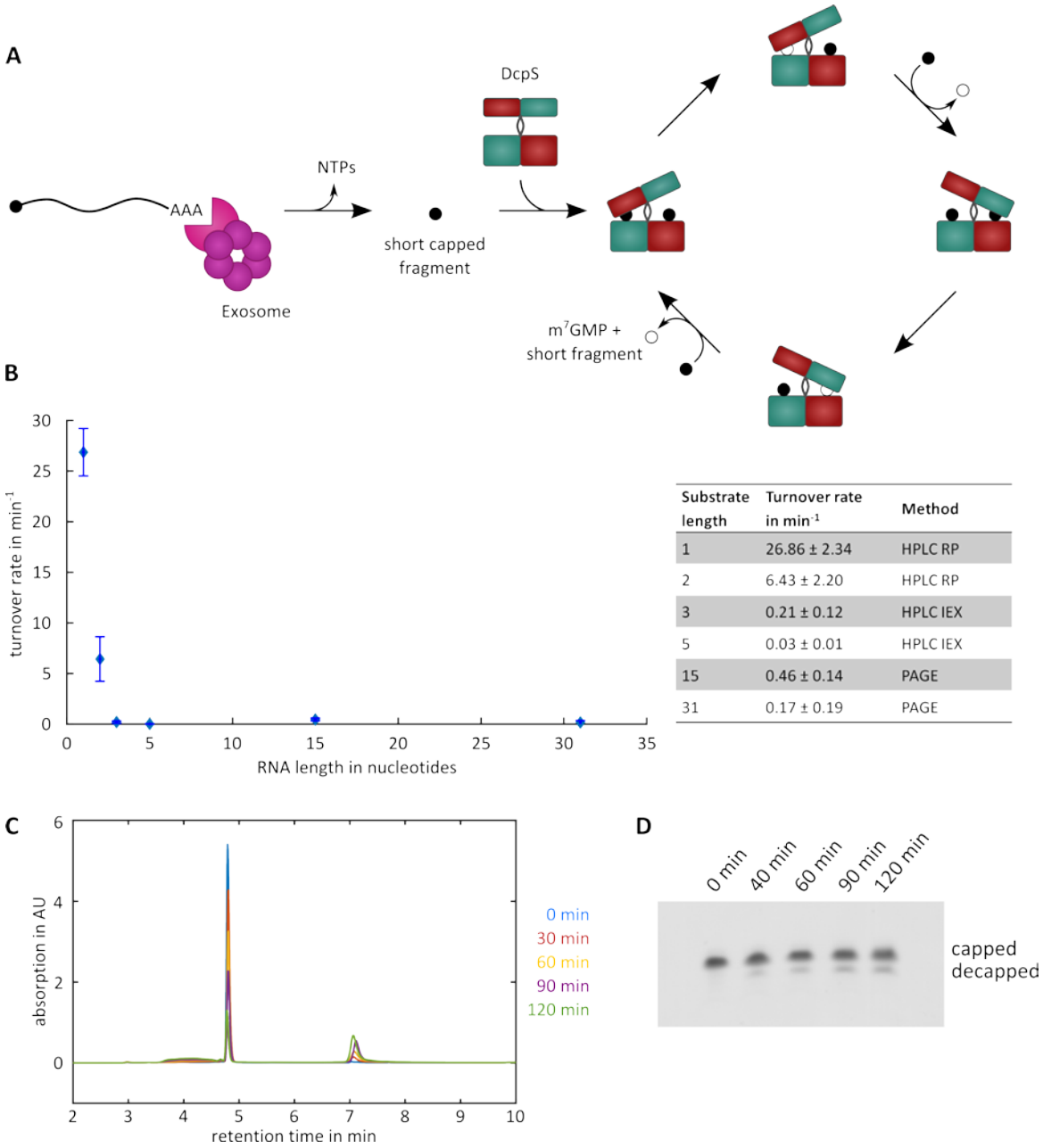
DcpS activity is inverse proportional to substrate length

To determine the relationship between DcpS activity and substrate length we made use of our recently developed method to prepare capped mRNA fragments of any length and sequence [17]. Activity assays of DcpS were performed with homogeneous, capped RNA and accurate turnover rates were determined (see Methods, [18]). We ensured that the activity of the enzyme was in the linear regime and that the enzyme activity was not influenced by substrate or product inhibition [11].

Based on these assays, we found that the activity of yeast DcpS on a capped RNA with a single nucleotide body is 26.9 ± 2.3 min^{-1} which is in agreement with previous reports (Figure 1C) [11]. For mRNA species that have an mRNA body that is 31 nucleotides, we extracted a turnover rate of 0.1 min^{-1}. From that we conclude that DcpS indeed has an internal molecular ruler and that long mRNA substrates are only processed with very low efficiency.

To determine the substrate length at which the catalytic efficiency of DcpS is reduced, we repeated the above experiments with substrates that have an mRNA body of 2, 3, 5, and 15 nucleotides. Interestingly, we found that DcpS activity already significantly drops upon going from a capped mono-nucleotide (26.9 ±2.3 min^{-1}) to a capped di-nucleotide (6.4 ± 2.2 min^{-1}). This trend continued with a capped tri-nucleotide, that showed an activity reduced to 0.1 min^{-1}. Interestingly, all mRNA substrates with an mRNA body longer than three nucleotides that we tested are processed by DcpS with the same low basal rate between 0.1 and 0.01 min^{-1} (Figure 1B-D).

In summary, we find that DcpS activity drops upon increasing the length of the mRNA substrate body from 1 to 2 to 3 or more nucleotides. It should be noted that DcpS is able to hydrolyze mRNA substrates that are three or more nucleotides long, but only in a highly inefficient manner. These results are qualitatively in agreement with published data that describe a decrease in DcpS activity with increasing substrate length [14]. However quantitatively, our data show that the range in length for eligible substrates is much tighter than was previously suggested, with a cutoff at two nucleotides.



Substrate length	Turnover rate in min^{-1}	Method
1	26.86 ± 2.34	HPLC RP
2	6.43 ± 2.20	HPLC RP
3	0.21 ± 0.12	HPLC IEX
5	0.03 ± 0.01	HPLC IEX
15	0.46 ± 0.14	PAGE
31	0.17 ± 0.19	PAGE

Figure 1: Substrate length dependency of DcpS turnover rates. A: mRNA is hydrolyzed by the exosome complex, leaving short capped RNA fragments, which are decapped by DcpS. The homodimeric DcpS enzyme has two active sites located between the N- and C-terminal domains. The substrate is enclosed between the domains by a conformational change from the symmetric to the asymmetric protein. After catalysis in one active site, the N-terminal domain flips to close the second active site and release the products m^7GMP and the decapped RNA fragment from the first active site. **B:** Activity assays of DcpS with capped RNA substrates of different length show a drastic drop in activity for substrates longer than two nucleotides. Analysis of the activity assays was performed with different methods depending on the substrate length (HPLC, high pressure liquid chromatography; RP, reverse phase chromatography; IEX, ion exchange chromatography; PAGE, polyacrylamide gel electrophoresis) **C:** HPLC-RP chromatogram of activity assay of DcpS with capped mononucleotide (ratio 1:10,000) **D:** Urea-PAGE of activity assay of DcpS with capped 15-nt RNA (ratio 1:200).

Longer mRNAs prevent the formation of an active DcpS conformation

The apo DcpS enzyme is a symmetric dimer in solution leading to identical chemical shifts of both protomers in NMR. Thus, we observe a single set of resonances in methyl TROSY NMR spectra for the dimeric enzyme (Figure 2A, black). When DcpS forms a catalytically competent state upon substrate binding, the symmetry in the enzyme is broken as the substrate is tightly embedded in one of the two active sites [11]. This structural change is observable in NMR spectra, as each protomer now gives rise to a unique resonance for all affected residues. Therefore, the formation of a functional active site in DcpS results in splitting of the NMR resonances that result from residues at the domain interface.

In agreement with our previous studies, we observe that DcpS adopts a catalytically active conformation upon interaction with a cap analogue, a capped substrate that has an RNA body of a single nucleotide (m^7GpppG) (Figure 2A, red). Here, we observe that the DcpS enzyme also adopts the catalytically active conformation upon interaction with a capped dinucleotide (m^7GpppGU). This is in agreement with our activity assays, where we showed that the enzyme is active on this substrate. On the contrary, DcpS does not form a stably closed active site upon interaction with a capped trinucleotide (m^7GpppGGU) (Figure 2B and Supplementary Figure 1). This observation is in full agreement with our activity assays that show that the DcpS enzyme is not able to efficiently remove the cap structure of this mRNA. It is important to note that the DcpS enzyme interacts with the capped trinucleotide, as we observe clear chemical shift perturbations (CSPs) in the enzyme. These CSPs resemble the interaction between the substrate and the open binding site in the enzyme. Our data thus shows that a capped trinucleotide is able to interact with DcpS, but that the enzyme is not able to properly close around the cap structure. In summary, we thus show that DcpS is only active on mRNA substrates that are compatible with the formation of a closed active site.

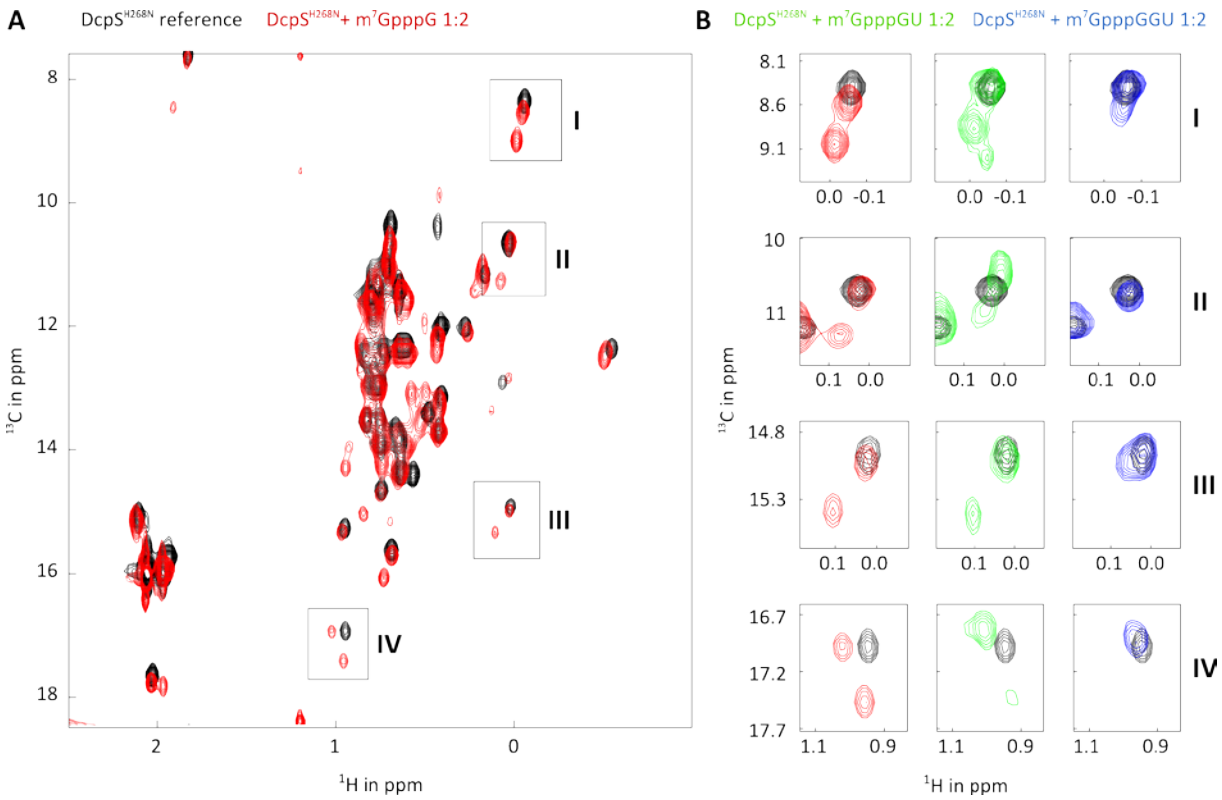


Figure 2: Substrates longer than two nucleotides do not induce a stable asymmetric conformation of DcpS. **A:** Methyl-TROSY-NMR spectra of apo DcpS (black) and with m^7GpppG (red). Peak splitting upon titration with m^7GpppG indicates the conformational change from the symmetric apo enzyme to the asymmetric substrate-bound enzyme. **B:** Zoom on peaks indicating the asymmetric conformation of DcpS (as in **A**). Titration with capped dinucleotide (green) induces the asymmetric conformation, while titration with capped trinucleotide (blue) retains the symmetric conformation.

The third nucleotide in the mRNA substrate induces steric clashes with the closed active site.

To obtain insights into the mechanism that prevents DcpS from forming a productive active site around mRNA substrates longer than two nucleotides, we determined the crystal structure of an inactive DcpS variant ($DcpS^{H268N}$) in complex with a capped dinucleotide. The complex crystallized in space group $P2_12_12_1$, with two dimers per asymmetric unit. The structure was determined by molecular replacement [19] with the structure of yeast enzyme in complex with m^7GDP (5BV3; [11]) as a search model. The final model was refined to 2.9 Å resolution with good geometry (Supplementary Table 1). In this structure, DcpS is in the asymmetric conformation, with the N-terminal lid domain flipped over to tightly embed the substrate in the closed active site (Figure 3A). In agreement with our activity assays, the inactive catalytic triad is properly oriented such that the triphosphate linkage

can be hydrolyzed. The methylated guanine cap structure is tightly embedded in the active site and stabilized by stacking interaction with Trp 161. Importantly, the structure also displays clear density for the mRNA body of the capped dinucleotide that fully occupies the active site cavity (Figure 3B). This allows us to follow the path of the mRNA inside the closed active site. The first nucleotide of the RNA body has no specific interaction partners, whereas the second nucleotide points further away from the cap structure, towards a narrow opening between the N- and C-terminal domains. This region has previously been proposed as a possible RNA exit channel [12]. Based on the path of the mRNA body, a third nucleotide would be placed inside this proposed exit tunnel. This tunnel has a diameter of ~8 Å and is thus too small to accommodate the third nucleotide. This is confirmed by modeling of a capped trinucleotide into the closed active site (Supplementary Figure 2) that shows that the vdW radii of the third nucleotide and the exit tunnel significantly overlap. From the DcpS side, these steric clashes are caused by sidechains in the loop opposing the exit channel and by side chains and backbone atoms of residues at the outer part of the channel. Significant conformational changes in the DcpS exit channel are thus required to accommodate a third nucleotide in the closed active site of DcpS. The energy that is associated with these conformational changes is likely too high to allow for effective catalysis of substrates longer than two nucleotides. This is in agreement with our NMR data that show that the stable asymmetric conformation in the enzyme is only formed upon the recruitment of substrates that are up to two nucleotides in length.

The length of the DcpS exit channel is short. This implies that nucleotides that follow the third nucleotide will no longer interfere with the DcpS closing mechanism. This is in agreement with our activity assays that show equal activity for mRNA substrates that have an RNA body of 3 nucleotides and those that are longer.

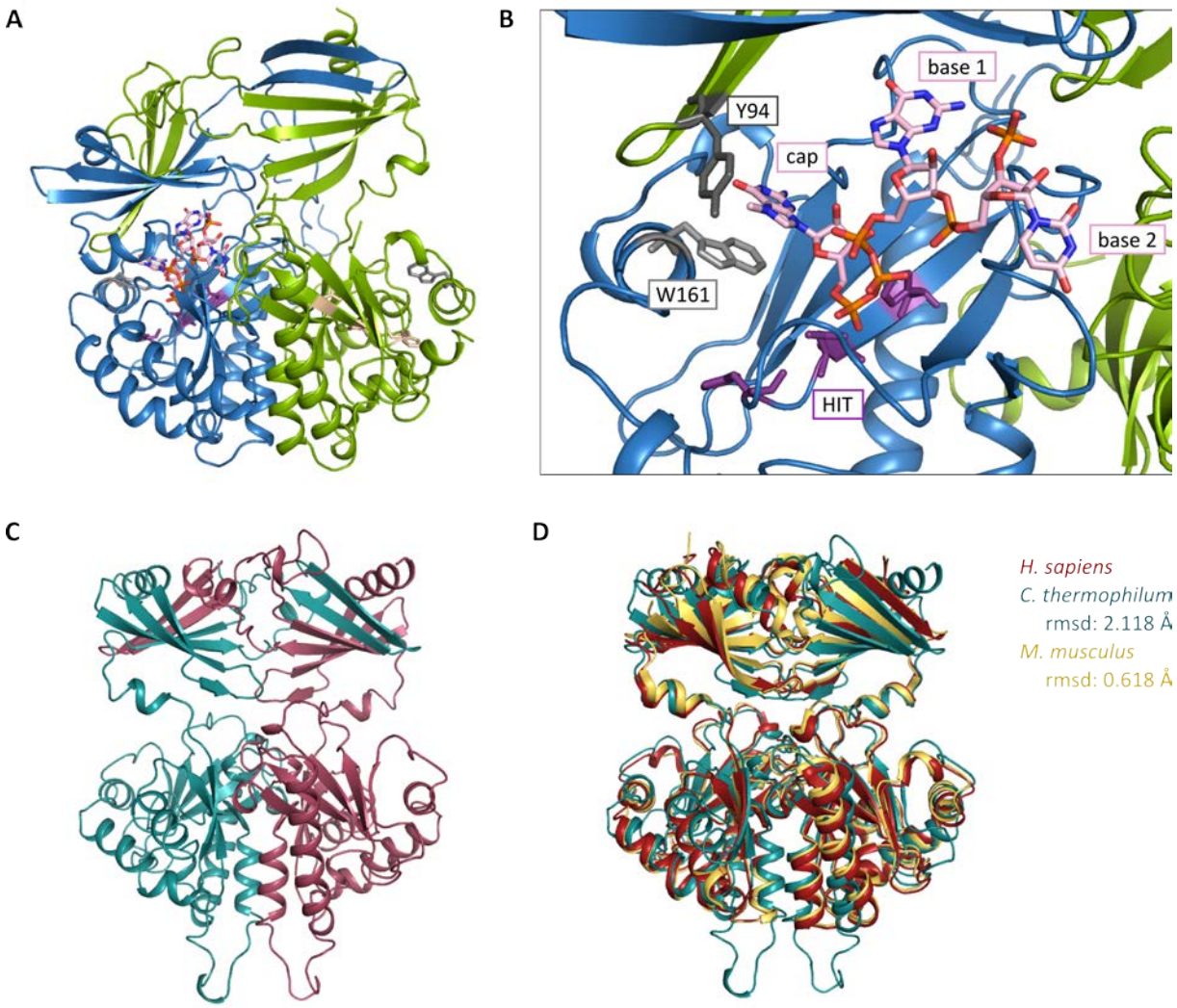


Figure 3: The threshold in substrate usage is based on the structure on DcpS and conserved among species. **A:** The structure of *S.c.* DcpS in complex with a capped dinucleotide is in the asymmetric conformation with substrate bound only in the closed active site. **B:** The cap guanine is bound in the active site of DcpS by stacking interaction with W161 and interaction of Y94 with the N7-methylation. The nucleotides of the RNA body show no specific interactions and the second base points towards the tunnel formed by N- and C-terminal domain leading towards the exterior of the enzyme. **C:** Structure of *C.t.* DcpS in the apo form. The enzyme crystallized in the symmetric conformation. **D:** DcpS of *C. thermophilum*, *H. sapiens*, *M. musculus* and *S. cerevisiae* (not shown) are structurally conserved.

The length sensing mechanism of DcpS is conserved among species

Above, we have focused on the DcpS enzyme from *S. cerevisiae* (Dcs1p) and identified the molecular basis by which the enzyme is able to selectively act on short mRNA substrates. Next, we aimed at determining if this preference in substrate usage is also present in the enzymes from *C. thermophilum* and humans. To that end, we determined the structure of *C.t.* DcpS in the absence of substrate (Figure 3C). The protein crystallized in space group P1 with one dimer per asymmetric unit, the

structure was refined to a resolution of 1.95 Å and displays good geometry. Apo *C.t.* DcpS forms a symmetric dimer with the same overall fold as the structure of the human and mouse protein (Figure 3D). To test whether the *C. thermophilum* and human proteins undergo the same conformational change as *S.c.* DcpS, we recorded methyl TROSY NMR spectra of catalytically inactive variants (C.t. DcpS H258N and H.s. DcpS H277N). These clearly show a single set of peaks for the dimeric apo enzymes which confirms the findings from static crystal structures. In summary, we conclude that the structure and conformation of the apo enzyme is well conserved and symmetric.

Next, we added mRNA substrates of increasing length to monitor the formation of the closed active site in the *C. thermophilum* and human derived enzymes by NMR. As for the *S. cerevisiae* enzyme, the addition of mRNA with a body of 1 or 2 nucleotides results in the formation of a closed active site for these enzymes. On the contrary, an mRNA substrate of three nucleotides does not result in the closed conformation (Supplementary Figure 3). From these data, we conclude that the threshold in substrate usage that we identified in the *S. cerevisiae* protein is fully conserved in the *C. thermophilum* and human enzymes. This shows that the mechanism for the preference in substrate usage by DcpS is conserved among species.

Discussion

In many biological pathways a set of enzymes acts successively. This e.g. applies to signaling pathways, where kinases are activated in an ordered successive manner or to metabolic pathways, where the product of one enzymatic reaction is the substrate of the next enzyme. Likewise, mRNA degradation takes place in a regulated manner, where the removal of the mRNA 5' cap structure by the DcpS enzyme takes places only after the exosome has degraded the mRNA body in the 3' to 5' direction. Here, we have addressed how the DcpS enzyme is able to determine to what degree the mRNA body has been shortened by the exosome complex.

Previous data by us and by others have shown that the scavenger decapping enzyme needs to undergo a large conformational change to form a catalytically competent active site. Based on methyl TROSY NMR data (Figure 2B), we here show that this active site cannot be formed when the substrates has an mRNA body of three or more nucleotides. Structurally, this inability of DcpS to properly close around the mRNA cap structure is caused by steric clashes between the third base of the substrate and amino acids in both the N- and C- terminal domains of the enzyme (Supplementary Figure 2).

Functionally, we have shown that the activity of the DcpS enzyme drops by two orders of magnitude when the length of the mRNA substrate increases from two or fewer to three or more nucleotides. From these data, we conclude that DcpS has a very strict molecular ruler that prevents long mRNA substrates to be processed by DcpS, whereas efficient turnover of short mRNA fragments such as those produced by the exosome can be efficiently degraded.

Currently, the lengths of the fragments that are produced by the exosome complex have not been accurately determined. However, published data suggest that the exosome products are shorter than four to five nucleotides [20–23]. It is tempting to speculate that the exosome produces mRNA fragments

with one or two nucleotides as these can then be directly handed over to DcpS for further processing. Further *in vitro* and *in vivo* experiments will need to reveal if a direct handover of exosome products to DcpS is functionally possible, or if mRNA fragments that are produced by the exosome, need to undergo further processing before DcpS can accept them as substrates.

In summary, our study shows how molecular mechanisms ensure that the DcpS enzyme is only fully active on short mRNA fragments that are produced by the 3’ to 5’ mRNA degradation pathway. This clearly displays that biological pathways have evolved in a way that ensures the correct and successive action of the involved enzymes. As the DcpS molecular ruler is conserved from yeast to humans (Figure 3D), we propose that the regulation of the 3’ to 5’ mRNA decay pathway is also conserved.

Methods

Protein preparation

BL21 (DE3) *E. coli* cells were transformed with a plasmid that codes for a given DcpS variant (see Table S1) with a TEV cleavable N-terminal His_6-tag or His_6-NusA-His_6-tag. For activity assays and crystallization, cells were grown in LB medium at 37 °C to OD_{600}=0.6. Subsequently, protein expression was induced with 0.5 mM IPTG and cells were shifted to 20 °C for 16 hours. Cells were harvested by centrifugation and resuspended in buffer A (50 mM NaPO4, 150 mM NaCl, 5 mM Imidazole) supplemented with 0.1 % Triton X-100, 1 mM EDTA and lysozyme. After lysis by sonication, 2 mM $MgSO_4$ was added and insoluble cell debris was removed by centrifugation, after which the supernatant was applied to Ni-NTA resin that was equilibrated in buffer A. The resin was washed with buffer A and the protein was eluted with buffer B (buffer A supplemented with 200 mM Imidazole). The eluted protein was supplemented with 0.25 mg TEV protease and simultaneously dialysed against 25 mM Tris pH 8.0, 75 mM NaCl, 1 mM DTT at 4 °C. The cleaved affinity-tag and TEV protease were removed by reverse Ni-affinity chromatography, after which the DcpS protein was subjected to size exclusion chromatography based on a Sephadex S200 column (*GE Healthcare*) in 25 mM Hepes pH 8.0, 25 mM NaCl.

RNA preparation

RNAs were prepared by *in vitro* transcription with T7 RNA polymerase and capped with vaccinia capping enzyme as described [17]. In brief, equimolar amounts of an anti-sense strand primer that contains the T7 promoter sequence and a sense strand primer that contains the (reverse complementary) target RNA sequence and the T7 promoter (in reverse compliment) were mixed at 1 µM concentration in 40 mM Tris pH 8, 5 mM DTT, 1 mM spermidine, 0.01 % Triton X-100, 4 mM NTPs, 20-60 µM $MgCl_2$ and 0.2 µM T7 RNA polymerase. Transcription reactions were incubated at 37 °C for 4 hours, after which insoluble pyrophosphate was removed by the addition of EDTA. The RNA product was precipitated at -20 °C in 0.3 M NaOAc pH 5.2 and 0.7 volumes of isopropanol or in

0.2 M NaCl and 3.5 volumes of EtOH (RNA below 10 nt), followed by centrifugation at 9000 g for 1 hour. The RNA pellet was washed with cold 75 % ethanol, dried and resuspended in buffer.

In vitro transcribed RNA was purified on an FPLC system using anion exchange chromatography (22 × 250 mm DNAPac PA100 column; Dionex). The RNA was applied onto a heated column (80 °C) that was equilibrated in 20 mM Tris pH 8.0 and 5 M Urea and subsequently eluted with an NaCl gradient. The RNA containing fractions were pooled, precipitated (see above), desalted using a PD-10 column (GE Healthcare) and concentrated.

Capping of the mRNA body

To add a 5' cap structure to the RNA, the purified RNA was dissolved in 50 mM Tris, pH 8, 5 mM KCl, 1 mM $MgCl_2$ and 1 mM DTT with 20 µM RNA. 0.5 mM GTP, 0.1 mM SAM, after which the capping enzyme was added [17]. After incubation at 37 °C the reaction was stopped by heating to 70°C for 10 minutes and subsequently centrifuged for 10 minutes at 4500 g to remove the insoluble precipitates. The capped RNA was subsequently precipitated as described above and resuspended in an appropriate buffer.

RNase A cleavage

Short capped RNAs were prepared by cleaving a longer capped RNA that contains, apart from a single U base, only GA bases. To that end 50 ng of RNase A was added per nmol of RNA, followed by incubation at 37 °C for 10 min. Subsequently, the RNase A enzyme was removed through a phenol-chloroform extraction and the 5' and 3' ends of the RNA were separated based on an ion exchange chromatography as described above.

DcpS activity

Decapping reactions were performed in 25 mM Hepes pH 8.0, 25 mM NaCl using a total volume of 110 µl. The total RNA concentration was 20 to 200 µM, to which the DcpS enzyme was added at a ratio of 1:200 to 1:10.000, depending on the capping efficiency with the respective mRNA substrate. Reactions were incubated at 37 °C for 2 hours and 20 µL samples for analysis were taken every 30 minutes. The decapping reaction in the samples was stopped by the addition of TFA to a final concentration of 0.1 % (v/v).

For reactions with mRNA substrates that contained an mRNA body of 1 or 2 nucleotides, the samples were analyzed with reverse chromatography on a C18 column (*Macherey Nagel*). The column was equilibrated in 25 mM NH_4Ac pH 5.3 at 40 °C and a gradient from 0 to 60% acetonitril was used to separate the reaction products. For reactions with mRNA substrates that contained an mRNA body between 3 and 5 nucleotides, the samples were analyzed by anion exchange chromatography on a DNAPac PA200 RS (*Dionex*). The column was equilibrated in 20 mM Tris pH 8.0 at 30 °C and the reaction products were eluted based on a NaCl gradient from 0 to 0.5 M and 0.3 to 0.8 M for 3 nt and 5 nt RNA respectively. For reactions with mRNA substrates that have an mRNA body over 10 nucleotides, the samples were analyzed on 20 cm long 16% denaturing PAA gels. To visualize the RNA, the gels were stained with methyleneblue and quantified using GelBandFitter (www.gelbandfitter.org) in Matlab.

The reported enzyme activities are the average of at least three fully independent reaction series, with enzymes that were independently expressed and purified. This ensured that our decapping rates are highly accurate.

NMR spectroscopy

DcpS samples for NMR spectroscopy were prepared by overexpression of the catalytically inactive protein in D_2O based minimal medium that contained $^{2}H^{12}C$ glucose as the sole carbon source. Methyl groups were labeled by the addition of 60 mg per liter of α-ketobutyric acid (methyl $^{13}CH_3$ labeled)

and 100 mg per liter of methionine (1H, 13C labeled) one hour prior to induction. The expression and purification of these proteins was performed as described above. NMR samples contained between 50 and 500 µM protein in 100% D_2O, 25 mM Hepes pH 8.0 and 25 mM NaCl. Methyl TROSY NMR spectra were recorded on Bruker Avance 600 and 800 MHz spectrometers at 20 °C or 25 °C. Typical NMR spectra were recorded with a total carbon chemical shift evolution time of 25 ms. NMR data were processed using the NMRpipe program-suit [24].

Protein Crystallization

The 1:1 complex of the DcpS enzyme from *S. cerevisiae* (residues 8-350; H268N) with m^7GpppGU RNA was crystallized at a concentration of 5 mg/ml in 10 mM Hepes pH 7.5, 100 mM NaCl, 1.6 M NH_4SO_4. DcpS from *C. thermophilum* was crystallized at a concentration of 13 mg/ml in 100 mM BisTRIS pH 5.5, 200 mM NH_4Ac, 25 % PEG 3350. Crystals were flash frozen in 30 % glycerol and diffraction data were collected at the beamline PXII of the Swiss Light Source (PSI, Villigen Switzerland) at a wavelength of 1 Å at 100 K. Data was processed using XDS [25] and the structures were determined by molecular replacement using Phaser [19]. For the structure of DcpS in complex with m^7GpppGU, the structure with PDB ID 5BV3 was used as a search model. For the structure of the *C. thermophilum* enzyme, the structure with PDB ID 1XMN was used. Iterative model building was performed using the program Coot [26], the structure refinement was performed using the programs Refmac [27] and Phenix [28]. All figures showing structural data were prepared using Pymol (www.pymol.org).

Acknowledgements

We thank Iris Holdermann, Janina Peters, Mira Schütz-Stoffregen and Johanna Stöfl for excellent technical assistance, members of the group for discussion and the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) for measurement time. This work was supported by the Max Planck Society, the Regensburg University and the Deutsche Forschungsgemeinschaft (SFB 960 TP/B12).

Author Contributions

A-L.F., A.N. and R.S. designed and performed experiments and analyzed data. A-L.F and R.S. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 6GBS (Apo-DcpS from *C. thermophilum*) and (to be submitted) (DcpS:m^7GpppGU complex from *S. cerevisiae*).

References

1. Wilusz, C. J., Wormington, M. & Peltz, S. W. The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* **2,** 237–246 (2001).
2. Parker, R. RNA degradation in Saccharomyces cerevisae. *Genetics* **191,** 671–702 (2012).
3. Parker, R. & Song, H. The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.* **11,** 121–127 (2004).
4. Van Dijk, E. *et al.* Human Dcp2: A catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* **21,** 6915–6924 (2002).
5. Wang, Z., Jiao, X., Carr-Schmid, A. & Kiledjian, M. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12663–12668 (2002).
6. Lykke-Andersen, J. Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay. **22,** 8114–8121 (2002).
7. Wurm, J. P., Holdermann, I., Overbeck, J. H., Mayer, P. H. O. & Sprangers, R. Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc. Natl. Acad. Sci.* **114,** 6034–6039 (2017).
8. Hsu, C. L. & Stevens, a. 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 (1993).
9. Wang, Z. & Kiledjian, M. Functional link between the mammalian exosome and mRNA decapping. *Cell* **107,** 751–762 (2001).
10. Nuss, D. & Furuichi, Y. Characterization of the m7G(5')pppN-pyrophosphatase Activity from HeLa Cells. *J. Biol. Chem.* **252,** 2815–2821 (1977).
11. Neu, A., Neu, U., Fuchs, A.-L., Schlager, B. & Sprangers, R. An excess of catalytically required motions inhibits the scavenger decapping enzyme. *Nat. Chem. Biol.* **11,** (2015).
12. Gu, M. *et al.* Insights into the structure, mechanism, and regulation of scavenger mRNA

decapping activity. *Mol. Cell* **14,** 67–80 (2004).

13. Chen, N., Walsh, M. a., Liu, Y., Parker, R. & Song, H. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J. Mol. Biol.* **347,** 707–718 (2005).
14. Liu, H., Rodgers, N. D., Jiao, X. & Kiledjian, M. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J.* **21,** 4699–4708 (2002).
15. Pervushin, K. V., Wider, G., Riek, R. & Wüthrich, K. The 3D NOESY- [1 H , 15 N , 1 H] -ZQ-TROSY NMR experiment with diagonal peak suppression. *PNAS* **96,** 9607–9612 (1999).
16. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-correlated relaxation enhanced 1H-13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125,** 10420–8 (2003).
17. Fuchs, A.-L., Neu, A. & Sprangers, R. A general method for rapid and cost-efficient large-scale production of 5’ capped RNA. *RNA* 1454–1466 (2016). doi:10.1261/rna.056614.116
18. Audin, M. J. C., Wurm, J. P., Cvetkovic, M. A. & Sprangers, R. The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation. *Nucleic Acids Res.* 1–12 (2016). doi:10.1093/nar/gkw062
19. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
20. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. *Cell* **127,** 1223–1237 (2006).
21. Wasmuth, E. V., Zinder, J. C., Zattas, D., Das, M. & Lima, C. D. Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *Elife* **6,** 1–24 (2017).
22. Zinder, J. C., Wasmuth, E. V. & Lima, C. D. Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. *Mol. Cell* **64,** 734–745 (2016).
23. Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A. & Conti, E. Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the

RNase II Nuclease Family. *Mol. Cell* **29,** 717–728 (2008).

24. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

25. Kabsch, W. *XDS*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

27. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).

28. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

Supplementary Information

Molecular basis for mRNA length sensing by the DcpS decapping enzyme

Anna-Lisa Fuchs[1], Ancilla Neu[2], Remco Sprangers[1,*]

[1] Department of Biophysics I, University of Regensburg, 93053 Regensburg, Germany.

[2] Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany.

* Correspondence should be addressed to remco.sprangers@ur.de

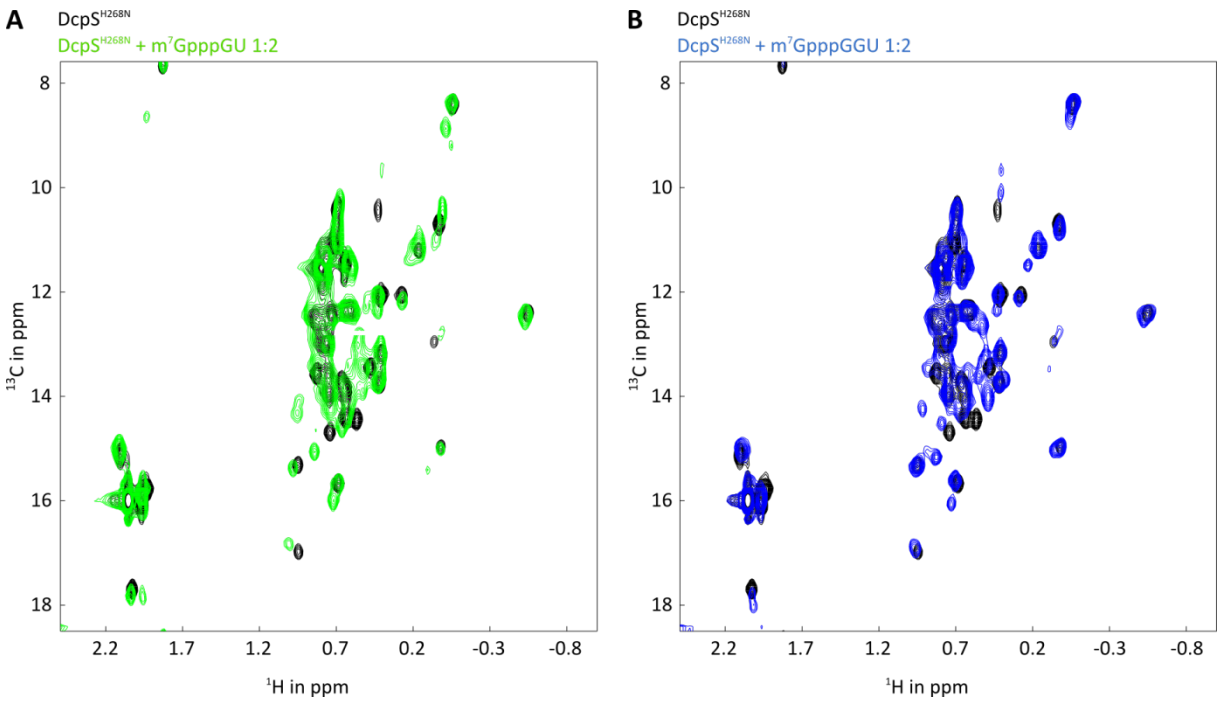
Keywords:

Enzyme regulation, mRNA decay, NMR spectroscopy, conformational changes, scavenger decapping enzyme, decapping.

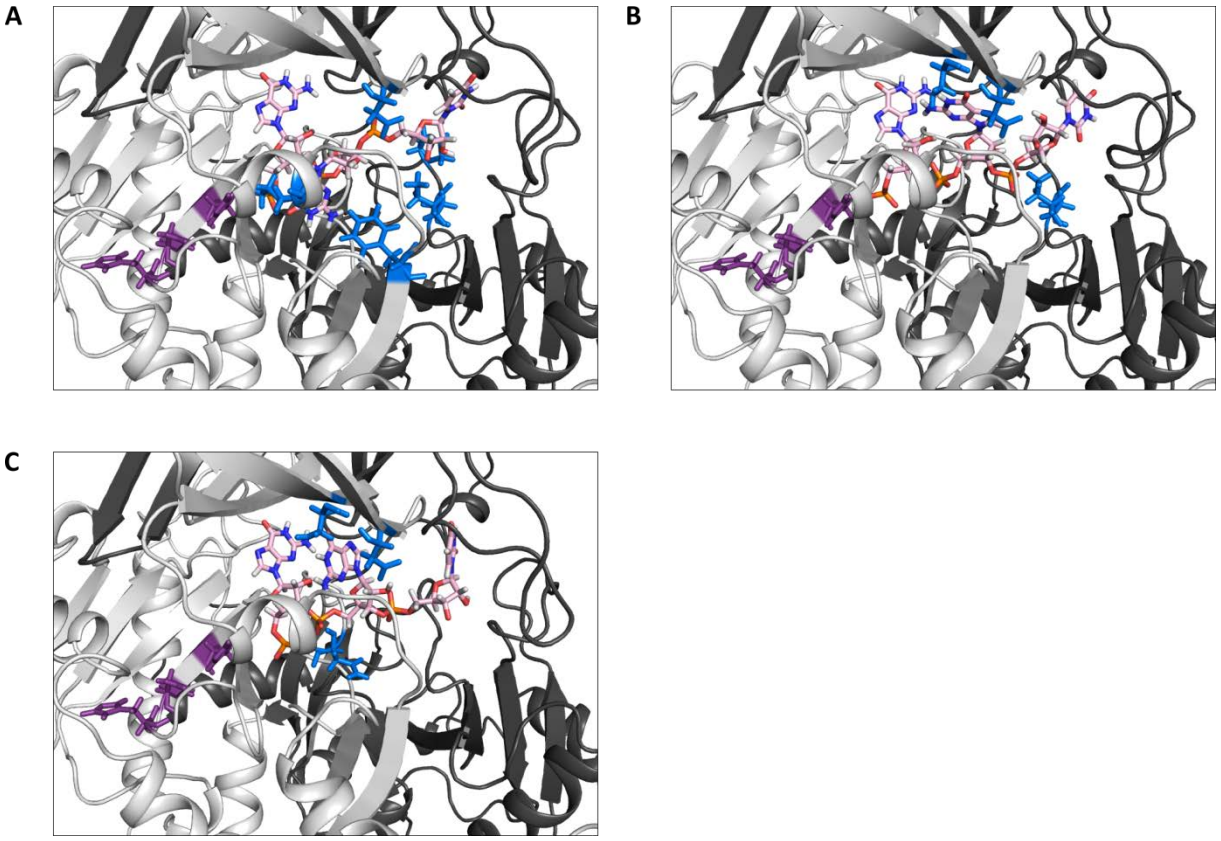
	***S.c.* DcpS^{H268N} + m^{7}GpppGU**	***C.t.* DcpS^{H258N} apo (6GBS)**
Data collection		
space group	P 21 21 21	P 1
cell dimensions		
a, b, c (Å)	87.76, 104,10, 189,22	52.59, 69.72, 70.95
α, β, γ (°)	90, 90, 90	104.48, 101.39, 111.61
resolution (Å)	47.31 (2.94)	48.50 (1.95)
completeness (%)	99.67 (97.45)	96.6 (94.1)
Refinement		
resolution (Å)	2.94	1.946
No. reflections	37376	115494
R_{work} / R_{free}		0.1828 / 0.2165
No. atoms		
protein		5402
ligand / ion		55
waters		257
B-factors		
protein		33.87
ligand / ion		57.09
water		34.95
r.m.s. deviations		
bond lengths (Å)		0.007
bond angles (%)		0.87

Supplementary Table 1: Statistics of crystal structures of *S.c.* DcpS^{H268N} in complex with m^{7}GpppGU (left column) and *C.t.* DcpS^{H258N} in the apo form. Values in parentheses are for highest-resolution shell.

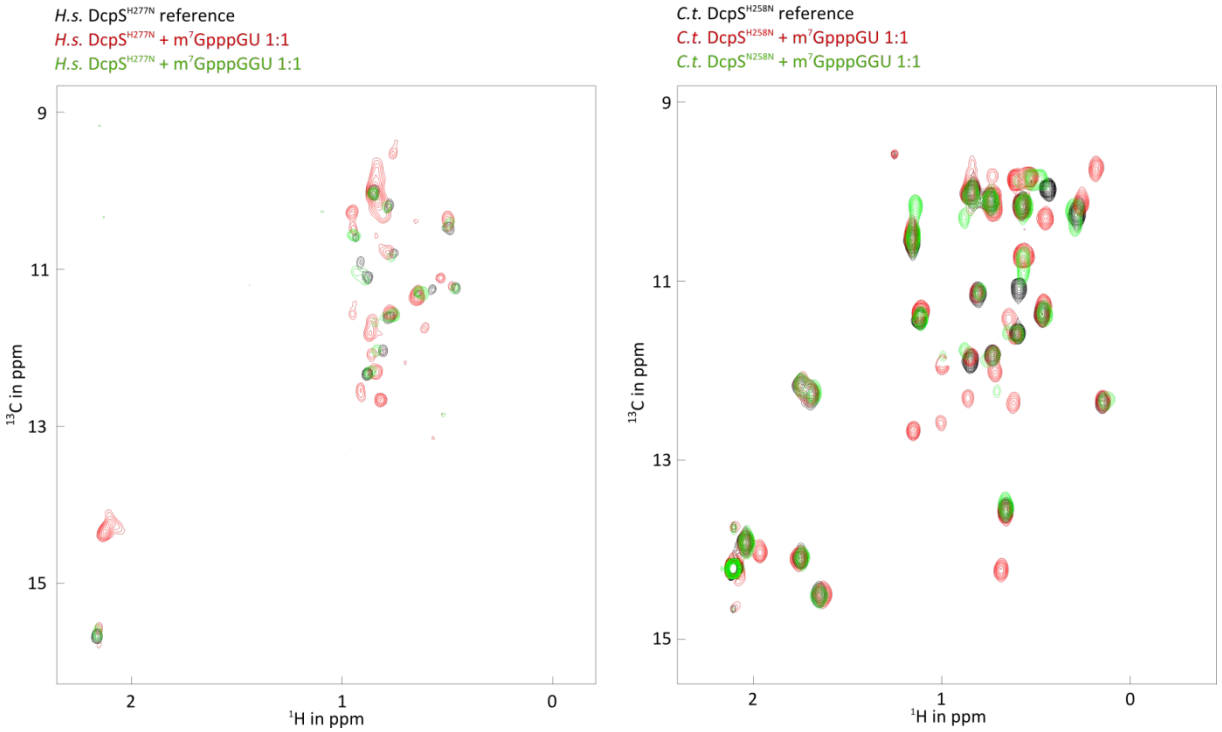
(The structure of *S.c.* DcpS in complex with m^{7}GpppGU is not yet validated by the PDB, so no final statistic is available.)



Supplementary Figure 1: Methyl-TROSY NMR spectra of *S.c.* DcpS with capped di- and tri-nucleotide. A: Titration of DcpS^{H268N} with m^{7}GpppGU at a ratio of 1:2. Peak splitting indicates the asymmetric conformation of the enzyme. **B:** Titration of DcpS^{H268N} with m^{7}GpppGGU at a ratio of 1:2. No stable asymmetric conformation can be detected.



Supplementary Figure 2: Modelling of a third nucleotide into the active site of DcpS. A-C: Three best results of modelling of a 3 nt-RNA into the active site of DcpS based on the structure of DcpS in complex with m^7GpppGU. Modelling was performed with the programme *CYANA.* RNA in pink, active site residues in purple, residues with overlapping vdW-radii with RNA in blue.



Supplementary Figure 3: methyl-TROSY NMR spectra of titrations DcpS with capped di- and tri-nucleotide. A: Titration of *H.s.* DcpS^{H277N} with m^{7}GpppGU and m^{7}GpppGGU. Capped dinucleotide induces the asymmetric conformation, capped trinucleotide does not induce the asymmetric conformation. **B:** Titration of *C.t.* DcpS^{H258N} with m^{7}GpppGU and m^{7}GpppGGU. Capped dinucleotide induces the asymmetric conformation, capped trinucleotide does not induce the asymmetric conformation.

Acknowledgements

This thesis would not have been possible without the support of many people, to whom I am exceedingly grateful:

Remco – Thank you for giving me the opportunity to work in your lab and on this great project that so nicely fell into place. Thank you for your support, your help, your understanding and the fact that you are a great boss.

Prof. Ralf Jansen, Prof. Dirk Schwarzer and Prof. Thorsten Stafforst – Thank you for agreeing to examine this thesis and for your many helpful and encouraging remarks, suggestions, and ideas in my TAC meetings.

The EU and SFB 960 – This project would not have been possible without the funding by the EU and the DFG - SFB960.

Sprangers and Wiesners – Thank you to all former and current members of the Sprangers and Wiesner groups. It was always fun to work with you and I hope we will always maintain the collegial atmosphere in our lab. Thank you for lots of discussions, mutual help in little lab tasks, and of course many, many cakes. It will be hard to find colleagues as great as you are.

Ancilla – I can only repeat myself: It has been a capital pleasure to work with you on this project and in general. I learnt so much from you. Thank you for a wonderful time in the lab and outside and many cups of tea!

Our amazing technicians – Iris, Johanna, Mira, Nina and Samira, thank you for keeping the lab running and reminding us to clean up our daily mess.

The 3rd floor – Thank you for the nice, relaxed, familial, crazy and friendly atmosphere, especially with the plant people. I miss you guys.

Mira and Stefan – Over the last years you became more than just colleagues. Thank you for many comforting talks.

Becka – Thank you for your friendship, our long-distance relationship and many talks in all situations.

My parents – Thank you for your constant support and encouragement, believing in me and my abilities, and never putting pressure on me.

André – Thank you for always being there for me and taking care of me, in all situations. I don't know what I would do without you!