Structure, dynamics and function of proteins in the mRNA decay pathway

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Diplom Biochemikerin Ancilla Neu

aus Sulz am Neckar

Tübingen

2015

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

Tag der mündlichen Qualifikation:	25.02.2015
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Dr. Remco Sprangers
2. Berichterstatter:	Prof Thilo Stehle

Summary

RNA decay is an important mechanism in regulating steady state mRNA levels and thus in orchestrating gene expression to match the changing requirements of a cell. This study is concerned with the molecular mechanisms of three processes in the two branches of mRNA decay: Firstly, the interactions that give rise to phase separation leading to processing body (P-body) formation are studied. Secondly, we investigate the assembly pathway of the heteroheptameric LSm1-7 complex that interacts with the 3' end of mRNAs in preparation of 5'-3' decay. The third and main project is aimed to elucidate the interplay between structure, dynamics and function in the scavenger decapping enzyme DcpS that performs the last catalytic step in 3'-5' mRNA decay.

In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery

Components of the eukaryotic mRNA decay machinery can accumulate in small cytoplasmic foci named processing bodies (P-bodies).These result from a phase transition process and have been observed in many eukaryotic species, yet the molecular mechanisms of their formation remain largely unknown.

In this work, we combine biophysical methods with microscopy to demonstrate that a minimal set of mRNA degradation machinery from *S. pombe* can undergo phase transition *in vitro*. The interactions permitting the formation of indefinitely expanding intermolecular networks is the contact between multiple helical leucine-rich motifs occurring in Dcp2 and Pdc1 and the LSm domain of Edc3. Interfering with this interaction blocks protein clustering *in vitro* and *in vivo*. Additionally, we describe the structure of the Pdc1 Ge-1_C domain and its binding to the decapping complex as an additional interaction within P-bodies.

Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings

LSm proteins are a varied protein family that is homologous to the Sm proteins in splicing These form a heteroheptameric ring around small nuclear RNA. The nuclear LSm2–8 (like Sm) complex and the cytoplasmic LSm1–7 complex play a central role in mRNA splicing and degradation, respectively. In contrast to the assembly pathway of the Sm complex, the assembly of the LSm complex is, has not been intensely studied yet.

Here, we solved the 2.5 Å-resolution structure of the LSm assembly intermediate that contains LSm5, LSm6, and LSm7. The three monomers display the canonical Sm fold and arrange into a hexameric LSm657–657 ring. We show that the order of the LSm proteins within the ring is consistent with the order of the related SmE, SmF, and SmG proteins in the heptameric Sm ring. Nonetheless, differences in RNA binding pockets prevent the prediction of the nucleotide binding preferences of the LSm complexes.

Using high-resolution NMR spectroscopy, we confirm that LSm5, LSm6, and LSm7 also assemble into a 60-kDa hexameric ring in solution and observe that the LSm domains adopt different angles depending on the number of monomers in the ring. With a combination of pull-down and NMR experiments, we show that the LSm657 complex can incorporate LSm23 in order to assemble further towards native LSm rings. In summary, our results identify LSm657 as a plastic and functional building block on the assembly route towards the LSm1–7 and LSm2–8 complexes.

An excess of catalytically required motions inhibits the scavenger decapping enzyme

The scavenger decapping enzyme (DcpS) hydrolyzes the 5’ cap structure of short capped mRNAs that are the product of the 3' to 5' mRNA decay pathway. Degradation of these fragments is important, as they otherwise could compete for translation initiation machinery.

Here, we solved the crystal structure of the yeast DcpS enzyme in complex with m^7GDP. The 80 kDa homodimer consists of an N-terminal and a C-terminal domain, where the two catalytic sites are located in between these domains. In the asymmetric enzyme:substrate complex one active site is arranged in a catalytically active conformation around the substrate, whereas the second active site adepts an open and unproductive state.

The static structure indicates that protein dynamics are essential for substrate recognition and product release. Using methyl TROSY NMR spectroscopy and ITC, we show that DcpS can interact with two substrate molecules in a sequential manner. The first nanomolar binding event results in a change of the symmetric apo homodimer into a highly asymmetric DcpS:substrate complex. The newly formed open binding site in this complex can then recruit a second substrate with micromolar affinity. Using longitudinal exchange experiments we show that this second binding event leads to DcpS domain flipping motions, where the second binding site closes and the first one opens. These dynamics increase with increasing substrate excess and are up to two orders of magnitude faster than the catalytic rate. To correlate these motions with function, we designed a mutant enzyme where the flipping motions are reduced. Interestingly, this more static enzyme displays increased catalytic activity. Our results thus indicate that DcpS flipping motions that are required for

product release can be disadvantageous for catalysis if they are much faster than substrate turnover. In summary, we present an example of how protein motions in enzyme complexes can modulate activity.

Zusammenfassung

mRNA Abbau ist ein wichtiger Mechanismus, um die Gleichgewichtskonzentrationen von mRNAs zu regulieren und somit die Expression von Genen an die variablen Bedürfnisse einer Zelle anzupassen. Diese Arbeit beschäftigt sich mit den molekularen Mechanismen von drei Prozessen in den beiden Teilen des mRNA Abbaus: Zunächst werden die Interaktionen, die zu einer Phasentrennung und der Bildung von processing bodies (P-bodies) führen, untersucht. Zweitens wird der Assemblierungsweg des heteroheptameren LSm1-7 Komplexes, der vor dem 5'-3' Abbau mit dem 3' Ende von mRNAs interagiert, erforscht. Das Hauptprojekt zielt darauf ab, das Zusammenwirken von Struktur, Dynamik und Funktion des Dcs1p Enzyms, das den letzten Schritt im 3'-5' Abbau katalysiert, zu beleuchten.

In vitro Rekonstitution eines zellulären Phasenübergangsprozesses, der die mRNA Abbaumaschinerie beinhaltet.

In Eukaryoten können sich Komponenten des 5'-3' mRNA Abbauweges in P-bodies, kleinen zytoplasmatischen Foken, ansammeln indem sie eine Phasentrennung durchlaufen. Diese Foki sind in vielen eukaryotischen Spezies beobachtet worden, allerdings sind die molekularen Mechanismen, die ihrem Zustandekommen zugrunde liegen, größtenteils noch unbekannt. In dieser Arbeit kombinieren wir verschiedene biophysikalische und struktubiolgische Methoden mit Mikroskopie und zeigen, dass eine kleine Anzahl an Komponenten der mRNA Abbaumaschinerie aus *S. pombe* ausreichend ist, um *in vitro*eine Phasentrennung zu erreichen. Die Interaktion, die es erlaubt unbegrenzte intermolekulare Netzwerke auszubilden ist die Bindung von kurzen Leucin-reichen Motiven (HLMs) aus Dcp2 und Pdc1 an die LSm Domäne von Edc3. Störung dieser Interaktion führt dazu, dass sowohl in vitro als auch in vivo die Anhäufung dieser Proteine verhindert wird. Wir charakterisieren außerdem die Struktur der Pdc1 Ge-1_C Domäne und ihre Interaktion mit dem Dcp1/Dcp2 Komplex, die ebenfalls in P-bodies stattfindet.

Die Struktur des LSm567 Ringes: Ein Intermediat der LSm1-7 und LSm2-8 Ringe

LSm Proteine sind eine diverse Proteinfamilie, die homolog zu den Sm Proteinen aus dem Spleißosom ist. Der nukleäre LSm2-8 Komplex und der cytoplasmatische LSm1-7 Komplex sind jeweils zentral für das Spleißen und den Abbau von mRNA. Im Gegensatz zu den Sm Proteinen ist der Assemblierungsweg der LSm Proteinen jedoch noch wenig verstanden.

In dieser Arbeit haben wir die Struktur eines LSm Assemblierungsintermediats bei einer Auflösung von 2.5Å gelöst, das LSm5, LSm6 und LSm7 enthält. Die drei Monomere weisen eine kanonischen LSm Faltung auf und bilden zusammen einen hexameren LSm657-

657 Ring. Wir zeigen, dass die Reihenfolge der Monomere im Ring konsistent mit der Reihenfolge der verwandten SmE, SmF und SmG proteine im heptameren Sm Ring ist. Allerdings verhindern Unterschiede in der RNA Bindestelle die Vorhersage der Nukleotidbindung der LSm Komplexe.

Mittels hochauflösender NMR Spektroskopie bestätigen wir, dass LSm5, LSm6 und LSm7 auch in Lösung einen 60kDa großen hexameren ring bilden und beobachten, dass die LSm Domänen abhängig von der Anzahl der Monomere im Ring andere Winkel annehmen. Wir zeigen außerdem mit einer Kombination von NMR und Pull-down Experimenten, dass der LSm657 Komplex LSm23 inkorporieren kann und sich damit weiter in Richtung der nativen LSm Ringe entwickelt. Insgesamt identifizieren unsere Ergebnisse den LSm657 Komplex als ein plastisches und funktionales Assemblierungsintermediat der LSm1-7 und LSm2-8 Ringe.

Ein Überfluss an katalytisch notwendigen Bewegungen des im Dcs1p Enzym hemmt seine katalytische Funktion.

Das Scavenger Decapping Enzym katalysiert die Hydrolyse der 5’ Cap Struktur an kurzen mRNA Abbauprodukten, die aus dem 3’-5’ Abbauweg hervorgehen. Diese Fragmente müssen abgebaut werden, da sie ansonsten unproduktiv Bestandteile der Translationsmaschinerie binden können.

Wir haben die Struktur des Scavenger Decapping Enzyms aus Hefe in Komplex mit seinem Inhibitor m^7GDP gelöst. Das 80kDa große Homodimer besteht aus einer N- und einer C-terminalen Domäne, die beiden aktiven Zentren befinden sich in den beiden Spalten zwischen den Domänen. In diesem asymmetrischen Enzym:Inhibitor Komplex umschließt ein aktives Zentrum ein Molekül m^7GDP in einem geschlossenen, aktiven Zustand, während das andere eine offene und inaktive Konformation annimmt.

Die statische Struktur des Proteins legt nahe, dass Konformationsänderungen im Protein für die Freisetzung des Produktes und die Erkennung des Substrates nötig sind. Wir zeigen mit TROSY NMR Spektroskopie und ITC, dass Dcs1p sequenziell zwei Substratmoleküle bindet. Die erste Bindung erfolgt mit nanomolarer Affinität und führt zu einer Konformationsänderung vom symmetrischen freien Protein zu einem asymmetrischen Enzym:Substrat Komplex. Die dabei entstehende offene Bindestelle interagiert mit micromolarer Affinität mit einem zweiten Substratmolekül. Mittels ZZ-Austausch Experimenten konnten wir zeigen, dass die zweite Bindung eine Klappbewegung der N-terminalen Domäne hervorruft, die die zweite Bindestelle schließt und die erste öffnet. Diese Bewegungen nehmen mit steigender Substratkonzentration zu und liegen um bis zu zwei Größenordnungen über der beobachteten Katalyserate. Um die Zusammenhang zwischen dieser Bewegung und der Enzymfunktion zu etablieren, haben wir eine Mutante entworfen, in

der die Klappbewegungen langsamer sind als im Wildtyp. Interessanterweise weist dieses Protein eine erhöhte Katalyserate auf. Unsere Ergebnisse deuten darauf hin, dass die Klappbewegungen, die für die Enzymfunktion notwendig sind, für die Aktivität des Enzyms von Nachteil sein können, wenn sie sehr viel schneller sind als der Substratumsatz. Zusammenfassend beschreiben wir hier ein neues Beispiel dafür, wie Proteinbewegungen die Aktivität eines Enzymes beeinflussen können.

Table of Contents

List of publications in this thesis

Peer-reviewed articles

- Mund M., Neu A., Ullmann J., Neu U., Sprangers R. (2011)
  Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings.
  J Mol Biol. 2011 Nov 25;414(2):165-76

- Fromm S.A., Kamenz J., Nöldeke E.R., Neu A., Zocher G. and Sprangers R. (2014)
  In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery.
  Angew Chem Int Ed Engl. 2014 Jul 7;53(28):7354-9

- Neu A., Neu U., Fuchs A.L., Schlager B. and Sprangers R.
  An excess of catalytically required motions inhibits the scavenger decapping enzyme
  Nat Chem Biol. 2015 Sep;11(9):697-704

Personal contributions

Chapter 1: In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery. I performed part of the structural refinement and partly supervised structural solution and the rest of the refinement. I co-wrote the manuscript and contributed ideas to the structural part of the project.

Chapter 2: Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings. I collected, processed and analyzed the x-ray diffracion data. I performed the structure determination and refinement of the LSm657 complex, contributed scientific ideas and co-wrote the manuscript.

Chapter 3: An excess of catalytically required motions inhibits the scavenger decapping enzyme. All experiments were planned, designed and performed by me with the exception of the activity assays and the isothermal titration experiment on the C-terminal domain. These

experiments were designed by me and performed by Anna-Lisa Fuchs under my supervision. I performed all data analysis and structure determination. I contributed the scientific ideas of the project and co-wrote the manuscript.

Abbreviations

ACh	acetylcholine
AChE	acetylcholine esterase
C.elegans	*Caenorhabditis elegans*
D. melanogaster	*Drosophila melanogaster*
Dcs1p	scavenger decapping enzyme from *S.cerevisiae*
DcpS	scavenger decapping enyzme from *H. sapiens*
DNA	desoxyribonucleic acid
HeLa	human cell line Helene Lang
LSm	Like-Sm
m^7GDP	7-methylguanosine 5'-diphosphate
m^7GMP	7-methylguanosine 5'-monophosphate
m^7GTP	7-methylguanosine 5'-triphosphate
m^7GpppG	cap analog
mRNA	messenger ribonucleic acid
miRNA	micro ribonucleic acid
NMR	nuclear magnetic resonance
NGD	no-go decay
NMD	nonsense mediated decay
NSD	no-stop decay
oligoA	oligo adenyl
P-bodies	processing bodies
polyA	poly adenyl
PABP	polyA binding protein
RNA	ribonucleic acid
S.c.	*Saccharomyces cerevisiae*
S.p.	*Schizosaccharomyces pombe*
SMN2	survival of motor neurons2
SMA	spinal muscular atrophy

1. Introduction

Throughout life, cells are confronted with shiftingphysiological requirements depending on cell cycle, developmental stage or external requirements due to a changing environment. Reactions to these needs are brought about by changes in gene expression that differentially alter the abundance of RNAs and proteins,which in turn govern the structure and metabolic state of the cell. Gene expression has thus to be highly variable and tunable over space and time but at the same time remain tightly regulated to elicit a controlled response. In eukaryotes, a plethora of mechanisms exists that act on different cellular levels, such as control of DNA structure and modification, transcription, RNA splicing, translation and RNA and protein localization and stability. Together, the variety of gene expression mechanisms allows the integration of diverse sources of information and appropriate reactions to external signals. mRNA degradation is a central mechanism in gene expression, and is used by multiple regulatory pathways[1].

1.1. Eukaryotic mRNA degradation

Overall cytoplasmic mRNA concentrations depend on the steady state equilibrium between synthesis rate,which isdeterminedby transcription and splicing, and mRNA degradation. Total mRNA levels are kept at normal levels by co-regulating these two processes[2].

Half-lives of individual eukaryotic mRNAs vary between species. In yeast, half-lives range between 2-3 minutes and more than 90 minutes [3, 4], whereas in mammalian cells they range between 15 minutes and more than 24 hours[5].mRNA concentrations have to be tightly controlled by the cell in order to match the requirements for differential expression of a given gene at a given time and location in the organism.

Eukaryotic mRNA degradation serves several purposes: Firstly, transcripts have to be routinely turned over. Secondly, it differentially targets transcripts for decay upon the need of up- or down-regulation of genes. Additionally, targeted degradation can be mediated by the action of miRNAs or siRNAs that promote either endonucleolytic cleavage with subsequent decay or recruitment of the decapping machinery. Both mechanisms target the transcripts for pathways of mRNA degradation. Thirdly, faulty transcripts have to be cleared from the cell to prevent synthesis of potentially deleterious protein variants or to release translation machinery that is unproductively trapped on a dysfunctional transcript. Surveillance pathways detect aberrant mRNAs and feed them into mechanisms of regular mRNA turnover. There are dedicated surveillance pathways to identify faulty transcripts containing premature stop

codons (Nonsense-mediated decay, NMD), transcripts that lack a stop codon (Non-stop decay, NSD) and transcripts with stalled ribosomes (No-go decay, NGD).

1.1.1. Structural features of eukaryotic mRNA

Eukaryotic mRNA transcripts are protected from unspecific exoribonuclease activity at both ends. Their 5' end is modified co-transcriptionally with a mRNA cap structure. This cap consists of a 7-methylated guanine nucleotide that is linked with the 5' end of the transcript by a 5'-5' triphosphate bridge. In higher eukaryotes this cap-0 structure is further expanded by methylations at the ribose moieties of the first and second translated bases (cap-1 and cap-2). Apart from protecting the mRNA from degradation, the 5' cap structure is required for RNA quality control, nuclear export and translation initiation.

At the 3' end, transcripts are protected from degradation by a polyA tail that is added after transcription termination and that is required for export from the nucleus. In the cytoplasm, the polyA tail is tightly occupied by the polyA binding protein (PABP). It enhances mRNA stability and the interaction between cap binding proteins and polyA binding protein promotes translation efficiency. The length of the polyA tail varies considerably between species and between individual groups of mRNAs and is altered by deadenylation and readenylation during the lifetime of the mRNA. Generally, greater length is associated with increased stability, whereas reduction of the polyA tail to a critical threshold is the initial and often rate-limiting step of mRNA decay. From shortened oligoA tails, degradation can further progress in 3'-5' direction by action of the cytoplasmic exosome complex or the oligoA tail is recognized by the LSm1-7-Pat1 complex that recruits the decapping machinery. Generally, mRNA degradation is initiated by deadenylation activity at the 3' end of a transcript. Subsequent degradation can take place from both ends of the mRNA, from 5' to 3' and from 3' to 5', however the determinants that assign a transcript to either of the pathways remain unclear. The decapping enzyme Dcp2 and the scavenger decapping protein DcpS eliminate the 5' cap structure in the two pathways respectively to prevent inhibition of cap dependent processes such as nuclear export, translation or RNA processing by excessive concentrations of free cap structure. The products of these hydrolytic events are 7-methyl guanosine mono- or diphosphate. In human and *Drosophila* cells, these compounds are dephosphorylated to yield the methylated base m^7G, however the details of its further degradation pathway remain unknown. [6]

1.1.2. 5′-3′ mRNA degradation

In the 5'-3' pathway, the 5' cap is hydrolyzed by the decapping enzyme Dcp2 between the Pα and Pβ yielding m^7GDP and 5' monophosphorylated RNA as products. Its activity is modulated by a variety of decapping factors, such as Dcp1, Edc3 or Dhh1. While the proteins

involved in mRNA decay are conserved throughout evolution, interestingly their interaction network is not. The decapping step is irreversible and leaves the mRNA body vulnerable for fast and processive degradation by the 5'-3' exoribonuclease Xrn1. The components of the 5'-3' degradation pathway localize to cellular compartments termed P-bodies in which the mRNA degrading enzymes interact in a phase separating net. The cellular function of P-bodies is still being debated.

1.1.3. 3′-5′ mRNA degradation

After initial deadenylation, the 3'-5' degradation pathway proceeds with hydrolysis of the mRNA from the 3' end by the decameric exosome complex. This complex degrades transcripts until they measure ~10nts in length. These short and capped mRNA molecules can compete efficiently for translation initiation machinery and are sequestered from the cell by the scavenger decapping enzyme DcpS that hydrolyses the triphosphate bridge in the cap between the Pβ and Pγ position yielding m^7GMP and 5'-dephosphorylated RNA.

1.2. LSm proteins

Sm proteins were initially named after Stephanie *Sm*ith, a patient suffering from systemic lupus erythematosus. In this autoimmune condition, antibodies are targeted against a set of proteins that share between them the two conserved sequence motifs Sm1 and Sm2 encompassing 32 and 14 amino acids respectively. The Sm proteins (B, D1, D2, D3, E, F, G) are assembled into heteroheptameric toroidal rings that associate with the five snRNAs at their Sm sites forming the snRNP core assemblies. These core snRNPs are part of the U1, U2, U4 and U5 snRNPs (reviewed in [7]). They are indispensible for formation of the spliceosome and thus for correct splicing of pre-mRNAs.

Sm-like proteins (LSm) have been identified bioinformatically by their shared Sm1 and Sm2 motifs [8]. They assemble into two distinct complexes: LSm1-7 and LSm2-8 differ in function and localization but share all but one of their subunits. The order of proteins in the rings has been predicted according to their homology to the Sm components and has been confirmed by structural studies of LSm1-7 and LSm2-8 [9, 10]. More proteins containing Sm domains have been identified since in all trees of life [11]. Bacterial Hfq rings mediate interactions with small RNAs via surfaces on both sides and in the pore of the ring. Also archaea utilize homohexameric rings of Sm proteins to interact with small RNAs.

In addition, divergent LSm domains have evolved to mediate interactions with short linear peptide sequences [12]. These moreover differ by their monomeric state and an interrupted β4 strand.

1.2.1. Eukaryotic LSm rings bind RNAs in diverse cellular contexts.

1.2.1.1. LSm1-7 acts in 5′-3′ mRNA decay.

The LSm1-7 ring is involved in cytoplasmic mRNA degradation and is necessary to couple deadenylation and decapping in 5'-3' decay via its interaction with Pat1p as in absence of the LSm proteins, oligoadenylated mRNAs accumulate while decapping is inhibited. [9,13-17]. Pat1p is a multifunctional multidomain protein that acts as a molecular platform by interacting with components of the 5'-3' RNA degradation machinery such as the decapping enzyme Dcp1/2, the decapping activator Dhh1, the exoribonuclease XrnI, the deadenylation complex Ccr4/Not and polyU RNA. Pat interacts with LSm2 and LSm3 with its and C-domain in *S. cerevisiae* and *H. sapiens* and with its mid-domain in *D. melanogaster*[9, 10, 18].

It has been reported to locate preferentially to the 3' end of oligoadenylated mRNAs in yeast, to the oligouridine tracts 5' of oligoA tails and to oligouridine sequences close to the 3' end of mRNAs *in vivo*. However, the precise sequence preference is unknown. Association of LSm1-7 can protect 20-30 nts of the message from trimming. Both the LSm1-7 complex and Pat1p locate to P-bodies *in vivo*.

1.2.1.2. LSm2-8 is part of the U6 snRNP.

LSm2-8 is localized in the nucleus. The snRNPs U1, U2, U4 and U5 are formed by their respective snRNA and an Sm ring. In contrast, the LSm2-8 ring provides together with Prp24 the protein moiety in the U6 snRNP. LSm2-8 complex binds specifically to the uridine-rich 3' end of the U6 snRNA [19] and remains bound throughout spliceosome assembly, catalysis, disassembly and subunit recycling. U6 confers both sequence specificity and catalytic activity to the spliceosome.

The assembly pathways of the LSm1-7 and LSm2-8 ring appear independent of eukaryotic chaperones and an RNA template as they can be reconstituted during expression in *E. coli*. Some evidence indicates that assembly of LSm1-7 and LSm2-8 complexes is coupled, as overexpression of LSm1 reduces LSm2-8 levels[20]. The mechanism by which biogenesis of these complexes is linked is currently unknown but might involve competition for an assembly intermediate that comprises the common subunits LSm2-7.

1.2.2. LSm proteins form RNA binding rings with a similar architecture as Sm proteins.

1.2.2.1. LSm structures

Structural data for a variety of LSm complexes is available [9, 10, 12, 21-25]. LSm domains adopt a short N-terminal helix followed by 5 antiparallel, highly bent beta strands.

Additional terminal sequences present in LSm1, LSm4 and LSm7 have been implied in functions as RNA binding (LSm1) or P-body association (LSm4). β-strands 2-4 are highly bent around a kinked residue in the middle of the strand. Strandβ5 crosses the open edge of the sheet leading to a roll topology. The hallmark Sm motifs encompassβ1-3 and β4+5 respectively.

Both Sm and LSm proteins assemble into toroid complexes. The most common assemblies are six or seven membered, but also eight-membered rings have been reported. Inter-subunit interaction occurs between β4 of one subunit and β5 of the adjacent subunit so that a continuous β-sheet extends throughout the ring. All N-terminal helices are located on one side of the ring termed the proximal side and pack against β2 and β4 of their own subunit and β2 of the adjacent subunits. Residues located in the loops 3 and 5 line the pore of the ring. These are highly conserved and mediate the interaction with RNA.

Subunit order in the ring is thought to be specified by the subunit-specific interactions. In addition to non-specific hydrogen bonds between the strands, these are mainly electrostatic contacts between side chains in β4 and β5 and in the N-terminal helix. However, subunit interaction is *in vitro*not highly specific as demonstrated by the existence of non-native interfaces in LSm protein heterooligomers [10, 26, 27]. Thus, subunit interaction is a compromise between affinity on one side and specificity and plasticity during assembly on the other side.

1.2.3. Assembly of Sm rings is different from LSm complexes

In vivo, the formation of heptameric Sm rings requires the appropriate snRNA template and the ATP dependent action of assembly machinery. As the sequences upstream and downstream of the ring are base-paired, the Sm ring has to be assembled on its RNA template. In most species, snRNP assembly takes place in the cytoplasm, so that the snRNA is exported from the nucleus after transcription. Its Sm binding site is specifically recognized by GEMIN5 and bound to the SMN (survival of motor neurons) complex consisting of SMN and a group of proteins collectively termed gemins. The Sm proteins form building blocks consisting of SmD1/SmD2, SmB/SmD3 and SmE/SmF/SmG and are Arg-N methylated at their C-termini by action of PRMT5. These preformed Sm oligomers are assembled into the final heptameric complex around the snRNA target in the SMN complex. First, SmD1/SmD2 and SmF/SmE/SmG are combined with pICln, a chaperone that occupies the position of SmD3/SmB in the ring shaped assembly intermediate. After insertion of the snRNA, the chaperone is expelled and the ring is completed by SmD3/SmB. After further processing as exonuceolytic trimming and hypermethylation of the cap structure, the snRNP is reimported into the nucleus.

Also *in vitro*, the Sm proteins multimerize into the stable subcomplexes SmD1/SmD2, SmB/SmD3 and SmE/SmF/SmG. These can assemble spontaneously into the native heptameric ring in presence of an appropriate snRNA [28, 29]. The action of the SMN complex in the cell is thought to provide specificity in RNA recognition and avoid the unproductive assembly of non-native intermediates[30-32].

In contrast to the extensive assembly machinery used for Sm proteins, no accessory factors for LSm ring formation are known. LSm complexes assemble in contrast to Sm complexes without template RNA [19, 33] and chaperone action. Both the homology between Sm and LSm proteins and the presence of stable LSm subcomplexes however suggest that LSm proteins retrace the assembly mechanism of Sm proteins.

1.3. Eukaryotic scavenger decapping proteins

1.3.1. Scavenger decapping enzyme function

Each of the two branches of mRNA degradation possesses a decapping activity to remove the 5' cap from mRNA during degradation. DcpS was first described as activity that degrades the triphosphate cap in capped RNAs of < 10 nts and characterized in human cells [34, 35]. The protein was later termed scavenger decapping enzyme (DcpS). It performs the decapping reaction at the end of 3'-5' decay, hydrolyzing the Pβ-Pγ phosphodiester. Human DcpShas been reported to interact with the exosome and suggested to coordinately decap mRNAs once degraded by the exosome[36].

Scavenger decapping enzymes are conserved from fungi to higher eukaryotes, however no homologue is known in plants. As part the HIT family of pyrophosphatases, they have been identified as such by the typical histidine triad motif (H-Φ-H-Φ-H-Φ, where Φ denotes a hydrophobic amino acid). While the C-terminal domain of DcpS is homologous to HIT proteins and contains the histidine triad motif, it is, in contrast to other HIT proteins, inactive on its own accord and requires the N-terminal domain in cis or trans for activity [37]. Disruption of yeast DcpS induces cellular accumulation of m^7GpppN [38].

1.3.2. Catalytic mechanism

The catalytic mechanism of scavenger decapping enzymes is thought to proceed via a S_N2 mechanism, which has been demonstrated for another member of the HIT family, the FHIT protein [39]. There, the deprotonated central histidine in the histidine triad carries out a nucleophilic attack on the nucleoside phosphate at the γ-position. The reaction proceeds through a bipyramidal transition state. Subsequently, pp-RNA is released while a covalent

nucleotidyl-phosphohistidine adduct is formed. Nucleophilic attack of a water molecule at the γ-position hydrolyzes the adduct and releases the other product m^7GMP from the enzyme.

1.3.3. Substrate preferences

The substrate range of scavenger decapping enzymes is mainly determined by three factors: Modifications of the capping base, length of the RNA body and the number of phosphates adjacent to the methylated 5' base.

Scavenger decapping enzymes are only active on N-7 methylated guanosine derivatives and thus exhibit intrinsic specificity for mRNA caps[34, 35]. Homologs from species that feature additional methylations in the capping base are able to hydrolyze these cap structures, however at different rates than the singly methylated bases [40].

Concerning mRNA length, DcpS enzymes are most active on isolated cap structure (m^7GpppN) without exhibiting specificity for the 5' base[41]. Activity rapidly decreases when RNA body length exceeds 10 nucleotidesfor human DcpS [42]. This requirement is in agreement with scavenger decapping enzymes acting downstream of the exosome,whichproduces short, capped mRNA fragments. Thus, the cap structures of longmRNAs are preserved by DcpS proteins. A notable exception of this prerequisite is the enzyme from *S. pombe* that has been reported to act on longer transcripts[43].

The presence of a base after the triphosphate bridge is not strictly required for scavenger decapping activity, as m^7GTP has been shown to be hydrolysable by both the yeast and nematode enzymes. Rates of catalysis are in the range of the hydrolysis of m7GpppG[44].

The requirement for three phosphates for catalysis is elucidated by the substrate m^7GDP. Activity on m^7GDP, the product of 5'-3' degradation, has been reported, but issubject to controversy. While it has been reported active and implied in cross-regulation between the two branches of mRNA decay [45, 46], more recent studies did not detect significant activity on this compound[42, 44, 47]. As m^7GDP can still bind to scavenger decapping enzymes, it acts as an inhibitor of DcpS activity. Thus, the minimal number of phosphate groups in the cap required for efficient catalysis is three.

1.3.4. Localization

The localization of DcpS proteins varies among species. Its namesake function in mRNA turnover suggests a cytoplasmic localization, however a predominantly nuclear localization has been reported in mammals[37, 48] and *S. pombe*[43]. Accordingly, a canonical nuclear localization signal (NLS) and nuclear export signal (NES) have been identified in DcpS from *H. sapiens*. In agreement with these findings, human DcpS is a shuttling protein with its bulk located in the nucleus[37]. In contrast, DcpS enzymes in in

nematodes [47, 49] and *S. cerevisiae*[50] feature no distinct NLS or NES and localize mainly to the cytoplasm. This may reflect a different range of functions that DcpS proteins perform in different organisms.

1.3.5. Other cellular functions

Many proteins involved in RNA metabolism moonlight in other cellular functions. While originally identified for their decapping activity, DcpS enzymes have been implied in a variety of cellular processes in different species, offering a rationale for their reported nuclear localization. The effects of DcpS function can be brought about either by a mechanism that is independent of its catalytic activity, by the effect of decapping activity on mRNAs or indirectly, by modulating the steady state levels of its substrate, inhibitor or product (m^7GpppG, m^7GDP and m^7GMP). These small molecules in turn affect the cell by different mechanisms: if not cleared from the cell, they may compete with functional mRNAs for cap binding proteins [51] or become incorporated into DNA[46].

1.3.5.1. RNA turnover

A first sector of action is modulation of miRNA and mRNA turnover independent of decapping activity. In *C.elegans*, DcpS has been reported to interact with XrnI, enhancing miRNA degradation independent of DcpS decapping activity [52]. In *S.cerevisiae*, Dcs1p has been found to activate 5'-3' mRNA degradation by interacting with Xrn1 and increasing its affinity for RNA [38, 53]. This effect is observable *in vitro* and *in vivo* and is independent of the decapping activity of Dcs1p. In the absence of Dcs1p, the degradation defect of XrnI leads to decreased expression of the mitochondrial Por1 protein. As respiratory genes are required for yeast to metabolize non-fermentable carbon sources, this might explain the inability of Δdcs1 yeast strains to grow on glycerol as a carbon source[53]. It is notable that Dcs1p, a protein at the end of 3'-5' degradation, co-regulates XrnI, an enzyme in 5'-3' decay.

1.3.5.2. Splicing regulation and gene expression

Furthermore, DcpS has been involved in splicing regulation. First-intron splicing is reduced in DcpS knock-down cells, which has been attributed to sequestration of the nuclear cap binding protein CBP20 by excess cap structure resulting from reduced DcpS activity[51]. This function is affiliatedwith the nuclear fraction of DcpS that has been observed in human cells. By a similar mechanism, sequestration of the cytoplasmic cap binding protein eIF4E by excess cap structure due to absence or reduced activity of DcpS is thought to influence translation[51].

Inhibition of scavenger decapping activity by C5-substituted quinazolines has been shown to increase SMN2 expression [54] and thus alleviate the symptoms of spinal muscular

atrophy in a mouse model [55, 56]. The mechanism of expression activation is yet unknown. However, this finding stresses that DcpS is involved in humans to gene expression.

1.3.5.3. Dcs1p acts in *S. cerevisiae* carbohydrate metabolism

Dcs1p has been described as a neutral trehalase inhibitor in yeast and has been reported to directly interact with it [57, 58]. Trehalose (α1-glucopyranosyl-α1-glucopyranosid) is accumulated under nutrient deprivation and stress conditions to protect the cell and organelles[59]. After alleviation of stress, it is hydrolyzed by the neutral trehalase Ntr1 yielding glucose for quick resuming of growth. Ntr1 is already expressed under stress conditions and needs to be inhibited in order to prevent futile cycles of concomitant trehalase synthesis and degradation by Ntr1. DcpS is highly expressed jointly with Ntr1 at the onset of stress and to physically interact with Ntr1, reducing its interaction with Ntr1 activators. In a Dcs1p deletion strain, Ntr1 is constitutively active, pointing to a inhibitory function of Dcs1p in trehalose metabolism in response to stress[50, 60, 61].

Furthermore, Dcs1p is regulated on a nutrient dependent basis. Dcs2p is a homologue of Dcs1p in S. cerevisiae that shares 65% sequence identity and 90% similarity with Dcs1p. Its expression is repressed in log growth phase but becomes induced both transcriptionally and translationally as soon as glucose becomes limited [46, 50]. Dcs2p localizes in stationary phase mostly to P-bodies with a fraction remaining distributed throughout the cytoplasm. Newly translated Dcs2p proteins can form inactive homodimers or heterodimers with Dcs1p [46], most likely adopting the same architecture as Dcs1p dimers. Dcs2p can be considered a nutrient-dependent modulator of Dcs1p activity as Dcs1/2p dimers exhibit reduced activity on m^7GpppG compared to the Dcs1p dimers. Inhibition of Dcs1p by Dcs2p may feed into modulation of the 5'-3' mRNA degradation but has also been implied in reducing life-span due to increased incorporation of methylated nucleotides into RNA [46].

Taken together, scavenger decapping enzymes are involved in mRNA and miRNA turnover, gene expression, stress response and metabolism. DcpS functions differ in cellular localization, in whether they are direct or nucleotide mediated and in whether Dcs1p acts as a homo- or heterodimer. If and how these functions are jointly coordinated is unknown at present, as are all molecular details of these additional activities.

1.3.6. Structure of scavenger decapping enzymes

DcpS enzymes from human and mouse[62] have been studied extensively by x-ray crystallography. Structures of human DcpS in free, substrate (m^7GpppG, m^7GpppA), product (m^7GMP) and inhibitor bound states (m^7GDP, D153249, D156844, D157493) are available and picture intermediate states of its molecular mechanism of action [41, 54, 63].

DcpS enzymes are homodimeric proteins with a distinct two-domain structurethat behave as rigid bodies in published structures. The N-terminal domain encompasses residues 1-134 in the human protein and is linked with the C-terminal domain (residues 148-336) by a hinge region that consists of an α-helix and amino acids without secondary structure. Both domains are dimeric on their own accord. The N-terminal NTF2-like domain dimerizes by a domain swap where one protomer donates an α-helix and two anti-parallel β-strands to complete the fold of the other protomer. In the C-terminal domain that is a homologous to the HIT protein family, the protomers interact via an extended shared surface that buries 1400Å^2. In other HIT family pyrophosphatases, this domain is sufficient to carry out catalysis, however for DcpS the presence of the N-terminal domain is required for activity[37, 64].

1.3.7. Conformational change between free and substrate bound DcpS

DcpS is a homodimer and adopts a symmetric conformation in the free form with a two-fold symmetry axis. The domains do not interact with each other in this conformation but are separated by the hinge region [62, 63].

In the substrate or inhibitor bound form, DcpS interacts with two molecules of m^{7}GpppN but undergoes a conformational change that results in different binding sites for its two substrate molecules. In a flip-over motion, the N-terminal domain and the hinge fold down onto the C-terminal domain forming a binding pocket around a substrate molecule. The active site residues arelocated on the C-terminal domain and correspond to a canonical HIT protein catalytic site. Interestingly, the N-terminal residues involved in substrate binding are situated on the domain-swapped β-strands and thus contributed by the other protomer. This active site tightly encloses the substrate and is thus termed ‘closed’.

The second binding site is occupied in crystal structures as well but compared to the apo structure, the domains have moved apart and binding residues from N- and C-terminal domain are remote from each other. The conformational change between free and bound states is mainly mediated by a series of changes in backbone torsion angles in residues in the hinge region, while the domains act as rigid bodies. This indicates that closure of one active site has to occur concomitantly with opening of the second site and that the productive arrangement of one active site is mutually exclusive with activity in the second. A hypothetical simultaneous closure of both sides would require a large distortion of either of the two domains, which is highly unlikely given their observed rigidity and stable domain interfaces.

1.3.8. Active sites in closed, open and free conformation

The active site in the closed conformation compromises the C-terminal catalytic residues and is completed by parts of the N-terminal domain and the hinge region surrounding the substrate. A van-der-Waals cage that is mainly formed by aromatic residues encloses the methylated base;the phosphate groups are specifically recognized by electrostatic interactions. The binding pocket for the second base is promiscuous and can accommodate both purines and pyrimidines or remain void if m^7GDP is bound. C5-substituted 5-quinazoline inhibitors mimic key features of natural substrates and contact the same residues in the active site for interaction[54].

Structural information is available for the open conformation in either the free form or occupied with either substrate m^7GpppG, inhibitor D153249 or the product m^7GMP. The C-terminal domain recognizes substrate molecules with their methylated base moieties in a similar way as in the closed conformation. The conformation of the substrate differs in the phosphate bridge in Pα and Pβ and the non-methylated base: The dinucleotide adopts a more extended conformation and the second base is bound more loosely by stacking interactions with the hinge region compared to several electrostatic interactions in the closed conformation. Together with higher B-factors in the second base compared to the first one, this indicates that the bulk of binding energy is provided by interaction with the C-domain. The inhibitor D153249 is bound with lower occupancy to the open active site. In the open conformation, the methylated base of the substrate is bound in a similar manner as in the closed active site. Substrate conformation however differs in the orientation of the Pβ and the second base. The open binding site in complex with the product m^7GMP bears more similarity with the free site than with the substrate bound one. The methylated base is recognized differently.

The catalytic residues in HIT proteins are sufficient to perform catalysis. This implies that the C-terminal domain would also be sufficient to hydrolyze the cap structure. The arrangement of the C-terminal catalytic site residues is very similar between open and closed active site yet there is a large difference in substrate conformation. It is tempting to speculate that the activation in the pocket is brought about by activation of the leaving group in the substrate, enforced by the closed position of the N-terminal domain.

1.3.9. Structure-function relationships in DcpS enzymes

Insight into the relationship between structure, function and dynamics in scavenger decapping enzymes is based on structural and kinetic investigations.

The structures of free and ligand bound DcpS enzymes demonstrate that molecular motions are required for DcpS activity, as conformational change is required to release the product from the closed active site [41, 63]. Aconcertedmechanism for substrate binding and product release in the two active sites is suggested: Both the closed site A and open site B

are occupied with substrate. Product release after catalysis in A requires opening of the closed site. Opening of site A leads to simultaneous closure in B and thus to productive active site formation in the previously open site. This mechanism proposes an increase in catalytic efficiency as substrate binding and product release can take place simultaneously.

DcpS mutants in whom the interaction between the domains in the closed conformation has been disrupted display increased activity, indicating that product release might be the rate-limiting step in the catalytic cycle[41]. However, the mutated residues are located in the hinge region and are interactwith the substrate in the open protomer. Thus, the mechanism by which the observed hyperactivity is realized remains to be investigated.

Additional information was gained from kinetic analysis of homo- and heterodimers of human DcpS protein[65]. This study demonstrated that DcpS is a negatively cooperative protein, as implied previously. Additionally, pre-steady state kinetic experiments demonstrated that Dcs1p activity is not rate-limited by product release rates, raising the question of the rate limiting step. Interestingly Dcs1p is substrate-inhibited, i.e. activity is fast under single turnover conditions but slower under multiple turnover conditions. This effect was reduced in heterodimers that were unable to interact with substrate in one protomer, indicating that the second active site is involved in substrate inhibition. Comparing the events in single and multiple turnover reactions, it was concluded that the step differing between them, the conformational change between the symmetric and asymmetric state, had to be the rate limiting one.

Structural and kinetic data demonstrate clearly that the active sites in Dcs1pinteract and that protein dynamics are required for Dcs1p function. These experiments have shed light on DcpS enzyme function. However, in order to understand the interplay between structure and conformational dynamics and how they give rise to its properties as allostery and substrate inhibition, experimental measurements of the rates of conformational exchange are still required.

1.4. Protein movements in allostery

1.4.1. Allostery

The term allostery describes the phenomenon that an interaction in one site of a biomolecule can influence activities such as binding or catalysis at a distant site. This phenomenon has long been known and recognized for its biological relevance. Allostery is relevant for cellular signaling, metabolism and disease; drugs acting at allosteric sites have successfully been designed. By integrating information from one site of a macromolecule with activity in another it is possible to tune activity according to physiological conditions.

Compared with a system, in which progression of a reaction is determined kinetically by the concentration of reactants or products, integration of information by allosteric mechanisms allows for additional levels of regulation between pathways for heterotropic systems or for other dynamic ranges of ligand for homotropic systems.

Allostery can be divided in homotropic and heterotropic effects. In homotropic allostery, molecules of the same compound interact with the allosteric site and regulate activity in the orthosteric site. In heterotropic allostery, different compounds are bound to the different sites. Homooligomers are particularly prone to homotropic allosteric phenomena because multiple binding sites are present in the molecule already. Additionally, residues in the interfaces that permit signal transduction in between the protomers exert their effect in all copies and are thusmore efficient in propagating conformational or dynamic changes. Allosteric behavior has however been demonstrated for monomeric proteins as well.

Classically, allostery in homooligomers has been described by the concerted (MWC) model or the sequential (KNF) model. The concerted modelassumes that all protomers in a complex adopt the same conformation, which can be either highand low affinity conformation. Ligand binding shifts the equilibrium between all-high and all-low affinity states and thus alters the affinity of subsequent binding events.In contrast, the sequential model is based on the principle of induced fit and states that single protomers undergo conformational change upon ligand binding. Thus, liganded protomers always occur in the binding competent conformation, and a change in affinity of subsequent binding sites is due to stabilization or destabilization of interfaces between protomers of different conformations. Both models have been successfully applied to describe protein allostery.

These models are deterministic insofar as a binding event always leads to conformational change elsewhere in the protein, either in all protomers in the MWC model or in only the binding subunit in the KNF model. These rely on a strict mechanical coupling between intermediate stages of conformational exchange. In contrast, probabilistic allosteric models model mutually dependent conformational equilibria for all protomers and invoke ensembles of proteins that statistically explore conformational space. This is in correspondence with the notion that proteins in solution exhibit considerable conformational diversity. Binding events differentially stabilize populations and thus alter the composition of the ensemble, ligand binding is thus interpreted as a remodeling of the free energy landscape of the ensemble. Probabilistic models still cover the concerted and sequential model as extreme cases.

These models describe allostery on a phenomenological level, however the underlying molecular mechanisms are understood for few examples. Underlying paradigms that apply include conformational change, change of protein dynamics and population shifts in protein ensembles. In studying a given allosteric system it is a challenge to determine which effects

govern allostery on the molecular level and which model reflects and describesthis mechanism appropriately. Recent workhas shown that the same protein can act according to different models for different ligands indicating the difficulties in predicting allosteric mechanisms[66].

1.4.2. Protein dynamics in allostery

First insights on the molecular details of allostery were obtained by structural investigations that provided static structures of end-point states of conformational change. Early theoretical considerations demonstrated that measurable allosteric effects can be generated by shifts in frequency and amplitude of thermal fluctuations without perturbing the average structure of a protein [67]. Indeed, using solution techniques as NMR, examples are accumulating that demonstrate that allostery in proteins can be brought about by changes in protein dynamics that increase or decrease the ability of a protomer to interact with a ligand. The mechanisms found expand the repertoire of mechanisms frompurely enthalpic structural changes to dynamically driven allostery where entropic contributions manifest as motions in side-chains, the backbone or local or complete unfolding of the protein.Further information is accessible by computational methods as transition state analysis, anisotropic thermal diffusion, normal mode analysis or statistical coupling analysis. Thesecontribute information on possible allosteric mechanisms on a range of timescales between femtoseconds and microseconds. As dynamics have been established as important factor in allostery, the complete description of protein movements has become relevant in discerning allosteric mechanisms. Thus, detailed molecular information on populated states, the conformational width of their ensembles, their relative populations and the rates of interconversion are required to complete the static description.

1.5. Substrate inhibition

Michaelis-Menten kinetics postulate that catalytic rates increase with increasing substrate concentration until reaching saturation, where the turnover is limited by the catalytic rate inherent to the enzyme and not by the supply of substrate molecules. However, a significant amount of enzymes defies this model as, after a maximum in activity, turnover rates decline with increasing reactant concentrations either to a constant or to zero. As increasing substrate concentration reduces activity, this phenomenon is termed substrate inhibition and is common to many enzymes [68]. The biological significance of substrate inhibition is controversial. While it is fairly common occurring in an estimated 20% of enzymes [69], it has often been argued to result from non-physiological substrate concentrations in assays. Nevertheless, a number of enzymatic reactions in the cell take

place under inhibiting substrate concentrations so that a biological function substrate inhibition has to be suspected.

1.5.1. Cellular functions of substrate inhibition

For some substrate-inhibited enzymes, the cellular function of their inhibition has been elucidated. Generally, in context of a metabolic pathway, substrate inhibition alters steady state levels of metabolites or modulates the substrate concentration ranges in which the pathway will be active as demonstrated by several examples.One is phosphofructokinase, the third enzyme in glycolysis, which is inhibited by its substrate ATP[70]. This inhibition is physiologically worthwhile as ATP is also the end product of the metabolic activities of glycolysis, citric cycle and ATP synthase so that an end-point inhibition is here conveyed by substrate inhibition.

Alternately, substrate inhibition can modulate the temporal response of a pathway to a given stimulus. For instance, substrate inhibition of acetylcholinesterase (AChE) provides for accelerating clearance of its substrate acetylcholine (ACh) from the synaptic cleft. Thus, the very fast build-up of acetylcholine after vesicle discharge leads to AChE inhibition and faithful signal transduction. The inhibited enzyme slowly degrades ACh until the threshold of inhibition is crossed and ACh is cleared quickly from the synaptic cleft. Thus, both robust signal transduction and fast ACh degradation are realized[71, 72].

1.5.2. Molecular mechanisms of substrate inhibition

The kinetic model of substrate inhibition invokes a doubly substrate bound state of the enzyme with decreased turnover. With increasing substrate, the occupancy of the enzyme is shifted to this less active state and its overall activity decreases. However how the reduction in turnover is realized on the molecular level is yet to be elucidated for most substrate inhibited proteins. Nevertheless, substrate inhibition has to be considered an allosteric effect as interaction at a remote site regulates activity at the active site.

Substrate inhibition can be realized by secondary, nonproductive binding sites that impede substrate access to the active site or reduce turnover [73-75], or hinder product release [72, 76]. Secondary binding sites can be close to the active site [77] or more distant and act via an allosteric mechanism[78]. Known mechanisms of substrate inhibition have been characterized by structural investigations and mostly rely on steric effects at the active site. The analogy with allostery in binding proteins however suggests that protein dynamics may have a role in mechanisms of substrate inhibition as well.

2. Interactions involved in P-body formation

2.1. Aims and significance

P-bodies are cellular structures resulting from phase transition in the cytoplasm that have been described in a variety of mammalian species. While the list of proteins localized there is expanding, their function is still being debated and the molecular mechanisms of their formation are not fully understood. Our aim was to address a variety of questions:

- Can a minimal reconstituted system of purified proteins undergo phase transition *in vitro* using the known interations between the Edc3 protein and HLM motifs from Dcp2 from S. pombe?
- Does disruption of an interaction *in vitro* mirror *in vivo* effects?
- What other properties of P-bodies are emulated by this system, such as dynamic fusion or recruitment of other mRNA decay factors?
- What is the influence of number and affinities of the protein components on the conditions where phase transition occurs?
- Pdc1 is a multi-domain protein has been implied in P-body integrity. Which of its interactions recruit it to P-bodies and what are their structural details?

In order to analyze the contribution of single interactions to the formation of an indefinitely extending network of interacting proteins that leads to phase transition, we set out to characterize affinities and interacting motifs in these proteins. In S. pombe, these include Edc3, the decapping complex Dcp1/Dcp2 and Pdc1. We use a combination of NMR, ITC and x-ray crystallography to study structural determinants of the relevant interactions and their affinities. Additionally, insights into the effect of these interactions on phase transition are gained by a bottom-up *in vitro*assay that uses fluorescence microscopy.

2.2. Results

2.2.1. The interaction between multivalent constructs of Dcp2 HLMs and Edc3 LSm domains can induce phase transition

Dcp2 contains helical leucine-rich motifs (HLMs) that interact with the LSm domain of Edc3 [79]. Here, seven interacting HLMs from Dcp2 have been confirmed by NMR. Their affinities range from the low µM to the mM order, as determined by ITC. When mixing the purified Dcp2 far C-terminal region containing four HLMs with purified with dimeric full-length Edc3, a clear phase transition can be observed under a bright field microscope. The resulting

droplets contain both proteins as Oregon green labeled Dcp2 (553-741) and Edc3 are significantly enriched in the droplets in individual assays.

In order to investigate the requirements for phase transition, a series of assays was carried out under varying concentrations of both proteins and with two Dcp2 constructs containing either four or five HLM sequences. Overall, phase separation was found depending on a number of conditions: The overall concentration of both proteins has to be sufficiently high. Secondly, the molar ratio between the proteins had to be within certain limits, as too much of an excess of either protein abolished phase separation. Thirdly, the valence, i.e. the number of interacting modules within one protein, and their affinities influence the concentration range in which phase transition can be observed.

Following these concepts, it was speculated that an increasing concentration of monomeric HLM peptide added to a phase transition assay should be able to interrupt phase transition. Monomeric HLMs should saturate the interactions sites of Edc3 without promoting the formation of an indefinitely branched continuous network. Indeed, addition of monomeric HLM-1 suppressed phase separation *in vitro*. When the same HLM peptide from Dcp2 is overexpressed *in vivo*, this interferes with the localization of Edc3 to P-bodies. Interestingly, P-bodies are still present in these cells underlining the redundancy of mechanisms for P-body formation. This experiment demonstrates that findings from the *in vitro*system are in agreement with proceedings in a cellular environment.

To additionally investigate how accurately our *in vitro* system could represent the properties of P-bodies, their dynamics and ability to incorporate interacting proteins were examined. Indeed, fusion events similar to those observed in the highly dynamic P-bodies [80-83] could be observed. Moreover, phase separation was also observed when mixing Dcp1/Dcp2 constructs (Dcp2 1-289 + 553-741 excluding the mid-domain) with full length Edc3. The droplets contained Dcp1 protein, demonstrating that as in the cellular foci, additional factors can be recruited into the droplets by additional interaction sites on the scaffolding proteins.

2.2.2. Dimeric Pdc1 can induce phase transitions via its HLM motifs

Pdc1 has been implied in P-body formation in S. pombe [84] and is related to metazoan Edc4 [85]. It encompasses a WD40 domain, a coiled-coil multimerization domain and a C-terminal domain with similarity to D. melanogaster Ge-1_c. Additionally, the N-terminal region of Pdc1 displays three HLM motifs. To assess the function of the HLM motifs in P-body formation, the interaction between the motifs and the Edc3-LSm domain was validated and quantified by NMR and ITC. The affinities range from high µM to mM. Pdc1 that can multimerize via its coiled coil domain can theoretically promote phase transition by a high valence of HLMs per molecule. As no soluble Pdc1 could be obtained, an engineered version

in which the coiled-coil domain was replaced by glutathione S-transferase (GST) to permit for dimerization was expressed. Phase transition assays demonstrated that also Pdc1 undergoes formation of an oil-like droplet phase in presence of Edc3 indicating the presence of an indefinite network of intermolecular interactions. Hence, Dcp2 and Pdc1 are redundant proteins in their ability to induce phase separation by interaction with Edc3.

2.2.3. The C-terminal domain of Pdc1 adopts a Ge-1_c fold and can directly interact with the Dcp1/Dcp2

In addition to the HLM motifs and the coiled-coil multimerization domain which are required for induction of phase transitions, Pdc1 encompasses in the far C-terminus a sequence with similarity to Ge-1_cdomains. We verified the fold of this domain by solving the x-ray crystal structure of Pdc1 932-1076 at a resolution of 1.35Å. The structure revealed close structural similarity to the C-terminal domain of *D. melanogaster*Ge-1. It consists of a very closely packed helical bundle with eight α-helices. The N-terminal three and C-terminal five helices are oriented in a 90-degree angle from each other.

The C-terminal 290 residues of Pdc1 interact with Dcp2 [84] and encompasses the Ge-1_C domain. To assess whether it is directly involved in this interaction, NMR titrations with labeled Pdc1 Ge-1_C and unlabeled Dcp1/Dcp2 complex were performed. Upon addition of the decapping complex, resonances of residues located in the C-terminal helices were perturbed indicating interaction with Dcp1/Dcp2 in this site. This finding demonstrates that at least two types of interaction take place between Pdc1 and components of P-bodies: Firstly, HLM motifs in the N-terminus interact with dimeric Edc3 and secondly, the C-terminal Ge-1_C domain directly contacts the decapping complex. Multimeric Pdc1 has for both interactions a valence of > 1 so that both interactions are expected to contribute to phase transition processes. These findings explain the described function of Pdc1 in P-body integrity and underline its importance as a scaffolding protein.

2.3. Discussion

2.3.1. Determinants of *in vitro* phase transition by HLM LSm interactions

Using a bottom-up approach to reconstitute a minimal system that can undergo phase transition from purified recombinant proteins, it was possible to characterize the determinants required for phase transitions *in vitro*. The droplet phase contained both recombinant proteins as demonstrated by the localization of Oregon Green dotted Dcp2 and Edc3. We investigated the influence of overall concentration of proteins, of relative excess of one protein over the other and of valence.

Overall concentrations to obtain phase separation were required to be in the high micromolar range and the modular concentrations needed to be similar to each other. A high excess of either of the components did not result in phase separation, as this component would saturate the binding sites of the other interaction partner and then does not allow for a branched network that can indefinitely expand within the protein phase. Also, addition of an excess of a monomeric HLM peptide resulted in disruption of the droplet phase as the interaction with the peptide would bind available Edc3 LSm sites without allowing for a branched network. *In vivo*, overexpression of the same peptide resulted in the loss of Edc3 localization to P-bodies, demonstrating that Edc3 is recruited to P-bodies by the interaction with HLM motifs. Moreover this experiment underlines the redundancy of mechanisms for P-body formation in the cell.

As expected, the valence of interactions was relevant, i.e. the number of interacting motifs per molecule. In Dcp2, there are seven identified HLMs whereas the valence of Edc3 is two, as the dimer encompasses two LSm domains. Increasing the valence of the Dcp2 construct lowered the overall concentrations needed to obtain phase separation, consistent with the concept of avidity.

The affinities of the interactions between Edc3 LSm and the various HLMs from Dcp2 and Pdc1 vary between low µM and the mM range. Due to the high overall concentrations in P-bodies even interactions that are considered weak may result in appreciable saturation under these conditions.

2.3.2. Droplet phase separation is an *in vitro* model for P-body formation

We here present a bottom-up approach to model the phase separation events occurring in P-bodies *in vitro*. Due to the redundancy of the process *in vivo*, efforts to study its mechanisms by genetic approaches are hampered [86]. Thus a reconstituted system of recombinant proteins allows building up a minimal essential system in which single events in phase transitions can be studied without redundancy.

This model emulates diverse properties of cellular P-bodies, such as fluid dynamics or incorporation of interacting proteins into the droplet phase, e.g. Dcp1 via its interaction with Dcp2. Additionally, we have shown that overexpression of monomeric HLM peptide interferes with Edc3 localization *in vivo* and *in vitro*.

Further work will explore how more components of mRNA decay can localize to the droplet phase. For example, the helicase Dhh1 has been shown to interact with the FDF repeat of Edc3 and could thus be recruited. mRNA can be bound by the catalytic domains of both the helicase and the decapping complex and in return provide a binding platform for other RNA interacting proteins.

Furthermore, this *in vitro* model can be used to investigate the mechanisms that underlie P-body formation upon stress. It has been speculated that protein concentration in the cellular environment are close to a phase transition in order to allow for rapid phase separation upon small changes in the system such as application of a posttranslational modification or small changes in protein concentration. Our model allows for the investigation of the effects that permit droplet formation, such as concentration, avidity and affinity of the interactors. It could also be used to study if and how phase separation influences the activity of the enzymes involved.

2.3.3. Interactions in P-bodies

As mentioned before, the mechanisms underlying P-body formation are highly redundant and variable among species [87]. Previous work has demonstrated that low complexity regions in the RNA binding protein FUS can induce the formation of amyloid-like fibrils and induce phase separation [88-90]. This mechanism is different from the interaction we observe between HLMs from Dcp2 and Pdc1 and the Edc3 LSm domain. However, also other proteins that localize to P-bodies display low complexity regions e.g. LSm4, Pat1p and XrnI. Both interactions are expected to occur concomitantly enhancing the phase separation process. Therefore, protein concentrations required for phase separation in the cell are expected to be much lower than in the simplified and less redundant *in vitro* system.

We have characterized an additional interaction within the network of P-body proteins. We show that Pdc1 contains a Ge-1_Cdomain and have solved its structure by x-ray crystallography. This part of the protein had been reported to interact with Dcp2 and we have shown by NMR that this domain indeed interacts with the decapping complex Dcp1/Dcp2 using its C-terminal helices. Any additional interaction between P-body components is expected to influence the requirements of the phase transition process so that direct protein-protein interactions have to be characterized and taken into account to predict conditions of phase transition.

It remains to be noted that interactions in the network of P-body components may not only serve the purpose of localizing proteins to these foci, but also to modulate their catalytic or binding functions. Thus, further work will be aimed at dissecting how molecular interactions lead to P-body formation and function.

3. The LSm657 complex: a functional and plastic assembly intermediate

3.1. Results

Structure and RNA sequence specificity have been well established for the Sm complex. For the closely related LSm proteins structural information was lacking at the time of the work. The sequence conservation between Sm and LSm proteins argues for similar ternary and quaternary structures for the LSm proteins. Particularly, the pairwise sequence similarity suggests that each single LSm protein is most similar to a single Sm protein (SmB-LSm1/LSm8; SmD1-LSm2; SmD2-LSm3; SmD3-LSm4; SmE-LSm5, SmF-LSm6 and SmG-LSm7). This argues for conserved functions and a common assembly pathway shared by the LSm ring. On the other hand, the complexes differ however in their cellular function, their stability in the absence of RNA, their requirement for RNA and the action of chaperones during assembly. Thus, experimental data is required to characterize LSm rings and their assembly on a structural and functional level.

3.1.1. The x-ray crystal structure of the S. pombe LSm657

3.1.1.1. The LSm657 hexamer displays the canonical features of LSm complexes.

We here present the structure of the LSm657 complex from *S. pombe* solved by x-ray crystallography at a resolution of 2.5 Å.

Two copies of each LSm protein are arranged in a ring-shaped heterohexamer, which is together with heptamers a typical oligomerization state for Sm and LSm complexes [29, 91, 92]. The order of the monomers in the ring is 657-657. Two hexameric rings are present in the asymmetric unit of the crystal. The dimensions of the torus are approximately 33 Å in height, 54 Å in outer diameter and 11 Å in pore size.

The structure comprises residues 4-79 of LSm5, 3-73 of LSm6 and 26-99 of LSm7 which cover the canonical components of the LSm fold for all proteins: an N-terminal α-helix followed by 5 highly bent β-strands. The strands form a continuous antiparallel β-sheet that extends across the subunits by the interactions between the second half of β-4 of one monomer with β-5 of the next one. These strands are contained in the hallmark Sm2 motif, while the Sm1 motif spans β-1, β-2 and β-3. The topology of the LSm domains is an SH3-like barrel that is closed by the crossover of the β-5 strand over the highly bent β-sheet. In addition to trans-β-strand contracts, intersubunit interactions are mediated by the N-terminal

helix that stacks onto the β-sheet of the neighboring subunit at the proximal side of the ring. Side chain contacts between subunits involve salt bridges and hydrophobic interactions.

Interestingly, the number of contacts in the two native interaction surfaces is higher than in the non-native interface between LSm6 and LSm7. Analysis of the structure of LSm657 by the PISA server (Protein Interfaces, Surfaces and Assemblies) supports this finding as fever hydrogen bonds and electrostatic interactions are detected.

The crystal structure includes eight Zn^{2+} ions that are bound to each hexameric complex in our structure. Four of these ions are located at the surface, one at the Lsm6-LSm5 interface and three are cluster at the interface between LSm6 and LSm7. This interaction most likely results from high concentrations of zinc acetate in the crystallization condition, which was present at 100 mM.

3.1.1.2. Differences between LSm proteins on the sequence level are not reflected in structure.

All copies of the LSm proteins adopt the canonical LSm fold. C^{α} RMSDs of the Sm motifs of identical LSm proteins are very small (0.20 Å, 0.58 Å, 0.13 Å respectively for LSm5, LSm6 and LSm7). The larger RMSD of the four copies of LSm6 results from a different conformation of loop 2. Also RMSD values between the mean structures of the individual LSm proteins are small ranging between 0.75 Å and 1.34 Å.

The similarity between corresponding pairs of Sm and LSm proteins that is easily observed on the sequence level is not mirrored in structural similarity. The RMSDs between pairs of corresponding LSm and Sm proteins were not significantly lower than these of unrelated pairs. Previously, RMSDs between human Sm, archaeal Lsm and bacterial Hfq proteins had been reported to be below 1.5 Å for the conserved Sm1 and Sm2 motifs, indicating a stable, rigid and conserved fold. Interestingly, the RMSDs are low independent from the oligomeric state of the complexes in question. The features that give rise to assembly of the LSm proteins in their specific order and to RNA specificity are thus not coded in overall topology. Rather, the side chains of contributing positions located on the rigid LSm scaffold dictate specificity by forming the complementary interaction surfaces of the native ring.

3.1.1.3. RNA binding properties of LSm657

The pore of the LSm657 is composed of residues from loop 3 and loop 5. It is positively charged suggesting a role in RNA binding. Previous structural studies on the archaeal homoheptameric Sm1 protein have revealed interaction with a uridine nucleotide in this region [92]. Similarly, in the complex between human Sm ring and RNA the RNA passes through the pore. All Sm proteins in this structure interact with consecutive bases in a slightly

different fashion. In addition to binding to 3' ends, the LSm complexes have also been reported to interact more transiently with RNA than their Sm counterpart. This is reflected in varying conservation of positions in the nucleotide binding pocket in Sm and LSm proteins: these range from virtually complete conservation to substitution to chemically similar residues. These differences in conservation support alternative sequence specificities for the Sm and LSm rings but at the same time complicate sequence prediction for LSm interacting RNAs.

3.1.2. NMR spectroscopic analysis of the LSm657 complex

NMR studies were carried out to confirm the quaternary structure of the LSm657 complex in solution. These experiments were prompted several considerations: Firstly, previous work has demonstrated that LSm proteins can adopt different oligomeric states in crystals than in solution [23]. Furthermore, the ring shaped assembly requires interruption of two interfaces upon formation of the native heptameric LSm rings from the LSm657 hexamer. Thus, the observed quaternary state might hamper assembly and should thus be confirmed independently. Finally, we asked whether the presence of zinc ions in the buffer influenced protein-protein interfaces and thus the composition of the complex.

3.1.2.1. LSm57 is a dimeric LSm subcomplex in solution

Following a divide and conquer strategy aimed to investigate LSm657 in solution, we first studied the LSm57 complex that does not assemble into a higher oligomeric state. LSm57 being dimeric demonstrates an inability of β-strand 5 of LSm7 to interact with the β-strand 4 of LSm 5 and to form a non-native LSm75 interface. A complete backbone assignment was obtained for this complex using a U-[^{13}C, ^{15}N,^{1}H] labeled sample. Analyzing secondary chemical shifts demonstrated that secondary structure elements in the complex are present as predicted for the canonical LSm fold. Additionally, short amide-amide distances observed in NOESY spectra are in very good agreement with the substructure of LSm57 that was extracted from the LSm657 structure. Together, these data demonstrate that both the fold and the mode of dimerization of LSm57 are similar in solution as in the crystal.

3.1.2.2. LSm657 is a canonical LSm hexamer in solution

In order to study the LSm657 complex, sample preparation and pulse sequences had to be adapted to a 60kDa protein complex. Therefore, fully deuterated samples and transverse relaxation optimized spectroscopy (TROSY) experiments were used. A full assignment of the LSm657 complex was performed, as assignments from the LSm56 complex could not be transferred due to a large number of chemical shifts between the

spectra. Also for the LSm657 complex, the secondary chemical shifts indicate the presence of the canonical LSm secondary structure elements. ^{15}N-edited NOESY and TROSY-^{15}N (t1)-NOESY-TROSY ^{15}N (t2) – ^{1}H(t3) spectra demonstrate that almost all short-range amide-amide distances predicted from the crystal structure are observable, including intermolecular cross-strand contacts. These experiments confirm the formation of a stable, hexameric LSm657 complex in solution in absence of zinc acetate. Hence, no artificial intramolecular contacts are introduced by interaction with the zinc ions. Together, we demonstrate the agreement between the crystallized quaternary structure and the solution state for LSm657.

3.1.2.3. LSm monomers adopt different angles in LSm rings.

The contacts that differ between the LSm57 dimer and the LSm657 hexamer are the interfaces between LSm7 and LSm6 and between LSm5 and LSm6. Thus, a change in the chemical environment and thus chemical shifts are expected for residues in these areas. Comparing the spectra of LSm57 and LSm657, significant chemical shift perturbations in these areas are detectable, yet also regions distant from these site display chemical shift changes. We argue that these small but significant effects result from an effect that globally affects protein structure. In a dimeric assembly, LSm proteins are predicted to adopt an angle of 51° as is the case for dimeric Sm intermediates [93]. This angle is optimal for their further assembly into a heptamer. Contrarily, in the hexameric LSm657 complex, the subunits are expected to interact in a 60° angle, which we confirm in our crystal structure. We argue, that the change in the spectra results from subunits interacting in different angles in both complexes. The strain induced by the hexameric assembly might be compensated for by formation of the two non-native LSm67 interfaces. This observed angular plasticity has also been reported for the *A. fuldigus*LSm2 protein, which can assemble in hexamers and heptamers depending of pH, ionic strength and the presence of RNA [91, 94].

3.1.2.4. Extensions at N- and C-termini of LSm7 are likely unstructured

LSm7 contains long presumably unstructured N- and C-terminal extensions. These are readily cleaved during proteolytic treatment during sample preparation for crystallization. NMR experiments show that these sequences are most likely disordered in solution, as resonances corresponding to these residues are absent from NMR spectra.

3.1.3. LSm657 is a functional assembly intermediate

3.1.3.1. LSm23 can be incorporated into LSm657

To probe whether LSm657 is a functional assembly intermediate, we examined whether a subsequent assembly step could be recapitulated *in vitro*. Pull-down experiments on His-tagged LSm657 incubated with untagged LSm23 were performed to assess the ability

of these LSm proteins to assemble into a stable subcomplex. Indeed, LSm23 could be copurified with LSm657 after incubation. This demonstrates that LSm23 strongly interacts with LSm657 and forms a complex containing all five LSm proteins. Interestingly, when using an excess of LSm23, the resulting complex contains an increased amount of LSm23 indicating the formation of a heptameric LSm23-23-657 assembly. Contrary, LSm56 is not able to copurify LSm23 indicating that the LSm7 interface is required to mediate the interaction with LSm23. This is in agreement with the position of LSm23 adjacent to LSm7 in the native ring.

To gain molecular detail on the assembly reaction we observed its procession by NMR. Spectra of U-[^{15}N,^{1}H] labeled LSm657 were recorded and the incorporation of unlabeled LSm23 monitored over time. We find that LSm23 interacts strongly with LSm657 as demonstrated by strong chemical shift changes in LSm657. These changes occur on a slow exchange regime over a period of hours indicating a process with high activation barrier. This argues for the fact that the observed interaction is indeed a remodeling of LSm proteins into a common ring and not mere binding of LSm23 on top of the hexameric LSm657 ring.

3.1.3.2. LSm657 experience angular rearrangements upon LSm23 incorporation

Again, upon incorporation of LSm23, large differences in the spectra of LSm657 occur. These are, as observed previously for the LSm56 and LSm657 complexes, affecting more resonances than the interfaces involved. There, we attributed the extent of the shifting resonances to a changing angle between subunits upon change of oligomeric state. Also upon incorporation of LSm23, the number of LSm proteins in the ring changes. This might induce a change in LSm angle resulting in chemical shift changes from angular rearrangement in LSm657 in addition to those from direct interaction. However, since no assignment is available for the LSm23-23-657 complex, the location of the interacting surfaces with LSm23 cannot be mapped to LSm657 to validate the topology in LSm23-23-657. Nevertheless the fewer contacts between LSm7 and LSm5 argue for a weaker interface that would more easily dissociate and accommodate LSm23 subunits. This is in agreement with the location of LSm23 next to LSm7 in the native complex.

3.2. Discussion

3.2.1. The LSm657 assembly intermediate is a canonical LSm complex.

We here report the structure of the hexameric LSm657 assembly intermediate. The structure displays the features of a canonical, ring-shaped LSm multimer in sequence, topology and tertiary and quaternary structure. The subunit order in the hexamer is 657-657. The Sm intermediate SmFEG that corresponds to the LSm657 complex also forms a hexamer *in vivo* and has been shown by biochemical experiments to adopt that subunit order (Plessel et al., 1997, Kambach et al., 1999). This complex has a physiological role in Sm ring formation. It is recognized by the SMN protein and by successive addition of other preformed assembly intermediates (SmBD3, SmD1D2), the functional Sm ring is assembled around the snRNA [7]. Our data demonstrate that the LSm657 complex is an intermediate in LSm assembly.

3.2.2. LSm interactions are promiscuous yet allow for specific assembly

The interfaces between LSm6 and LSm5 and between LSm5 and LSm7 in the LSm657 complex are present in the final LSm1-7 and LSm2-8 rings. In contrast, the interface between LSm7 and LSm6 in LSm657 occurs only transiently during assembly. Other non-native LSm interfaces are present in stable complexes and have been reported. These include LSm6-LSm7, Lsm3-LSm2, LSm2-LSm7 and LSm3-LSm3 interactions [10, 23, 27]. Other interfaces, however, do not stably associate. In this work, we have shown that LSm57 is a dimer in solution indicating that β-4 of LSm5 is not able to stably interact with β-5 from LSm7. The main interactions between LSm subunits are hydrogen bonds between the β-4 and β-5 of neighboring subunits and are thus sequence unspecific. Yet, a specific order of subunits is maintained in Sm and LSm rings. The information must thus be coded in other contacts in the interface, which constitute complementary protein interfaces.

Unfortunately, comparing the native and non-native interaction surfaces between the LSm proteins in the LSm657 complex doesn't reveal the residues that are required for specific assembly. Previous work has suggested, that different interactions occur between subunits in assembly intermediates compared to the native complex [29, 93]. Hence, direct comparison with the final LSm1-7 or LSm2-8 complex from S. pombe might elucidate the molecular details of specificity, however these structures are currently not available. Recent structural studies have characterized the *S. cerevisiae* LSm1-7 and LSm2-8 rings that differ from *S. pombe*.

Additionally, direct measurement of interface stability would provide data on the correlation between stability and specificity in LSm assembly. Mutational analysis can point

out crucial residues that specify the sequence of assembly events and final subunit order in LSm and Sm complexes.

3.2.3. LSm complexes accommodate their monomers at different angles

Sm and LSm proteins have previously been shown to assemble into diverse multimeric complexes. The native LSm1-7 and LSm2-8 complexes are heptameric rings, but stable rings with different stoichiometries are known as well. The closely related bacterial Hfq, SmFEG and LSm657 are both hexameric while the structure of an octameric LSm3 complex has been solved [23, 26, 27, 95]. Angular plasticity is a common feature among Sm and LSm proteins of which the molecular basis is yet unknown.

In this work we observe that the spectra of the subunits in the LSm57, LSm657 and LSm23-657 complexes differ substantially. Chemical shift changes occur not only at resonances corresponding to sites close to the interaction surface but also in more remote locations. We interpret these smaller but significant chemical shifts changes as changes in the angular rearrangement upon forming a ring of different size. As the angle in the dimeric LSm57 is not restrained, it is possible that it will adopt the same angle as the dimeric protein complexes SmBD2 and SmD1D2. These have been reported to adopt angles close to 51°, which is the angle in a native heptamer. The ring-shaped hexameric LSm657 adopts an angle of 60° between its subunits. Thus a change of oligomeric state between dimer and hexamer would be consistent with angular change and thus with the remote chemical shift perturbations between LSm57 and LSm657.

The oligomeric state of the intermediate LSm23-657 can be either pentameric or heptameric depending on whether one or two units of LSm23 are incorporated. Both our biochemical data and the absence of pentameric LSm assemblies suggest the heptameric state. In both cases however, the intersubunit angle changes upon incorporation of LSm23, which is consistent to the additional chemical shift changes we observe in the spectra upon LSm23 incorporation.

How do LSm proteins achieve this angular plasticity? The rearrangement could occur via changes in the arrangement of secondary structure elements or by repacking the side chains in the interfaces. Our NMR data suggests that changes upon incorporation of subunits are propagated to more distant locations in the protein. This would argue for a rearrangement throughout the secondary structure elements. However, if the oligomeric state was mainly coded in the secondary structure, the C^{α} RMSD values in heterooligomers should reflect mainly the oligomeric state of a complex. There, subunits in the same oligomeric state, e.g. in heptamers, should be similar and differ from protomers in a hexameric state. This is however not the case.

Interestingly, it has been shown that the interactions contributing to dimerization in Sm subcomplexes are different to those present in the final Sm ring [29, 93]. This finding argues for a side chain rearrangement in the interfaces although the angular change is small between these complexes. In LSm complexes, the angular change of 9° per interface upon transition from a hexameric to a heptameric complex results in a maximal change in distance of 4 Å between two subunits. Repacking side chains can span this distance. Furthermore, side chain rearrangement may occur additionally to a tighter packing of secondary structure elements.

It remains to be seen how the transition from hexameric to heptameric intermediates influences the stability of the complex and whether the change in oligomeric state serves functional purposes during assembly.

3.2.4. LSm657 is a functional assembly intermediate - the assembly order of Sm proteins is conserved in LSm complexes.

The assembly of the Sm ring has been well studied [7]. After the snRNA is transported into the cytoplasm, preformed subcomplexes (SmFEG, SmBD3, SmD1D2) of the seven Sm proteins are deposited successively on the snRNA by the SMN protein. The order of assembly in Sm proteins is SmFEG, followed by SmD1D2 and SmBD3. SmFEG has been shown to be a hexameric assembly intermediate [96]. This suggested that the homologous hexameric LSm657 complex performed a similar function during LSm assembly. Indeed, in our vitro experiments have shown that under physiological conditions, LSm657 can incorporate LSm23, the homologue of SmD1D2 *in vitro* and thus retrace the assembly pathway of the Sm ring. This finding is in agreement with the notion that proteins are under evolutionary pressure to retrace ordered assembly pathways [97].

In vitro, the binding of LSm23 into LSm657 takes place on a timescale of hours. This indicates that the process has a relatively large activation barrier, arguing for an incorporation event rather than simple binding on a preexisting surface. Incorporation of LSm23 into hexameric LSm657 would require the opening of two LSm interfaces, which would be in agreement with a slow process. The remodeling of the ring that we describe is rather slow, but could *in vivo*be accelerated by the presence of RNA, chaperones and LSm1/LSm4/LSm8 that are incorporated into the LSm ring at a later stage.

Considering the conservation between Sm and LSm proteins, other features of the complexes may be similar, too. Our structure of the LSm657 assembly intermediate shows that N- and C-terminal extensions that are predicted to be unstructured are indeed not visible in electron density. NMR experiments support the conclusion that these sequences are unstructured as no resonances corresponding to these amino acids were observed, in agreement with fast exchange of amide protons with the solvent in unstructured regions. In

the SmD1D2 and SmBD3 subcomplexes, regions outside the Sm fold also appear unstructured but fold and interact with RNA in the context of the complete Sm ring [93]. Hence, the extensions of LSm7 could serve a similar purpose in the LSm ring.

3.2.5. Prediction of RNA binding properties of the LSm657 assembly intermediate

In the LSm657 assembly intermediate, the pore of the ring is positively charged, suggesting a role in RNA interaction. While the smallest Sm assembly with reported RNA binding ability is SmD1D2-FEG [32] corresponding to LSm23-657, the putative RNA binding surface of LSm657 can be inspected and compared with the known mode of RNA interaction of Sm proteins.

Sm and LSm rings bind different RNAs in the cell: The Sm ring is locked on the snRNA by virtue of double-stranded segments on both sides of the recognition sequence. The interaction of LSm rings with RNA is at the 3' termini and is thought to be more transient. These mainly contain a cyclic phosphate in case of the Lsm2-8 ring [98] that interacts with U6 snRNA or a hydroxyl group for the LSm1-7 complex that recognizes the 3' end of mRNAs. The rings thus differ widely in the details of their mode of interaction impairing the prediction of binding preference of the LSm657 segment.

The crucial residues for RNA interaction in the LSm domain are located in loops 3 and 5 on the inside of the pore. In the canonical nucleotide binding site, an RNA base is stacked between an aromatic residue (H37, numbering according to *A. fuldigus* LSm) and a highly conserved arginine residue (R63). The Watson-Crick edge of the base is read out by hydrogen bonds to a conserved asparagine (N39) conveying specificity for uracil bases. R63 also contacts the phosphate moiety of the previous protomer whereas D65 sandwiches between the phosphate groups of the backbone. These hallmark residues are highly conserved in the ring-forming LSm proteins 1-8 with two exceptions: R63 in LSm5 is substituted for an asparagine. Additionally, the canonical aromatic residue that stacks with the base in LSm7 is substituted for a leucine.

These observations suggest, that RNA interaction differs in LSm5 and LSm7 compared to the Sm binding mode. In the other LSm proteins the canonical sites are conserved, suggesting that their mode of interaction is more similar to that of the Sm ring, which features sequential interaction with internal bases. Overall, however, prediction of RNA interaction based on our structure remains elusive.

Recently, the structure of the LSm2-8 ring from *S. cerevisiae* in complex with 5'-GCUUUU-3' and 5'-GCUUU-3' has been determined [24]. These structures show that the 3'-terminal uridine nucleotide is bound by LSm3 and the subsequent bases interact with LSm2, LSm8 and LSm4. This finding has been contested as the native RNA contains five terminal

uridine nucleotides and terminates often in a 2'-3'-cyclic phosphate [99]. It has been speculated that the additional uridine and the cyclic phosphate could be accommodated by LSm6. Both models however agree that the non-canonical LSm5 and LSm7 binding sites are not involved in sequential interaction with RNA.

3.3. Outlook

Future work on LSm ring assembly can be categorized in two parts: Theoretical considerations on the assembly mechanism and studies of assembly intermediates and assembly reactions *in vitro* and *in vivo*.

3.3.1. Prediction of order in cyclic LSm complexes

The specificity of the order in the ring is coded in the complementarity of the interfaces and thus has to be predictable *a priori*from the sequence. Data on which interaction surfaces form in solution or in a crystal, and which do not, is available. However, there is no quantitative information on the energies of the interfaces that might be related to the specific formation of complexes. This hinders prediction of assembly intermediates. However, other information can be exploited to narrow down the residues that mediate specific interaction: LSm proteins and Sm proteins show pairwise similarity on the sequence level, but have not been reported to form mixed complexes. Thus, specificity has to be achieved with residues that differ between corresponding Sm and LSm proteins. Also, LSm1 and LSm8 act in different complexes but integrate at the same location in the ring. Thus, residues relevant for specificity have to be the same in these proteins ensuring their identical location. This work can make use of computational methods as well as assessment of the capacity of mutant proteins to assemble in the ring. This endeavor will profit from high-resolution structures of assembly intermediates and native rings from the same species to compare the contributions of single residues in the interfaces throughout assembly.

3.3.2. Angular plasticity in LSm proteins

Along these lines, it will be worth investigating, how the angular flexibility of LSm domains is mediated. The oligomeric state of Sm and LSm proteins ranges from bacterial Hfq that constitutively form hexamers to the archaeal Sm2 protein can form homohexamers and homoheptamers and eukaryotic Sm and LSm proteins that vary in oligomerization during assembly. It is unknown at present, what determines the oligomerization state of LSm and Sm rings and what its physiological function is. It is also unknown, whether the different oligomerization states of LSm subcomplexes is solely a result of interface promiscuity or whether it has a function in defining the energy of the intermediates along the pathway

towards the native ring. Again, high-resolution structures of different intermediates will help uncover the molecular details behind angular plasticity.

3.3.3. LSm complex assembly and functions

Having established that LSm rings can undergo assembly *in vitro*, it will be interesting to investigate the further assembly pathway. We have shown that LSm657 can incorporate LSm23 subcomplexes. The molecular structure of LSm23-657 is currently unknown, although the known assemblies comprise hexamers and heptamers indicating that an LSm23-23-657 complex is more likely than a pentameric LSm23-657. Subsequent work will characterize the following assembly pathway leading to the native LSm1-7 and LSm2-8 rings. Previous data demonstrates that the formation of these rings is correlated suggesting a coordinated assembly or a common precursor [20]. Future insight into the assembly mechanisms of both rings will indicate whether availability of LSm1-7 and LSm2-8 links RNA degradation with splicing.

Our work also forms the basis of future structural analysis of LSm rings. While the molecular structure of the LSm2-8 ring in complex with RNAs and of the LSm1-7 ring in complex with Pat1p from *S. cerevisiae*have been solved, the mode of 3' end recognition for both rings are still unclear. For LSm2-8 the mode of 2'-3' cyclic phosphate recognition is unknown as are the features defining a 3' end of mRNA that are recognized by LSm1-7.

LSm proteins are building blocks for a vast variety of rings. In addition to LSm1-7 and LSm2-8, also the hexameric LSm2-7 has been described that is involved in pseudouridination of rRNA [100]. Moreover, a variety of additional LSm proteins exist that have been reported to assemble into rings and are unrelated to RNA binding [101]. Studies addressing their assembly in rings will facilitate the investigation of their function and reveal principles of specific multimer assembly.

4. An excess of catalytically required motions inhibits the scavenger decapping enzyme

4.1. Aims and significance

This thesis is concerned with understanding the molecular details underlying two steps in mRNA decay: The specific assembly of the LSm1-7 and LSm2-8 ring and the relevance of conformational dynamics for enzymatic function in the scavenger decapping enzyme Dcs1p. Addressing these questions will shed light on both mechanistic details of mRNA decay and on the principles governing specific protein complex assembly and enzyme dynamics in general..

4.1.1. How do Dcs1p dynamics give rise to its function?

At the beginning of this thesis, the structure and function of scavenger decapping enzymes had been characterized. These data and MD simulations implied a crucial role for conformational change in its catalytic cycle. Models for the DcpS catalytic cycle were however limited by a lack of experimental data of its protein dynamics: The structural characterization relied on interpolation between static states and the kinetic analysis on interpolation between known rates of binding and hydrolysis to describe the conformational change, yet no direct measurements of its rate or its trigger were available.

The absence of quantitative data of enzyme dynamics is a common hindrance in understanding their molecular mechanism. Especially data on large proteins is scarce. NMR spectroscopy provides experimental means to investigate protein dynamics on various timescales and recent advances allow for application to large proteins. Thus, the goal of this part of my thesis was to measure and characterize the dynamics in the yeast scavenger decapping enzyme during catalysis. Thus, several questions are addressed:

- What are the structural and kinetic details of Dcs1p conformational dynamics?
- In which state (substrate-bound, product bound, singly or doubly occupied) do Dcs1p motions occur?
- What are the molecular determinants in the protein and the ligand that give rise to Dcs1p motions?
- What is the binding mode of the substrate m^7GpppG to Dcs1p?
- Do Dcs1p dynamics contribute to its enzymatic function, in particular to the phenomena of allostery, substrate inhibition and mutant hyperactivity?

To address these questions, a biophysical approach was used that combined NMR spectroscopy with x-ray crystallography, ITC and activity assays to obtain a comprehensive picture of Dcs1p function via dynamic, structural, functional and binding data.

The significance of this work is twofold: Understanding the relevance of Dcs1p conformational dynamics will shed light on the mechanism of an important enzyme in mRNA degradation, how its substrate preferences and modes of inhibition are mediated. Also it will improve the understanding of its other cellular functions. Moreover, this work will improve the general notion of how protein dynamics can shape enzyme function. The concept of dynamically mediated allostery or substrate inhibition may be a general mechanism and thus expand our understanding of protein function.

4.2. Results

4.2.1. The structure of S. cerevisiae Dcs1p in complex with m^7GDP

We solved the structure of inactive *S. cerevisiae* Dcs1p H268N in complex with the inhibitor m^7GDP by x-ray crystallography to a resolution of 2.3 Å. Initial phases were obtained by molecular replacement using a search model based on the structure of human DcpS in complex with m^7GpppG (1ST0). Phases were improved by iterative rounds of model building and refinement. The model is in good agreement with experimental data and has good stereochemistry.

The model of yeast Dcs1p in complex with m^7GDP encompasses residues 7-343 of both chains. The overall structure is very similar to asymmetric structures of the human and mouse proteins. The dimeric C-terminal domain is adopts a HIT pyrophosphatase fold and dimerizes via a large shared surface. The N-terminal domain dimerizes via a strand swap where one protomer donates two β-strands to the β-sheet in the N-domain of the other protomer. The topology is identical to human and mouse proteins except for a 14 amino acid insertion in the N-terminal domain that is structured in the crystal but does not adopt a secondary structure. This stretch participates in crystal contacts and includes a disulfide bond. We conclude that without the stabilizing contacts and in the reducing environment of the cell, this loop is very likely unstructured.

Yeast Dcs1p in complex with m^7GDP adopts an asymmetric conformation that encloses the ligand tightly in a closed active site by interactions with the C-terminal domain, the hinge region and the N-terminal domain. Both domains behave as rigid bodies so that the residues involved in the ligand binding in the closed site are located at a distance in the open protomer. Interestingly, the second active site is empty.

For the human protein the C-terminal domain contributes most of the interaction surface with the ligand,as reported previously. Comparing the conformations of the interacting residues in the C-terminal domain for the open and closed active sites, the C-domain of the closed active site resembles most to that of either the closed or open m^7GpppN bound state. In contrast, the empty open site is most similar to the either empty open site in inhibitor bound structures or to the product bound open site. The differences between these conformations mainly affect Y264 the nearby residues S263 and W161. Y264 is part of the triad that mediates recognition of the substrate methyl group and domain closure. Interestingly, the conformations of the catalytic residues are similar, raising the question why the open protomer is inactive. Additionally, the model of the open protomer lacks the loops 169-173 and 196-200 for which no interpretable electron density was obtained.

The structure features three ordered sulfate ions. Due to the high concentration of ammonium sulfate, which is used as a precipitant during crystallization, lower affinity binding sites of sulfate are occupied in this condition. Two sulfate ions are located in the active sites respectively and make contacts with residues that in the substrate bound human DcpS enzyme interact with the triphosphate bridge. The third sulfate ion is located beyond the binding site of the second base of the substrate and might thus interact with a site that would coordinate the phosphate moiety of the RNA body in longer substrates. The residues interacting with the first two sulfates are conserved, while the residues that interact with the third sulfate are not.

Overall the structure of yeast Dcs1p is very similar to these of human and mouse scavenger decapping enzymes. It is noteworthy that in this structure, only the closed active site is occupied with m^7GDP.

4.2.2. Labeling and assignment strategies for NMR experiments on Dcs1p

NMR experiments on larger proteins require specialized strategies to reduce signal loss due to relaxation and to reduce spectral complexity due to spectral overlap. Methyl-TROSY pulse sequences select for less slowly decaying component of magnetization and thus improve signal intensity for larger proteins. The methyl TROSY experiments we exploit here were carried out on uniformly $^1H^{13}C$ Ile-δ and Met-ε labeled and otherwise deuterated samples. The benefit of using methyl groups as probes is threefold: Three equivalent protons provide triple signal to noise, and using methyl groups reduces spectral overlap while still providing well distributed probes throughout the protein. Moreover, the effective correlation time of the nuclei is reduced due to the high flexibility of the hydrophobic residues used for this labeling scheme, resulting in slower relaxation.

To relate information from recorded spectra to the molecular events on the protein level, each peak needs to be assigned to the residue that gives rise to it. In case of isoleucine and methionine residues, assignment was carried out by a mutational approach. Mutation of relevant residues to a similar amino acid, such as leucine, which does not get specifically labeled, results in disappearance of the respective peak, which is thus assigned.

In case a peak of interest was to be assigned, prior knowledge was used to speculate on the identity of the corresponding amino acid: Comparing spectra from separate domains, the location of a residue in either the N- or C-terminal can be deduced. Additionally, chemical shift perturbation of a residue upon mutation of another one argues for proximity of the two. Finally, the x-ray crystal structure of yeast Dcs1p we solved previously was used for structure based chemical shift prediction with shiftx [102] in order to make an educated guess on residue identity. Subsequent validation by mutating the candidate residues allowed for unambiguous assignment.

Using this strategy, a set of residues was assigned that covers 49% of isoleucine and 33% of methionine methyl groups. The assigned residues are distributed over both domains, the hinge region and the active site and thus provide a comprehensive set of probes throughout the protein.

4.2.3. Dcs1p binds two substrate molecules sequentially in non-equivalent sites

The presence of only one molecule of m^7GDP in our structure of yeast Dcs1p argues for non-identical mechanism of ligand binding to the two sites. Also, the structures of DcpS proteins in complex with two ligand molecules have revealed strikingly different modes of interaction for the two sites. The closed site is formed by N-terminal, C-terminal and hinge domain and encloses the substrate m^7GpppN tightly. In contrast, the open site consists of the C-terminal domain interaction surface with few additional contacts between the flexible second base and the hinge domai. Differing both in number and character of contacts, the structures suggested different strength of interaction. Given the anti-cooperative nature and substrate inhibition of scavenger decapping enzymes, a more complex mode of interaction than independent binding to the two active sites had to be suspected.

In order to elucidate the order of events in DcpS substrate binding, we conducted NMR and ITC experiments to obtain information on both affinities and the residues involved in binding.

NMR titration experiments on inactive Dcs1p H268N protein and the substrate m^7GpppG revealed that substrate binding occurs in two subsequent steps.

In the free state, the number of peaks corresponds to the number of isoleucines and methionines in one Dcs1p chain. This demonstrates that Dcs1p is a symmetrical homodimer

in which a residue encounters the same chemical environment in both chains. Additionally, the free spectra correspond well to an overlay of the separate N- and C-terminal domains indicating no significant interaction between the domains in the free state. Both findings are in agreement with the published crystal structures of free human and mouse DcpS proteins.

During titration of Dcs1p with m^7GpppG, at first large overall changes in the spectrum occur until reaching a molar ratio of 1:1 between Dcs1p dimer and ligand. The number of peaks increases significantly in a slowly exchanging process. The intensity of newly appearing peaks increases while a portion of previously present peaks decreases in intensity. These findings indicate a global conformational transition upon binding of one substrate molecule per Dcs1p dimer, consistent with the reported asymmetric state of ligand-bound Dcs1p. The conformational change influences a larger part of the protein than that involved in ligand interaction. As the protein adopts an asymmetric conformation, the same residue in the binding and free protomer experiences a different chemical environment. Interestingly, some peaks split into two new positions, indicating that the symmetrical state differs from both the open and closed conformation in the asymmetric state. In contrast, some peaks split, with one peak displaying identical chemical shift as the symmetric peak. This indicates that, in many respects, one conformation is similar to the symmetric one. Referring to the structure, this is the case for the open protomer.

When titrating to an excess of more than one substrate molecule per Dcs1p dimer, again, chemical shift perturbations occur. These affect few peaks and take place on a fast exchange regime, while the chemical shift changes occurring due to protein asymmetry are still present. The newly occurring chemical shift perturbations can be mapped to residues close to the open active site in the C-domain and to interacting residues in the hinge and in the N-terminal domain close to the hinge. In contrast, the resonances corresponding to closed active sites do not experience further chemical shift changes. These findings indicate a secondary substrate-binding event in which a second molecule of m^7GpppG interacts with asymmetric Dcs1p and is bound to the open active site.

The NMR titration experiments reveal a sequential binding mode for Dcs1p with a nanomolar first binding event that is concomitant with symmetric to asymmetric transition and a second binding event to the open active site that is estimated to be of micromolar affinity.

In order to obtain thermodynamic parameters, isothermal titration calorimetry (ITC) was used to further characterize the sequential binding of m^7GpppG to inactive Dcs1p H268N. This inactive but otherwise wt protein did not yield interpretable isotherms as an additional endothermic process hindered analysis. Rationalizing that the cause might be an aggregation event, mutants and conditions were screened for interpretable isotherms. Isothermal titration data of Dcs1p K126A H268N were free of the additional endothermic event and were interpreted using a sequential binding model. In this mutant, a lysine residue

contacting the β-phosphate group in the closed active site, is replaced by an alanine. It displays the same conformational change and secondary binding site as the wild-type. As the overall properties of the protein are the same, the thermodynamic parameters are a good estimate for the wild-type although they may differ and underestimate the true values. The fitted affinities for the first and second binding site are 240 ± 90 nM and 280 ± 30 µM respectively.

The C-terminal domain contributes the main surface for the second binding. To investigate the relative contribution of N- and C-terminal domains to overall binding, we also characterized the interaction between separately expressed C-domain (residues 130-350) and m^7GpppG by NMR spectroscopy and ITC. NMR demonstrated that the same residues mediate substrate binding to the C-domain in the separate domain as in the full-length context (residues I165 and I210). Some additional shifts in the full length Dcs1p protein are present in the N-terminal domain (I42 and I116). These are located in the hinge and adjacent side chains. The ligand in the open site cannot make direct contact with all of these residues in the N-terminal domain. The hinge region interacts directly with the substrate but structural changes expand into the N-terminal domain (see later paragraphs).

Taken together, our x-ray structure of Dcs1p in complex with one molecule of m^7GDP suggests the presence of a singly occupied state. Subsequent titration experiments via NMR agree with published structures of free, singly and doubly occupied DcpS proteins. We show here that Dcs1p follows a sequential binding mechanism with affinities in the nano- and micromolar range.

4.2.4. The interaction between the methyl group and Y94 is required for formation of the asymmetric conformation and catalysis.

The previously determined structures of human DcpS in complex with substrates are all captured in the closed asymmetric conformation. The proximity of catalytic residues and of residues with influence on activity suggest that the formation of this state is required for catalytic activity, The network mediating the interaction between the domains encompasses a set of hydrogen bonds around tyrosine 94 that recognizes the methylation on the RNA cap structure. In order to assess the relevance of this interaction for the formation of the closed conformation we recorded spectra of the Y94A mutant in presence and absence of m^7GpppG and of wt Dcs1p in presence and absence of the unmethylated cap analog GpppG. In both instances, we observed an interaction of the ligand with the enzyme in the C-domain part of the active site as demonstrated by shifting resonances of reporter peaks (I165, I210). However, we found no indication of an asymmetric state upon disruption of the interaction between Y94 and the methyl group. We then determined the activity of wt Dcs1p on GpppG and of Dcs1p Y94A on m^7GpppG and found them to be inactive. Thus we demonstrate that

both establishment of a closed conformation and activity depend on the interaction between the methyl group of the cap substrate and Y94.

4.2.5. The substrate m^7GpppG can displace the product m^7GMP from Dcs1p

Asymmetric crystal structures contain either substrates (m^7GpppG, m^7GpppA) or inhibitors (m^7GDP, 5-quinazolines) in the closed active site. To test whether completed catalysis is sufficient to open the closed active site and release the product, we performed NMR titration experiments with the product m^7GMP on Dcs1p H268N. The hallmark peaks that report on domain rearrangement display similar shifts in presence of m^7GMP indicating that Dcs1p undergoes symmetric to asymmetric transition upon binding of m^7GMP. Not all shifts are identical to the m^7GpppG titration experiment, which is expected for a chemically different interaction partner. Thus, m^7GMP can sustain the asymmetrical conformation of Dcs1p indicating that hydrolysis of the substrate is not sufficient to induce conformational change back to the symmetric state and thus product release from the closed active site.

To address whether interaction with m^7GpppG in the second catalytic site can induce product release from Dcs1p, we introduced a reporter isoleucine residue (L279I) in the active site of Dcs1p that probes the presence of m^7GMP or m^7GpppG in the active site. We then conducted experiments in which Dcs1p is converted to the asymmetric product bound form by addition of m^7GMP. Then, the titration is continued with m^7GpppG and the reporter peak clearly indicates that the product is displaced from the closed binding site and that then substrate is bound in a closed active site.

These results demonstrate that m^7GpppG can displace m^7GMP from the closed active site. This can occur via a mechanism where m^7GMP leaves the closed active site according to its off-rate via a symmetric state and is replaced by the more tightly binding m^7GpppG. Alternatively, m^7GpppG could induce a flip-over motion of the N-terminal domain upon interaction with the open protomer. This result is consistent with a model in which product release from a recently active catalytic site is coordinated with substrate binding in the open active site.

4.2.6. Dcs1p undergoes millisecond timescale motions under substrate excess

Scavenger decapping proteins in complex with substrates or inhibitors enclose their ligand in a secluded closed active site. It is evident from their crystal structures that they have to undergo conformational change after turnover to release products and bind substrate molecules for the next round of catalysis. Quantitative data on Dcs1p motions are required to understand the catalytic cycle and to evaluate molecular models for its cooperativity and substrate inhibition.

Amongst other methods, protein dynamics on the millisecond time-scale can be observed by longitudinal exchange NMR experiments. In this experiment, peaks originating from different conformations of a single NMR active residue that interconvert on the millisecond time-scale display cross peaks. When a residue changes conformation from open to closed during a delay time between detection of chemical shifts in the direct and indirect dimension in the NMR experiment, its gives rise to a cross peak. This peak displays the chemical shift of the closed conformation in the dimension that is measured first (here ^{13}C) and of the open conformation in the other dimension (^{1}H). Observation of the cross peaks unambiguously correlates peaks arising from the same residue in different conformations. Additionally, the frequency of interconversion and the relative populations of interconverting states can be extracted by examining cross peak intensity in dependence of delay time. Together, this method allows for detection of protein dynamics, identification of the residues involved and quantitation of protein motions.

To examine the motions in Dcs1p occurring during the catalytic cycle, we conducted longitudinal exchange experiments on the inactive mutant Dcs1p H268N under varying concentrations of the substrate $m^{7}GpppG$. The inactive mutant was used to prevent substrate turnover during data acquisition, which is required to obtain meaningful data on kinetic parameters. Experiments were carried out at 298 K, which is the optimal growth temperature of yeast, to ensure physiological conditions.

Indeed, cross peaks could be observed for all well separated and exchanging residues,indicating molecular motions under conditions of substrate excess. Peaks exhibiting cross peaks, and thus a conformational change, are located both in the N- and C-terminal domain. These protein movements are energy independent, as no source of chemical energy was present in the experimental conditions.

4.2.7. Millisecond timescale motions correspond to N-domain flipping

Close examination of the longitudinal exchange data provides information on the nature of the observed molecular motions in Dcs1p.

Peaks corresponding to the open and closed protomer in the asymmetric state display cross peaks in between each other. Hence, the protein is undergoing a conformational change in which a closed active site is opened while the other one is closed. This constitutes a flipping motion of the N-domain. Several other lines of argumentation consolidate this interpretation: Firstly, all exchange rates fitted to well separated sets of auto and cross peaks are similar. Thus, the molecular motions constitute one global process in which all residues exchange conformation at the same rate, consistent with N-domain flipping.

Secondly, the populations of interconverting states robustly fit to $pA = pB = 0.5$. By definition, in a process in which the two protomers in a homodimer exchange conformations,

the two states each have a relative prevalence of 0.5. This argues against an additional state with which the two asymmetric conformations might interchange.

Moreover, the position of cross peaks argues against an exchange between asymmetric and symmetric states. In peaks that display different chemical shifts for both asymmetric states and the symmetric state, exchange peaks appear between the positions of the asymmetric closed and asymmetric open peak and not between these two and the position of the symmetric protein.

Taken together, position of the observed cross peaks and fitted exchange parameters clearly demonstrate that Dcs1p undergoes flipping motions of its N-terminal domain under conditions of substrate excess. The movements correspond to an alternating opening and closing of the both active sites. Interestingly, the motions observed occur in presence of excess substrate and thus do not correspond to the release of product with concomitant substrate binding. Contrarily, these motions occur in absence of product and thus reflect events during substrate binding.

Dcs1p is a homodimer, so the end-point states of this motion are structurally identical and only differ in which of the identical protomers is closed. Structurally, this end-point state is likely very similar to the asymmetric conformation that has been trapped in crystals of human DcpS in complex with m^7GpppN and yeast Dcs1p in complex with m^7GDP.

Moreover, this finding has implications for the interpretation of previous titration experiments. The observed motions are present under excess substrate conditions including NMR titration experiments. Hence, the chemical shifts that were observed in the N-terminal domain upon interaction with the second site result from the dynamic domain rearrangement and not from direct interaction.

4.2.8. The frequency of Dcs1p motions depends on substrate excess

We performed longitudinal exchange experiments at various substrate excesses to investigate the influence of substrate concentration on protein dynamics, particularly in the light of Dcs1p substrate inhibition. Experiments were carried out at Dcs1p dimer to m^7GpppG ratios of between 1:1 and 1:20. Millisecond timescale flip-over motions occurred at all concentrations. Quantitation of these motions revealed that with increasing ligand, also protein dynamics increased up to a level of saturation. Measured frequencies of flipping range from below 5Hz to 35 Hz.

Flipping motions occur in asymmetric Dcs1p and their frequency increases with substrate concentration. Previous titration experiments have shown that Dcs1p interacts sequentially with two molecules of substrate. Interaction with the first interaction site does not result in Dcs1p motions. At the substrate concentration where flipping-motions were observed, titration experiments demonstrated that the substrate binds to the second binding

pocket of Dcs1p. Thus we conclude that interaction with the second binding site allosterically induces conformational change in Dcs1p.

4.2.9. Flipping is faster than catalysis and mutually exclusive with it

The human scavenger decapping enzyme DcpS has been shown to be inhibited under conditions of excess substrate. It is noteworthy that inhibition coincided with conditions of increased flip-over motions. To investigate this relationship, we determined catalytic rates of Dcs1p under multiple turnover conditions were determined using ^{31}P NMR experiments.

The progress of m^7GpppG hydrolysis can be monitored by following substrate and product concentrations over time. In the phosphorous 1D spectrum, Pα and Pγ of m^7GpppG exhibit identical chemical shifts, but are well resolved from the peaks resulting from Pβ and the phosphates in the products m^7GMP and GDP. Thus, substrate turnover can be monitored by following the decrease in substrate peak intensity and the increase in product peak intensity over time. Turnover rates were obtained by fitting the progress curve with an integrated Michaelis-Menten equation.

The resulting v_{max} value for wt Dcs1p is 0.12/s, which is in good agreement with previously published values for human DcpS. Under single-turnover conditions, where flipping motions are absent, human DcpS can perform catalysis an order of magnitude faster, with rates of >1.2 s^{-1}. We thus asked whether frequent conformational change could decrease catalytic activity.

Comparing rates of dynamics and catalysis, the flipping of Dcs1p is 300 times faster than catalysis under comparable experimental conditions. Structurally, the N-domain flipping motion opens an occupied closed active site to the open conformation, while concomitantly a liganded open site is closed. As the closed conformation is required for catalysis, its temporary disruption could be inhibiting protein activity. Under maximally inhibited conditions, the mean time that Dcs1p remains in one asymmetric conformation is 30 ms. If a following step in catalysis is slower and requires substantially more time than 30 ms for completion, it will likely be interrupted by domain rearrangement. In the Michaelis-Menten model of enzyme catalysis, ligand binding is followed by the chemical step. Catalysis itself however is commonly a fast process. Thus, we speculate that excess protein motions hamper an additional slow step that takes place after binding. This rate limiting process might be the active site rearrangement in which ligand and the closed active site form a productive complex.

4.2.10. In Dcs1p K126A activity is increased while dynamics are decreased

Previously, we established that in wild-type Dcs1p, frequent conformational change induced by substrate excess is correlated with reduced catalytic activity. If an excess of

protein motions is detrimental for Dcs1p activity, a mutant with reduced dynamics should exhibit increased activity. Thus, we investigated the correlation between dynamics and function further by examining the dynamics in a designed hyperactive variant of Dcs1p.

For DcpS, a number of hyperactive mutants have been reported before. Here, we additionally describe the hyperactive Dcs1p K126A mutant. In this variant, a lysine residue in the hinge region is exchanged to an alanine residue. In the closed protomer, the ε-amino group makes an electrostatic interaction with the β-phosphate group; in the open protomer, it is close to the second base of the substrate and exposed to the solvent. We established the hyperactivity of Dcs1p K126A in a ^{31}P NMR degradation experiment and determined a maximal turnover rate of 0.48 s^{-1}, which is four times as active as the wild type.

NMR spectra of the inactive variant Dcs1p K126A H268N recorded in the presence of m^7GpppG show that the mutant undergoes symmetric to asymmetric transition and interacts with two substrate molecules. Additionally, ITC experiments have demonstrated a sequential binding mechanism for Dcs1p K126A H268N. Regarding its substrate binding mode, the K126A mutant thus behaves as the WT protein. To obtain information on protein dynamics, we performed longitudinal exchange experiments at excess substrate concentrations equivalent to the wt experiments. Even at the highest ligand concentration, no measurable cross peaks between the two asymmetric conformations could be detected. Their absence indicates that the exchange rate lies below the detection limit of the experiment and thus is less than 0.5 Hz.

This result confirms the inverse relationship between flipping motions and activity under substrate excess in Dcs1p. While asymmetry and flipping motions are required for substrate binding and product release, their excess abundance reduces activity. This mechanism of substrate inhibition can however be alleviated by reduction of protein motions as shown in the mutant Dcs1p K126A.

4.3. Discussion

In this work, we characterize and quantify the dynamics occurring during the Dcs1p catalytic cycle. From structural information, a conformational change during substrate binding and product release had been postulated. The trigger of this conformational change as well as its timescale and its function in the catalytic cycle however remained unknown and subject to debate.

Using near-residue resolution solution NMR, we are able to address several mechanistic questions of Dcs1p function on the molecular level. Firstly, we investigated the mode of substrate interaction and its connection to Dcs1p cooperativity. In the course thereof we demonstrate that Dcs1p conformational change is a rigid body domain movement. The main focus of this work however is on the quantitation of protein motions and their function in catalysis and substrate inhibition.

4.3.1. The conformational change of Dcs1p upon substrate binding

Previous structural studies suggested that both domains in DcpS proteins act as rigid bodies when undergoing conformational change. This notion was expanded by molecular modeling simulations in absence of substrate or product [103]. In the free form, DcpS displays fluctuating domain motions on the nanosecond timescale by distortion of its flexible hinge. The simulations demonstrate that DcpS domains move towards an asymmetric and more closed arrangement but do not close the active site as in the substrate bound state. Additionally, the rigid body model postulates negative cooperativity in DcpS as activity in one site is mutually exclusive with the second, which was verified by kinetic studies [65].

Here, we present experimental evidence for the assumed N-terminal domain rigidity upon substrate interaction throughout the catalytic cycle. Our data demonstrate that upon substrate or product binding, Dcs1p transitions to an asymmetric state. NMR spectra confirm that the asymmetric conformation is present in solution in complex with m^7GMP, m^7GDP and m^7GpppG. In contrast, the product GDP does not induce asymmetry or any measurable interaction in NMR. This finding stresses the requirement for the N-7 methylated base for conformational change and the major contribution of the methylated base to affinity.

4.3.2. The binding mode and affinities of Dcs1p for substrates and products

Numerous studies have reported affinities of scavenger decapping enzymes for its substrate, substrate analogues, inhibitors and products [44, 54, 65, 104-107]. Affinities for substrates and substrate analogues are in the very low nanomolar range, for the inhibitor m^7GDP in the low nanomolar range and for the product m^7GMP in the high nanomolar range. The structure of inactive human DcpS in complex with its substrates displays strikingly

different binding modes for the two active sites and thus indicates different affinities for the first and second binding site. Nevertheless, occupancies of both sites were equal in the crystal structures. Thus it remained an open question, how tightly Dcs1p interacts with substrate molecules in the closed and open site. In the light of Dcs1p substrate inhibition, it is important to obtain quantitative data on individual substrate binding to both active sites. This information is relevant to obtain a complete mechanistic description of the Dcs1p catalytic cycle.

Here, we applied NMR and ITC to this question. Using NMR, we found that m^7GpppG, m^7GDP are indeed bound with different affinities in the two active sites in inactive Dcs1p H268N and follow a sequential binding mechanism. Affinities in the first site are in the nanomolar range for substrate and inhibitor and in the high micromolar range in the second site. The interaction sites are in agreement with crystal structures of doubly liganded DcpS proteins except for some peaks in the N-terminal domain that shift upon ligation of the second active site that are distant to the binding site. This effect is discussed further in the following section.

To obtain quantitative information, we conducted ITC experiments. Isotherms of inactive Dcs1p H268N titrated with m^7GpppG displayed three enthalpic events, including an additional endothermic process that prevented data analysis. We speculated that this process corresponds to an aggregation event during titration. The Dcs1p K126A H268N mutant did not display aggregation or the endothermic process and. yielded interpretable isotherms. To assess the comparability of this mutant, we demonstrated by NMR that overall topology, conformational change and the sequential binding mechanism are identical to wt protein, however this mutant displays increased activity. Fitting of ITC data yielded a Kd of 240 ± 90 nM for the closed site and 280 ± 30 µM for the second site of Dcs1p H268N K126A using a sequential model for fitting. Interestingly, the sequential binding model is a special case of binding based on anti-cooperativity as only upon ligation of a first site the second site becomes available.

In comparison to previous affinity measurements, our value for first-site affinity for inactive Dcs1p K126A towards m^7GpppG is in the high nanomolar range compared to the low nanomolar values reported for wt human DcpS towards non-hydrolysable cap analogs. The reported affinities were determined in experiments on active DcpS protein and non-hydrolysable cap analogs. Thus, these affinities are expected to diverge from ours due to the different chemical nature of ligand and active site. Indeed, the affinities of wt and inactive DcpS H277N have been reported to differ [65]. More importantly however, our ITC experiment was set up to quantify a nanomolar and a micromolar affinity in parallel. In ITC, the experimental protein concentration needs to be matched to the strength of the interaction. Here, the concentration was chosen to match both interactions acceptably and

thus is not ideal for obtaining data of the high affinity first site, which is reflected in the standard deviation of the Kd of the closed site.

Regardless, the experiments clearly demonstrate the sequential binding mode in yeast Dcs1p and yield a quantification of the second site interaction. Our findings are in accordance with structures that feature equally occupied open and closed active sites. The high protein and ligand concentrations present in protein crystals allow for interaction at even lower affinity sites such as the open active site of scavenger decapping proteins. Our structure of Dcs1p was obtained with preformed complex from which less tightly bound ligand was removed in a gel-filtration step, resulting in occupancy at the closed site only.

We thus confirm previous results with our binding data on closed binding sites, and further demonstrate that affinities for open active sites of scavenger decapping enzymes are in the high micromolar range following a sequential binding mode.

4.3.3. Substrate m^7GpppG can displace the product m^7GMP from Dcs1p

We have shown by NMR that interaction with m^7GMP induces the asymmetric conformation in Dcs1p, but no interaction with the open site could be detected. This is further corroborated by ITC experiments that detected interaction between product and Dcs1p only in one site with a Kd of 1 ± 0.64 µM. The Kd of the product to the open active site is undetectable and thus above the detection limit of both NMR and ITC. Thus, the product m^7GMP is readily released from an open active site. This is in agreement with a mechanism in which conformational change simultaneously allows for substrate binding in a closed site and for product release from an open one. Interaction of m^7GMP with an open site has been found in the crystal structure of human DcpS in complex with m^7GDP. Again, high concentrations of protein and ligand in crystals can account for the saturation of very weak binding sites.

Additionally, these data demonstrate that the presence of the Pβ conveys drastically increased affinity of a ligand to both closed and open binding site, with nanomolar vs. micromolar affinities in the closed site and micromolar vs. undetectable affinities in the open site. The residues interacting with the Pβ mediate the discrimination between the product m^7GMP and the inhibitor m^7GDP. In the closed site, these residues comprise K126, K196, S263, K287 and H259 which are highly conserved. This is indeed expected for resdues defining the substrate range of an enzyme.

Having demonstrated that m^7GMP can induce the asymmetrical conformation in Dcs1p, we went on to establish that the substrate m^7GpppG could release the product. In a mutant that contains a reporter peak to distinguish between substrate- and product-bound proteins, we showed that product-bound Dcs1p is converted to substrate-bound Dcs1p upon addition of m^7GpppG. Two mechanisms of exchange are conceivable:, Firstly, product release can

occur due to conformational change induced by interaction with m^7GpppG in the open active site, so that this open site is closed and m^7GMP can exit the then open site. Alternatively, in a mechanism of competition, m^7GMP can be released according to its off-rate and the free protein binds m^7GpppG preferentially due to its higher affinity. While our data are consistent with both models, NMR experiments on substrate bound Dcs1p have shown that interaction of m^7GpppG with the open site can induce conformational change. Thus it is likely, that the substrate can also induce conformational change when m^7GMP, a weaker ligand is bound to the first site.

Taken together, these results demonstrate that Dcs1p has to undergo conformational change in presence of its substrate and products during the catalytical cycle, and suggest that this conformational change might be induced by interaction of the substrate with the open active site.

4.3.4. Futile motions in Dcs1p mediate substrate inhibition

Dcs1p displays substrate inhibition [65], a phenomenon that is common among enzymes [69]. The mechanism by which substrate inhibition is conferred is known for only a small number of enzymes and mostly associated with steric interference of a secondary, either orthosteric or allosteric binding site with catalysis or product release. These insights are usually obtained by structural investigations. Here, we explore the contribution of protein dynamics to Dcs1p substrate inhibition.

Our work provides qualitative and quantitative data on Dcs1p domain movements during the catalytic cycle. We demonstrate that interaction of substrate with the second binding site induces flip-over motions of the N-terminal domain in a concentration dependent manner. The frequency with which the conformational change occurs increases with substrate concentration. Conceptually, these motions are required for substrate binding and product release after catalytic turnover. However, the motions are likely mutually exclusive with ongoing catalysis as the productive active site is dismantled during each flip-over movement.

We determined the frequencies of flip-over motions in inactive Dcs1p H268N protein and find that they occur under substrate excess at magnitudes higher than catalysis under comparable conditions, 35 Hz k_{ex} vs. 0.12 Hz k_{cat}. This finding suggests that overly frequent conformational change as it occurs under substrate excess reduces catalytic efficiency compared to the single turnover condition. This finding provides a rationale for the mechanism of substrate inhibition in Dcs1p. By induction of N-domain flipping at an unproductive frequency, interaction of m^7GpppG with the open active site reduces catalytic efficiency. This interpretation is supported by previous experiments that demonstrated that a heterodimeric DcpS protein that is unable to interact with substrate in one protomer displays

reduced substrate inhibition [65]. The closed active site of this protein has to be constituted by the binding-competent protomer while the open site is the binding incompetent one. Hence, due to strongly reduced interaction with the substrate in the active site, also flipping would be reduced in this protein variant, leading to increased activity under multiple turn-over conditions.

We support our argumentation by analysis of the hyperactive mutant Dcs1p K126A. This mutant exhibits strongly reduced flipping motions compared to the wild-type protein (<0.5 Hz vs. 35 Hz) under the same conditions. This finding confirms the inverse correlation between fast flipping motions in Dcs1p and the catalytic activity under multiple turnover conditions. Additional supporting data for this correlation stems from a previous study that has examined how rigidifying the hinge region by mutation to proline affects activity [41]. Three amino acids in the hinge region (human DcpS: aa 145-147) mediate the conformational change between open and closed state by changes to their φ and ψ angles. Clearly, reducing their flexibility to an extent that Dcs1p cannot adopt the asymmetric conformation will result in loss of activity. However, attenuating flip-over motions to a smaller extent would abolish futile motions and increase DcpS activity according to our model. Indeed, DcpS R145P and DcpS Q146P exhibit 113% and 142% activity compared to the wild type, respectively. Thus, reduced flexibility in the hinge region is not associated with decreased, but with increased activity. These data support our model, however the actual rates of conformational change remain to be determined. Also, the DcpS R145A mutant yields an even higher activation than R145P, for an additional function of this residue. This aspect is further discussed in section 4.1.6.

Together, these data argue that movements that are essential to catalysis can impede enzyme function in scavenger decapping enzymes when occurring in excess.

4.3.5. The rate limiting step in the catalytic cycle of Dcs1p

In substrate-inhibited enzymes, the rate-limiting step varies between low-substrate and high-substrate conditions. In the classical model of substrate inhibition, the doubly substrate-bound enzyme ES_2 complex exhibits a different k_{cat} than the ES complex and thus features different activity. It has been shown that the rate-limiting step in Dcs1p differs between single and multiple turnover conditions. Identifying the respective rate limiting steps and elucidating how the presence of substrate excess can induce a switch between them is necessary to understand substrate inhibition in Dcs1p on the molecular level.

The rate-limiting step under single-turnover conditions is the initial binding step [65]. Interestingly, under multiple-turnover conditions this shifts to a later event in the catalytic cycle. In presence of excess substrate, the observed rate constant was reduced by a factor of 7.7 compared to single turnover conditions.

Our data supports the finding from pre-steady state kinetics that product release is not the rate-limiting step in Dcs1p [65]. If product release were the slowest step, enhancing protein flipping motions would activate DcpS and limiting dynamics would further inactivate DcpS. In contrast, a mutant with reduced N-domain mobility (Dcs1p K126A) and thus a reduced capacity to release product, is more active than the wild-type protein.

By means of elimination, a step between substrate binding and product release had been proposed as the limiting step. The conformational change before catalysis in the doubly liganded protein had been considered a candidate and assumed to be either slow or unfavorable [65]. This hypothesis was supported by the reduced substrate inhibition in heterodimeric DcpS that contained one binding-compromised protomer arguing that this protein could not undergo the conformational change and thus increase activity by using only one active site. This experiment supported that inhibition takes place via a state with occupied second active site.

We complement and expand the kinetic work by structural and dynamic data on the domain movement of doubly liganded protein in solution. Our data indicates that the asymmetric to asymmetric transition (35 Hz) is faster by two orders of magnitude than the observed rate constant (0.12 Hz). Hence, conformational rearrangement can be excluded as slowest step. Additionally, NMR spectra indicate that inactive Dcs1p protein adopts an asymmetric conformation in presence of excess substrate. Thus, conformational change cannot be unfavorable, as then the symmetric conformation would be significantly populated and observable in the spectra. Using NMR spectroscopy we can exclude that a symmetrical conformation is populated significantly in solution. Thus, the conformational change is not slow or unfavorable.

We therefore propose an additional rate-limiting step in the Dcs1p catalytic cycle that occurs after substrate binding, yet before product release. The catalytic step cannot be excluded, yet chemical steps are often faster than protein rearrangements. The time required for completion of the rate-limiting step has to be significantly larger than the mean time that Dcs1p samples the asymmetric state so that substrate inhibition can occur.

The side-chain conformation of the active sites is notably different between the closed and open substrate bound form on one side and the empty or product bound state on the other side. Also the substrate adopts an extended conformation in the open site as opposed to a kinked conformation of the triphosphate bridge in the closed site. Thus it is tempting to speculate that the arrangement of a productive active site might be the slowest step. These structural differences between empty open and closed site are evident from the crystal structures of DcpS in complex with various ligands, however the time they require for completion is unknown. We speculate that the rate of the reverse flipping reaction disassembling the closed active site (17.5 Hz) might be faster than the subsequent side-

chain rearrangement so that the protein would not remain in the closed conformation for a sufficient period to proceed to the productive active side state.

These rearrangements are equally necessary for catalysis under single turnover conditions. However under single turnover conditions no disassembly of the active site is induced so that then the substrate-binding step could remain limiting.With the NMR labeling scheme that we used in the current studies, we have not been able to observe this rate-limiting step experimentally.

All longitudinal exchange experiments were conducted using inactive Dcs1p H268N protein in which the catalytic histidine is replaced with an asparagine residue. Thus, progression through the catalytic cycle is not possible, as the catalytic step cannot take place. Hence, this set-up allows observation of all steps prior to catalysis. Catalysis itself however cannot be ruled out as slowest step so that our speculations remain to be tested experimentally.

In summary, our work has introduced a new step to the catalytic scheme of Dcs1p: futile flipping that follows double substrate binding and precedes successful catalysis. Thus we expand the previous notion that assuming the asymmetric conformation before catalysis is the rate-limiting step in Dcs1p under multiple turnover conditions. We suggest that this conformational change is not direct and that futile motions prevent Dcs1p from stably adopting the productive asymmetric conformation.

4.3.6. The molecular trigger for allosterically induced conformational change

Binding of one molecule of substrate leads to a conformational change and locks Dcs1p rigidly in the asymmetric state. Interaction with a second substrate molecule triggers conformational exchange between the two asymmetric states and increases flexibility in Dcs1p. These effects have opposed impacts on catalysis, as noted previously. How does the interaction of m^7GpppG with the open active site induce motion of the N-terminal domain? In classical models of allostery, conformational change can be transduced by a series of structural changes that stabilize the other state. This information could be gained from structural data capturing intermediate states or from molecular modeling. Unfortunately, no data is available on this matter for Dcs1p yet. However, existing structural and mutational data can be examined to hypothesize on the molecular mechanism of allostery in Dcs1p.

Examination of the interactions between m^7GpppG and the open and closed active site in the crystal structure of human DcpS and of residues that experience chemical shift perturbations upon substrate interaction in NMR experiments and might thus contribute to the allosteric path reveals two main sites: the C-terminal domain and the hinge region.

4.3.6.1. Allostery in the C-terminal domain of Dcs1p

The methylated base and the β- and γ-phosphate groups interact with multiple residues in the C-terminal domain. Compared with the closed binding site, the interactions of the first base and the triphosphate bridge are very similar in both sites differing only in the side-chain positions of L206 and K207.

If flipping motions were triggered and transduced through the C-terminal domain only, different the C-terminal domain should also display different binding properties for its two binding sites as interaction with the two binding sites would yield different effects on the protein. Titration of C-terminal domain with m^7GpppG however shows independent and identical binding to the two sites in the protomers in isothermal titration calorimety. Also NMR titration experiments with C-domain and m7GpppG display only one set of shifting peaks, indicating independent and identical interaction for both active sites.

These data argue against an allosteric pathway whose molecular determinants are located solely in the C-domain.

4.3.6.2. Allostery transmitted via the hinge region of Dcs1p

The second base of m7GpppG is stacked between two arginine side chains in the C-domain and the hinge region while the ribose moiety makes few contacts at all. Mutating residues in the hinge area leads to increased activity in yeast Dcs1p, as shown in this work. As hyperactivity can be achieved by relieving substrate inhibition, this could indicate involvement of the hinge region in mediating allostery in Dcs1p. However, it is unclear if the effect of these mutations is mediated by action in the closed or open protomer. This question is, generally, difficult to answer as most residues in substrate interaction in Dcs1p have different functions in both conformations. Considering the Dcs1p K126A mutant we studied, the residue contacts the β-phosphate group in the closed protomer. In the open side, it is solvent exposed but located in proximity of the location of the second base interaction site in the human protein. The second base in the open site manifests a higher level of flexibility compared to the methylated base moiety of the same cap structure as demonstrated by decreased occupancy and higher B-factors of this part in structures of human DcpS in complex with m^7GpppN [41]. Flexibility in the second base of the substrate can allow the second base to interact with additional residues in the hinge region, including K126A. For this mutant we have shown that catalytic activation occurs via decreased dynamics. To discern the effects of K126A in the open and closed site, we argue that the effects in the two sites are opposite: Disruption of the interaction between lysine 126 and the substrate in the closed site would likely not result in decreased motion, but rather in weakening of the closed conformation. In contrast, disrupting the putative interaction between the second base and K126 in the open protomer could reduce stabilization of the flipping motion. Thus, we hypothesize that in this mutant, the reduction of the dynamics is mediated by the mutation in

the open active site. A positively charged surface extends in Dcs1p from the hinge region into the N-terminal domain to K112 that contacts the carbonyl group of the second base in the closed conformation. It is conceivable that this stretch provides a continuous interaction surface along which the second base of the substrate can move. This succession of interactions could gradually stabilize the protein during conformational change as enthalpy is lost from the opening other active site. Flipping is the interconversion between two identical states that are thus of equal potential energy. As by definition, one state cannot be stabilized in respect to the other, enhancement of flipping has to occur by lowering the energy of the transition states. It is thus tempting to speculate that the activation energy of the interconversion is lowered by interaction with K126 and other charged residues. Its disruption would penalize the conformational change between the opposite asymmetric states and thus reduce flipping and ultimately alleviate substrate inhibition.

This hypothesis is supported by the finding thatin the human DcpS protein, mutations of several positively charged residues in the hinge region result in hyperactive proteins [41]. These residues however also have functions in the closed active site so that their action in the allosteric path needs to be verified.

4.3.7. The function of the N-terminal and C-terminal domains of Dcs1p

The Dcs1p C-terminal domain is similar to proteins of the HIT family of pyrophosphatases. Other HIT and HINT proteins are nucleoside phosphoramidate hydrolases, dinucleotide hydrolases or nucleotidylyl transferases. The turnover rates of HIT proteins are rather low compared with other enzymes. The k_{cat} of the most closely related FHIT protein 73 $s^{-1}$for its substrate ApppA [108], however other hydrolases of this family exhibit lower activities that are in the range of Dcs1p activity [39]. It has to be noted that HIT proteins are active on their own accord while the Dcs1p C-domain is inactive on its own even though it comprises the catalytic triad. Only in context with the N-domain in cis or trans can Dcs1p perform catalysis [37].

The N-terminal domain of Dcs1p is structurally related to NTF2-like proteins. While these proteins form heterodimers via their exposed β-strand, the homodimerization of the Dcs1p N-domain via a β-strand swap at that site is unique. The function of the N-terminal domain is twofold: Firstly, its interaction with the C-terminal domain requires the N7-methylation on the substrate to stably form the asymmetric state. The N-terminal domain thus provides specificity for methylated di-nucleotides and nucleotide polyphosphates. We have shown that Dcs1p enzymes in which the interaction between the N7-methyl group and the recognizing tyrosine residue has been impaired do not form stably closed conformations. This was achieved by either using unmethylated cap analog or the Dcs1p Y94A mutant. An interaction between enzyme and ligand can be observed, but is limited to mostly the C-

terminal domain and occurs in a fast exchange regime. The asymmetric state is either scarcely or not populated. Additionally, we demonstrated that combinations of enzyme and substrate with disrupted methyl recognition are inactive consistent with data on human DcpS [41]. These findings underline the connection between the prolonged formation of the closed conformation and catalysis. According to our model, substrate inhibition exploits the same mechanism as the Y113A mutant, by preventing the formation of a closed conformation for a sufficient period of time. In contrast, this is not achieved by prohibiting interdomain interactions, but by stabilizing the opposite asymmetric conformation. Interestingly, the triad of residues that recognizes the N7-methyl group and mediates closure is highly conserved, with the exception of *C. elegans*. In this species, Y113 and Y273 are exchanged to phenylalanine residues and H139 to leucine and are thus unable to establish the conserved hydrogen bonds. The *C.elegans* Dcs1 protein can accommodate larger functional groups in the N7 position [107]. Thus it has to specifically recognize its substrates and lock the domains onto each other by virtue of a different mechanism.

Secondly, the substrate has to adopt a different conformation in the closed site compared to in the open site due to steric constraints. It has been speculated that catalytic turnover is achieved by activation of the leaving group by distortion of the substrate. Thus, the N-terminal domain could mediate and couple closure and activity and limit both to methylated substrates.

Evolutionary acquisition of the N-terminal domain provided for a mechanism to ensure specificity and activity for methylated substrates. Here we demonstrate an additional function of the N-terminal domain, which is the allosteric mediation of substrate inhibition by undergoing futile motions in the yeast Dcs1p protein.

4.3.8. Implications of protein motions and substrate inhibition of Dcs1p *in vivo* and *in vitro*

We show that a tightly coordinated balance between necessary and unproductive motions governs Dcs1p activity. Interestingly, scavenger decapping enzymes have not evolved for optimal turnover rates as demonstrated by the presence of hyperactive variants both in human and yeast. This suboptimal turnover is realized by substrate inhibition in Dcs1p. The residues in Dcs1p that maintain dynamic substrate inhibition are conserved; also human DcpS is readily inhibited by substrate excess. Conservation of the residues involved in N-domain flipping suggests that this behavior is common to scavenger decapping enzymes. Hence, a cellular function of substrate inhibition can to be suspected. Interestingly, substrate inhibition by excessive dynamics would allow for reduced activity as well as for increasing activity when quenching these motions.

The significance of substrate inhibition in the cell is still subject to speculation. In order to establish a cellular function of Dcs1p substrate inhibition in the cell, it will be necessary to obtain data on Dcs1p dynamics in presence of short, capped mRNAs and on the abundance of the different species of substrate in the cell. While the concentration of Dcs1p in yeast under normal growth conditions is ~0.3 µM [109], the concentration of short capped RNAs and cap structure in the cell can only be estimated hampering a prediction of cellular function. As Dcs1p substrate inhibition is likely conserved between yeast and human, a cellular function can be suspected. Assuming substrate inhibition of Dcs1p in the cell, conditions where high amounts of Dcs1p substrates are generated by the exosome complex will reduce Dcs1p activity and maintain high levels of cap structure. Catalytic turnover would accelerate only once substrate concentrations lessen. Thus, both temporally and spatially, larger differences in cap structure concentrations would result from substrate inhibition. This mechanism could provide switches in cross talk to other cap dependent processes such as translation initiation or splicing.

In addition to prompting experiments to explore Dcs1p cellular functions, our findings have practical implications for future structural analyses of scavenger decapping proteins. The fact that substrate excess can induce unproductive motions in Dcs1p provides a rationale for crystallization strategies. Molecular motions and conformational heterogeneity are expected to be detrimental for crystallization. Thus, using an equimolar ratio of substrate has likely improved our crystallization results and we expect it to facilitate future crystallization of scavenger decapping enzymes from other species.

4.3.9. Futile protein motions in catalysis

With our work we describe a unique molecular mechanism of substrate inhibition by inducing motions that are crucial for catalysis at a frequencies that reduce overall activity. This result challenges the idea that protein dynamics, and conformational change in particular, are necessarily coordinated for optimal catalytic turnover. We rather find that also the dynamic properties of a protein, i.e. the frequency with which it transits between stable states, can be utilized to regulate its activity. Moreover, not only by decreasing, but also by increasing protein motions, activity can be modulated.

Our work underlines the necessity to expand our understanding of proteins and particularly enzymes on the molecular level, as conformational change is common during protein action. In addition to describing series of static states, measuring and taking into account the flexibility around these states and the rates and triggers of transition between them will be required.

4.4. Outlook

Our work prompts future experimentation in three areas: the details of substrate inhibition in Dcs1p, the relevance thereof in the interplay between Dcs1p and its other interaction partners and most generally the contribution of dynamics in other substrate inhibited proteins.

4.4.1. The molecular details of Dcs1p dynamics in allostery

Our results demonstrate that unter multiple turnover conditions, conformational change is two orders of magnitudes faster than the overall catalytic rate in Dcs1p and thus not rate-limiting. Considering previous investigations ruling out substrate binding and product release, we hypothesize that the rearrangement of the closed active site into a productive state or the catalytic step could be rate limiting. Future work would need to probe these hypotheses in order to gain complete understanding of the catalytic cycle. Complete description ideally would also provide rate constants for the transitions between all observable states and allow a statistical description of the catalytic cycle.

To further confirm the dynamic mechanism of substrate inhibition, longitudinal exchange experiments on other hyperactive mutants could further establish that the general potential for hyperactivity of Dcs1p is due to futile motions in excess substrate states. Thus, the timeframes in which Dcs1p motions are beneficial or detrimental could be determined. Additionally, single-turnover activity assays on hyperactive mutants are required to confirm that their hyperactivity is founded in alleviation of substrate inhibition and not in generally enhanced catalytic turnover.

Considering the molecular basis of the interplay between structure, dynamics and function in Dcs1p, we have proposed a model in which conformational change is induced by electrostatic interaction between a the second base of a substrate molecule in the secondary active site and the charged residues in the hinge region. Mutations in this region have been shown others and us to increase Dcs1p activity. However, activity of these mutations has been measured in homodimeric protein so that despite rational argumentation, the effect of the mutation cannot be pinpointed to either the active or the allosteric site. Thus, future experimentation on heterodimeric Dcs1p samples will provide protomer-specific data on functional residues. Furthermore, determining the flipping rates in presence of more Dcs1p substrates, such as m^7GTP or m7GpppA will reveal molecular details of the trigger for unproductive dynamics on the substrate side.

To study the events underlying allostery in Dcs1p, molecular dynamics simulations of the rearrangements in the second binding site upon interaction with the substrate will be very instructive. Results thereof can subsequently be tested and validated for both activity and dynamic properties in established experiments.

4.4.2. The effect of Dcs1p dynamics on its cellular functions

Scavenger decapping proteins interact in the cell with their substrates and a number of inhibitors, such as m^7GDP and 5'-quinazolines. Moreover, they contribute to other cellular processes including sugar metabolism, splicing and miRNA degradation. Future work will have to address whether the dynamic mode of substrate inhibition we describe here influences other Dcs1p functions than mRNA decapping.

4.4.2.1. Dcs1p dynamics in substrate and ligand interactions

Recent work has shown that DcpS is inhibited effectively by 5'-quinazolines that mimic the RNA cap structure and that its inhibition influences SMN splicing [54-56]. Combined with our mechanistic results on substrate inhibition, this poses the question, whether these inhibitors also exert their function via futile protein motions.

In the cell, Dcs1p acts on short, capped mRNA fragments whose mRNA moieties are shorter than 10bp. The next step in understanding the dynamic nature of Dcs1p action would be to determine how Dcs1p structure and dynamics are affected by its larger substrates. A previous model states that the presence of an increasing number of bases poses an entropic penalty on active site closure. Evidently, crystallographic analysis of a state of enhanced dynamics is difficult, however the system may be amenable to NMR. As methyl group assignments are available, both the interaction surface of the additional bases as well as the effect on protein dynamics could be investigated. The positively charged surface in the hinge region of the open protomer is located in the proximity of the negatively charged RNA and would be a plausible interaction surface.

4.4.2.2. The significance of Dcs1p dynamics for protein-protein interactions

In yeast, Dcs2p is a closely related homolog of Dcs1p that is expressed under stress conditions. Even though the relevant catalytic residues are conserved in Dcs2p, it is inactive and inhibits Dcs1p by heterodimerization. Interestingly, two of these residues are mutated in Dcs2p a conserved homologue of Dcs1p that has been shown to inhibit Dcs1p function by dimerization.

The most obvious differences between Dcs1p and Dcs2p are the mutations K126G and K287M. Both residues interact with the β-phosphate group of the substrate. We expect that Dcs2p differs in its binding preferences for substrate and product from Dcs1p and may thus exhibit different kinetic properties and thus modulate these of the inhibited heterodimer Dcs1/2p. However, mutation of K126 residue to alanine in Dcs1p results in decreased motions and increased instead of activity. The effect of this mutation differently affects the two protein variants and is, based on our results, likely to modulate protein dynamics.

Dcs1p moonlights in other cellular functions. It has been reported to contribute to modulation of RNA degradation by Xrn1 activation, sugar metabolism and regulation of splicing events. In the light of the mobility of scavenger decapping enzymes and its relevance for enzymatic turnover, it would be interesting to further investigate how its protein dynamics influence these other functions. *In vitro* experiments can evaluate the requirement of asymmetric, mobile or static Dcs1p for these reported interactins.

In this work we have demonstrated advances in understanding the molecular basis of substrate inhibition of scavenger decapping enzymes. Its function in the cell, however, is unclear. It has been suggested previously, that action of Dcs1p reduces the concentration of methylated nucleotides that can be incorporated in nucleic acids and thus reduce the lifespan of yeast. By expanding the repertoire of mutants that lack inhibition under multiple turnover conditions, mutant yeast can be generated and analyzed for methylated bases in nucleic acids, accumulation of cap structure and activity of cap dependent process. Thus, the physiological significance of Dcs1p substrate inhibition can be elucidated.

5. References

1. Schoenberg, D.R. and L.E. Maquat, *Regulation of cytoplasmic mRNA decay.* Nature reviews. Genetics, 2012. **13**(4): p. 246-59.
2. Sun, M., et al., *Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels.* Molecular cell, 2013. **52**(1): p. 52-62.
3. Wang, Y., et al., *Precision and functional specificity in mRNA decay.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(9): p. 5860-5.
4. Herrick, D., R. Parker, and A. Jacobson, *Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae.* Molecular and cellular biology, 1990. **10**(5): p. 2269-84.
5. Shyu, A.B., M.E. Greenberg, and J.G. Belasco, *The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways.* Genes & development, 1989. **3**(1): p. 60-72.
6. Buschmann, J., et al., *Identification of Drosophila and human 7-methyl GMP-specific nucleotidases.* The Journal of biological chemistry, 2013. **288**(4): p. 2441-51.
7. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome.* Nature reviews. Molecular cell biology, 2014. **15**(2): p. 108-21.
8. Seraphin, B., *Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs.* The EMBO journal, 1995. **14**(9): p. 2089-98.
9. Sharif, H. and E. Conti, *Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover.* Cell reports, 2013. **5**(2): p. 283-91.
10. Wu, D., et al., *Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation.* Cell research, 2014. **24**(2): p. 233-46.
11. Albrecht, M. and T. Lengauer, *Novel Sm-like proteins with long C-terminal tails and associated methyltransferases.* FEBS letters, 2004. **569**(1-3): p. 18-26.
12. Tritschler, F., et al., *A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting.* Molecular and cellular biology, 2007. **27**(24): p. 8600-11.
13. Bouveret, E., et al., *A Sm-like protein complex that participates in mRNA degradation.* The EMBO journal, 2000. **19**(7): p. 1661-71.
14. Tharun, S., et al., *Yeast Sm-like proteins function in mRNA decapping and decay.* Nature, 2000. **404**(6777): p. 515-8.
15. Tharun, S., *Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay?* RNA biology, 2009. **6**(3): p. 228-32.
16. Haas, G., et al., *HPat provides a link between deadenylation and decapping in metazoa.* The Journal of cell biology, 2010. **189**(2): p. 289-302.

17. Totaro, A., et al., *The human Pat1b protein: a novel mRNA deadenylation factor identified by a new immunoprecipitation technique.* Nucleic acids research, 2011. **39**(2): p. 635-47.
18. Braun, J.E., et al., *The C-terminal alpha-alpha superhelix of Pat is required for mRNA decapping in metazoa.* The EMBO journal, 2010. **29**(14): p. 2368-80.
19. Achsel, T., et al., *A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro.* The EMBO journal, 1999. **18**(20): p. 5789-802.
20. Luhtala, N. and R. Parker, *LSM1 over-expression in Saccharomyces cerevisiae depletes U6 snRNA levels.* Nucleic acids research, 2009. **37**(16): p. 5529-36.
21. Tritschler, F., et al., *Similar modes of interaction enable Trailer Hitch and EDC3 to associate with DCP1 and Me31B in distinct protein complexes.* Molecular and cellular biology, 2008. **28**(21): p. 6695-708.
22. Collins, B.M., et al., *Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs.* Journal of molecular biology, 2001. **309**(4): p. 915-23.
23. Naidoo, N., et al., *Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae: implications for Lsm ring organisation and recruitment.* Journal of molecular biology, 2008. **377**(5): p. 1357-71.
24. Zhou, L., et al., *Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA.* Nature, 2014. **506**(7486): p. 116-20.
25. Zhou, L., et al., *Crystal structure and biochemical analysis of the heptameric Lsm1-7 complex.* Cell research, 2014. **24**(4): p. 497-500.
26. Wu, D., et al., *Crystal structures of Lsm3, Lsm4 and Lsm5/6/7 from Schizosaccharomyces pombe.* PloS one, 2012. **7**(5): p. e36768.
27. Mund, M., et al., *Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings.* Journal of molecular biology, 2011. **414**(2): p. 165-76.
28. Leung, A.K., et al., *Use of RNA tertiary interaction modules for the crystallisation of the spliceosomal snRNP core domain.* Journal of molecular biology, 2010. **402**(1): p. 154-64.
29. Leung, A.K., K. Nagai, and J. Li, *Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis.* Nature, 2011. **473**(7348): p. 536-9.
30. Pellizzoni, L., J. Yong, and G. Dreyfuss, *Essential role for the SMN complex in the specificity of snRNP assembly.* Science, 2002. **298**(5599): p. 1775-9.
31. Chari, A., et al., *An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs.* Cell, 2008. **135**(3): p. 497-509.
32. Raker, V.A., G. Plessel, and R. Luhrmann, *The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro.* The EMBO journal, 1996. **15**(9): p. 2256-69.
33. Salgado-Garrido, J., et al., *Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin.* The EMBO journal, 1999. **18**(12): p. 3451-62.
34. Nuss, D.L., et al., *Detection in HeLa cell extracts of a 7-methyl guanosine specific enzyme activity that cleaves m7GpppNm.* Cell, 1975. **6**(1): p. 21-7.
35. Nuss, D.L. and Y. Furuichi, *Characterization of the m7G(5')pppN-pyrophosphatase activity from HeLa cells.* The Journal of biological chemistry, 1977. **252**(9): p. 2815-21.

36. Wang, Z. and M. Kiledjian, *Functional link between the mammalian exosome and mRNA decapping.* Cell, 2001. **107**(6): p. 751-62.
37. Liu, S.W., et al., *Functional analysis of mRNA scavenger decapping enzymes.* RNA, 2004. **10**(9): p. 1412-22.
38. Liu, H. and M. Kiledjian, *Scavenger decapping activity facilitates 5' to 3' mRNA decay.* Molecular and cellular biology, 2005. **25**(22): p. 9764-72.
39. Lima, C.D., M.G. Klein, and W.A. Hendrickson, *Structure-based analysis of catalysis and substrate definition in the HIT protein family.* Science, 1997. **278**(5336): p. 286-90.
40. Wypijewska, A., et al., *Structural requirements for Caenorhabditis elegans DcpS substrates based on fluorescence and HPLC enzyme kinetic studies.* The FEBS journal, 2010. **277**(14): p. 3003-13.
41. Gu, M., et al., *Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity.* Molecular cell, 2004. **14**(1): p. 67-80.
42. Liu, H., et al., *The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases.* The EMBO journal, 2002. **21**(17): p. 4699-708.
43. Salehi, Z., et al., *A nuclear protein in Schizosaccharomyces pombe with homology to the human tumour suppressor Fhit has decapping activity.* Molecular microbiology, 2002. **46**(1): p. 49-62.
44. Wypijewska, A., et al., *7-methylguanosine diphosphate (m(7)GDP) is not hydrolyzed but strongly bound by decapping scavenger (DcpS) enzymes and potently inhibits their activity.* Biochemistry, 2012. **51**(40): p. 8003-13.
45. van Dijk, E., H. Le Hir, and B. Seraphin, *DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(21): p. 12081-6.
46. Malys, N. and J.E. McCarthy, *Dcs2, a novel stress-induced modulator of m7GpppX pyrophosphatase activity that locates to P bodies.* Journal of molecular biology, 2006. **363**(2): p. 370-82.
47. Cohen, L.S., et al., *Nematode m7GpppG and m3(2,2,7)GpppG decapping: activities in Ascaris embryos and characterization of C. elegans scavenger DcpS.* RNA, 2004. **10**(10): p. 1609-24.
48. Cougot, N., S. Babajko, and B. Seraphin, *Cytoplasmic foci are sites of mRNA decay in human cells.* The Journal of cell biology, 2004. **165**(1): p. 31-40.
49. Lall, S., F. Piano, and R.E. Davis, *Caenorhabditis elegans decapping proteins: localization and functional analysis of Dcp1, Dcp2, and DcpS during embryogenesis.* Molecular biology of the cell, 2005. **16**(12): p. 5880-90.
50. Malys, N., et al., *The 'scavenger' m7GpppX pyrophosphatase activity of Dcs1 modulates nutrient-induced responses in yeast.* Nucleic acids research, 2004. **32**(12): p. 3590-600.
51. Shen, V., et al., *DcpS scavenger decapping enzyme can modulate pre-mRNA splicing.* RNA, 2008. **14**(6): p. 1132-42.
52. Bosse, G.D., et al., *The decapping scavenger enzyme DCS-1 controls microRNA levels in Caenorhabditis elegans.* Molecular cell, 2013. **50**(2): p. 281-7.
53. Sinturel, F., et al., *Activation of 5'-3' exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast.* Proceedings of the National

Academy of Sciences of the United States of America, 2012. **109**(21): p. 8264-9.
54. Singh, J., et al., *DcpS as a therapeutic target for spinal muscular atrophy.* ACS chemical biology, 2008. **3**(11): p. 711-22.
55. Gogliotti, R.G., et al., *The DcpS inhibitor RG3039 improves survival, function and motor unit pathologies in two SMA mouse models.* Human molecular genetics, 2013. **22**(20): p. 4084-101.
56. Van Meerbeke, J.P., et al., *The DcpS inhibitor RG3039 improves motor function in SMA mice.* Human molecular genetics, 2013. **22**(20): p. 4074-83.
57. Uetz, P., et al., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. **403**(6770): p. 623-7.
58. Ito, T., et al., *A comprehensive two-hybrid analysis to explore the yeast protein interactome.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4569-74.
59. Piper, P., *Differential role of Hsps and trehalose in stress tolerance.* Trends in microbiology, 1998. **6**(2): p. 43-4.
60. De Mesquita, J.F., A.D. Panek, and P.S. de Araujo, *In silico and in vivo analysis reveal a novel gene in Saccharomyces cerevisiae trehalose metabolism.* BMC genomics, 2003. **4**(1): p. 45.
61. Schepers, W., et al., *In vivo phosphorylation of Ser21 and Ser83 during nutrient-induced activation of the yeast protein kinase A (PKA) target trehalase.* The Journal of biological chemistry, 2012. **287**(53): p. 44130-42.
62. Han, G.W., et al., *Crystal structure of an Apo mRNA decapping enzyme (DcpS) from Mouse at 1.83 A resolution.* Proteins, 2005. **60**(4): p. 797-802.
63. Chen, N., et al., *Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping.* Journal of molecular biology, 2005. **347**(4): p. 707-18.
64. Martin, J., M.V. St-Pierre, and J.F. Dufour, *Hit proteins, mitochondria and cancer.* Biochimica et biophysica acta, 2011. **1807**(6): p. 626-32.
65. Liu, S.W., et al., *Mechanistic and kinetic analysis of the DcpS scavenger decapping enzyme.* The Journal of biological chemistry, 2008. **283**(24): p. 16427-36.
66. Freiburger, L., et al., *Substrate-dependent switching of the allosteric binding mechanism of a dimeric enzyme.* Nature chemical biology, 2014. **10**(11): p. 937-42.
67. Cooper, A. and D.T. Dryden, *Allostery without conformational change. A plausible model.* European biophysics journal : EBJ, 1984. **11**(2): p. 103-9.
68. Reed, M.C., A. Lieb, and H.F. Nijhout, *The biological significance of substrate inhibition: a mechanism with diverse functions.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2010. **32**(5): p. 422-9.
69. Chaplin, M.F. and C. Bucke, *Enzyme technology*1990, Cambridge England ; New York: Cambridge University Press. xvi, 264 p.
70. Shirakihara, Y. and P.R. Evans, *Crystal structure of the complex of phosphofructokinase from Escherichia coli with its reaction products.* Journal of molecular biology, 1988. **204**(4): p. 973-94.
71. Rosenberry, T.L., et al., *A steric blockade model for inhibition of acetylcholinesterase by peripheral site ligands and substrate.* Chemico-biological interactions, 1999. **119-120**: p. 85-97.

72. Colletier, J.P., et al., *Structural insights into substrate traffic and inhibition in acetylcholinesterase.* The EMBO journal, 2006. **25**(12): p. 2746-56.
73. Faucher, F., et al., *Crystal structures of human Delta4-3-ketosteroid 5beta-reductase (AKR1D1) reveal the presence of an alternative binding site responsible for substrate inhibition.* Biochemistry, 2008. **47**(51): p. 13537-46.
74. Chou, C.Y. and L. Tong, *Structural and biochemical studies on the regulation of biotin carboxylase by substrate inhibition and dimerization.* The Journal of biological chemistry, 2011. **286**(27): p. 24417-25.
75. Di Costanzo, L., et al., *Crystal structure of human liver Delta4-3-ketosteroid 5beta-reductase (AKR1D1) and implications for substrate binding and catalysis.* The Journal of biological chemistry, 2008. **283**(24): p. 16830-9.
76. Sekulic, N., M. Konrad, and A. Lavie, *Structural mechanism for substrate inhibition of the adenosine 5'-phosphosulfate kinase domain of human 3'-phosphoadenosine 5'-phosphosulfate synthetase 1 and its ramifications for enzyme regulation.* The Journal of biological chemistry, 2007. **282**(30): p. 22112-21.
77. Gamage, N.U., et al., *Structure of a human carcinogen-converting enzyme, SULT1A1. Structural and kinetic implications of substrate inhibition.* The Journal of biological chemistry, 2003. **278**(9): p. 7655-62.
78. Cabrera, R., et al., *Crystallographic structure of phosphofructokinase-2 from Escherichia coli in complex with two ATP molecules. Implications for substrate inhibition.* Journal of molecular biology, 2008. **383**(3): p. 588-602.
79. Fromm, S.A., et al., *The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex.* The EMBO journal, 2012. **31**(2): p. 279-90.
80. Kedersha, N., et al., *Stress granules and processing bodies are dynamically linked sites of mRNP remodeling.* The Journal of cell biology, 2005. **169**(6): p. 871-84.
81. Brangwynne, C.P., et al., *Germline P granules are liquid droplets that localize by controlled dissolution/condensation.* Science, 2009. **324**(5935): p. 1729-32.
82. Brangwynne, C.P., T.J. Mitchison, and A.A. Hyman, *Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(11): p. 4334-9.
83. Aizer, A., et al., *The P body protein Dcp1a is hyper-phosphorylated during mitosis.* PloS one, 2013. **8**(1): p. e49783.
84. Wang, C.Y., W.L. Chen, and S.W. Wang, *Pdc1 functions in the assembly of P bodies in Schizosaccharomyces pombe.* Molecular and cellular biology, 2013. **33**(6): p. 1244-53.
85. Fenger-Gron, M., et al., *Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping.* Molecular cell, 2005. **20**(6): p. 905-15.
86. Teixeira, D. and R. Parker, *Analysis of P-body assembly in Saccharomyces cerevisiae.* Molecular biology of the cell, 2007. **18**(6): p. 2274-87.
87. Jonas, S. and E. Izaurralde, *The role of disordered protein regions in the assembly of decapping complexes and RNP granules.* Genes & development, 2013. **27**(24): p. 2628-41.
88. Hyman, A.A. and K. Simons, *Cell biology. Beyond oil and water--phase transitions in cells.* Science, 2012. **337**(6098): p. 1047-9.

89. Kato, M., et al., *Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels.* Cell, 2012. **149**(4): p. 753-67.
90. Han, T.W., et al., *Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies.* Cell, 2012. **149**(4): p. 768-79.
91. Toro, I., et al., *Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile Archaeoglobus fulgidus.* Journal of molecular biology, 2002. **320**(1): p. 129-42.
92. Toro, I., et al., *RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex.* The EMBO journal, 2001. **20**(9): p. 2293-303.
93. Kambach, C., et al., *Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs.* Cell, 1999. **96**(3): p. 375-87.
94. Achsel, T., H. Stark, and R. Luhrmann, *The Sm domain is an ancient RNA-binding motif with oligo(U) specificity.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(7): p. 3685-9.
95. Sauter, C., J. Basquin, and D. Suck, *Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli.* Nucleic acids research, 2003. **31**(14): p. 4091-8.
96. Plessel, G., R. Luhrmann, and B. Kastner, *Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (E.F.G) protein heteromer and the intact snRNP core.* Journal of molecular biology, 1997. **265**(2): p. 87-94.
97. Marsh, J.A., et al., *Protein complexes are under evolutionary selection to assemble via ordered pathways.* Cell, 2013. **153**(2): p. 461-70.
98. Licht, K., et al., *3'-cyclic phosphorylation of U6 snRNA leads to recruitment of recycling factor p110 through LSm proteins.* RNA, 2008. **14**(8): p. 1532-8.
99. Weichenrieder, O., *RNA binding by Hfq and ring-forming (L)Sm proteins: A trade-off between optimal sequence readout and RNA backbone conformation.* RNA biology, 2014. **11**(5).
100. Fernandez, C.F., et al., *An Lsm2-Lsm7 complex in Saccharomyces cerevisiae associates with the small nucleolar RNA snR5.* Molecular biology of the cell, 2004. **15**(6): p. 2842-52.
101. Mura, C., et al., *Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function.* RNA biology, 2013. **10**(4): p. 636-51.
102. Neal, S., et al., *Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts.* Journal of biomolecular NMR, 2003. **26**(3): p. 215-40.
103. Pentikainen, U., O.T. Pentikainen, and A.J. Mulholland, *Cooperative symmetric to asymmetric conformational transition of the apo-form of scavenger decapping enzyme revealed by simulations.* Proteins, 2008. **70**(2): p. 498-508.
104. Kalek, M., et al., *Enzymatically stable 5' mRNA cap analogs: synthesis and binding studies with human DcpS decapping enzyme.* Bioorganic & medicinal chemistry, 2006. **14**(9): p. 3223-30.
105. Rydzik, A.M., et al., *Synthetic dinucleotide mRNA cap analogs with tetraphosphate 5',5' bridge containing methylenebis(phosphonate) modification.* Organic & biomolecular chemistry, 2009. **7**(22): p. 4763-76.

106. Rydzik, A.M., et al., *Synthesis and properties of mRNA cap analogs containing imidodiphosphate moiety--fairly mimicking natural cap structure, yet resistant to enzymatic hydrolysis.* Bioorganic & medicinal chemistry, 2012. **20**(5): p. 1699-710.
107. Wypijewska del Nogal, A., et al., *Analysis of decapping scavenger cap complex using modified cap analogs reveals molecular determinants for efficient cap binding.* The FEBS journal, 2013. **280**(24): p. 6508-27.
108. Draganescu, A., et al., *Fhit-nucleotide specificity probed with novel fluorescent and fluorogenic substrates.* The Journal of biological chemistry, 2000. **275**(7): p. 4555-60.
109. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.

6. Appendix

Peer-reviewed articles

- Mund M., Neu A., Ullmann J., Neu U., Sprangers R. (2011)
  Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings.
  J Mol Biol. 2011 Nov 25;414(2):165-76

- Fromm S.A., Kamenz J., Nöldeke E.R., Neu A., Zocher G. and Sprangers R. (2014)
  In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery.
  Angew Chem Int Ed Engl. 2014 Jul 7;53(28):7354-9

- Neu A., Neu U., Fuchs A.L., Schlager B. and Sprangers R. (2015)
  An excess of catalytically required motions inhibits the scavenger decapping enzyme
  Nat Chem Biol. 2015 Sep;11(9):697-704

doi:10.1016/j.jmb.2011.09.051 J. Mol. Biol. (2011) 414, 165–176

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMUNICATION

Structure of the LSm657 Complex: An Assembly Intermediate of the LSm1–7 and LSm2–8 Rings

Markus Mund[1]†, Ancilla Neu[1]†, Janina Ullmann[1], Ursula Neu[2] and Remco Sprangers[1]*

[1]*Max Planck Institute for Developmental Biology, Spemannstrasse 35, D-72076 Tuebingen, Germany*
[2]*Interfaculty Institute of Biochemistry, University of Tuebingen, Hoppe-Seyler-Str. 4, D-72076 Tuebingen, Germany*

Received 14 July 2011; received in revised form 23 September 2011; accepted 29 September 2011
Available online 6 October 2011

Edited by J. Doudna

Keywords:
mRNA degradation;
mRNA splicing;
NMR spectroscopy;
TROSY;
macromolecular complexes

The nuclear LSm2–8 (*l*ike *Sm*) complex and the cytoplasmic LSm1–7 complex play a central role in mRNA splicing and degradation, respectively. The LSm proteins are related to the spliceosomal Sm proteins that form a heteroheptameric ring around small nuclear RNA. The assembly process of the heptameric Sm complex is well established and involves several smaller Sm assembly intermediates. The assembly of the LSm complex, however, is less well studied. Here, we solved the 2.5 Å-resolution structure of the LSm assembly intermediate that contains LSm5, LSm6, and LSm7. The three monomers display the canonical Sm fold and arrange into a hexameric LSm657–657 ring. We show that the order of the LSm proteins within the ring is consistent with the order of the related SmE, SmF, and SmG proteins in the heptameric Sm ring. Nonetheless, differences in RNA binding pockets prevent the prediction of the nucleotide binding preferences of the LSm complexes. Using high-resolution NMR spectroscopy, we confirm that LSm5, LSm6, and LSm7 also assemble into a 60-kDa hexameric ring in solution. With a combination of pull-down and NMR experiments, we show that the LSm657 complex can incorporate LSm23 in order to assemble further towards native LSm rings. Interestingly, we find that the NMR spectra of the LSm57, LSm657–657, and LSm23–657 complexes differ significantly, suggesting that the angles between the LSm building blocks change depending on the ring size of the complex. In summary, our results identify LSm657 as a plastic and functional building block on the assembly route towards the LSm1–7 and LSm2–8 complexes.

Introduction

Sm proteins were initially identified as the target of autoantibodies from patients with systemic lupus erythematosus.[1] These antibodies recognize an epitope that is present in seven different Sm proteins common to the U1, U2, U4, and U5 small nuclear ribonucleoproteins (snRNPs) that perform the splicing reaction (reviewed by, e.g., Wahl *et al.*[2]). In snRNPs, the Sm proteins arrange into a seven-membered ring around the uridine-rich sequence in the small nuclear RNA (snRNA).[3–6] The Sm proteins

*Corresponding author. E-mail address: remco.sprangers@tuebingen.mpg.de.

† M.M. and A.N. contributed equally to this work.

Abbreviations used: snRNP, small nuclear ribonucleoprotein; snRNA, small nuclear RNA; NOESY, nuclear Overhauser enhancement spectroscopy; TROSY, transverse relaxation optimized spectroscopy; LSm, Like Sm.

0022-2836/$ - see front matter

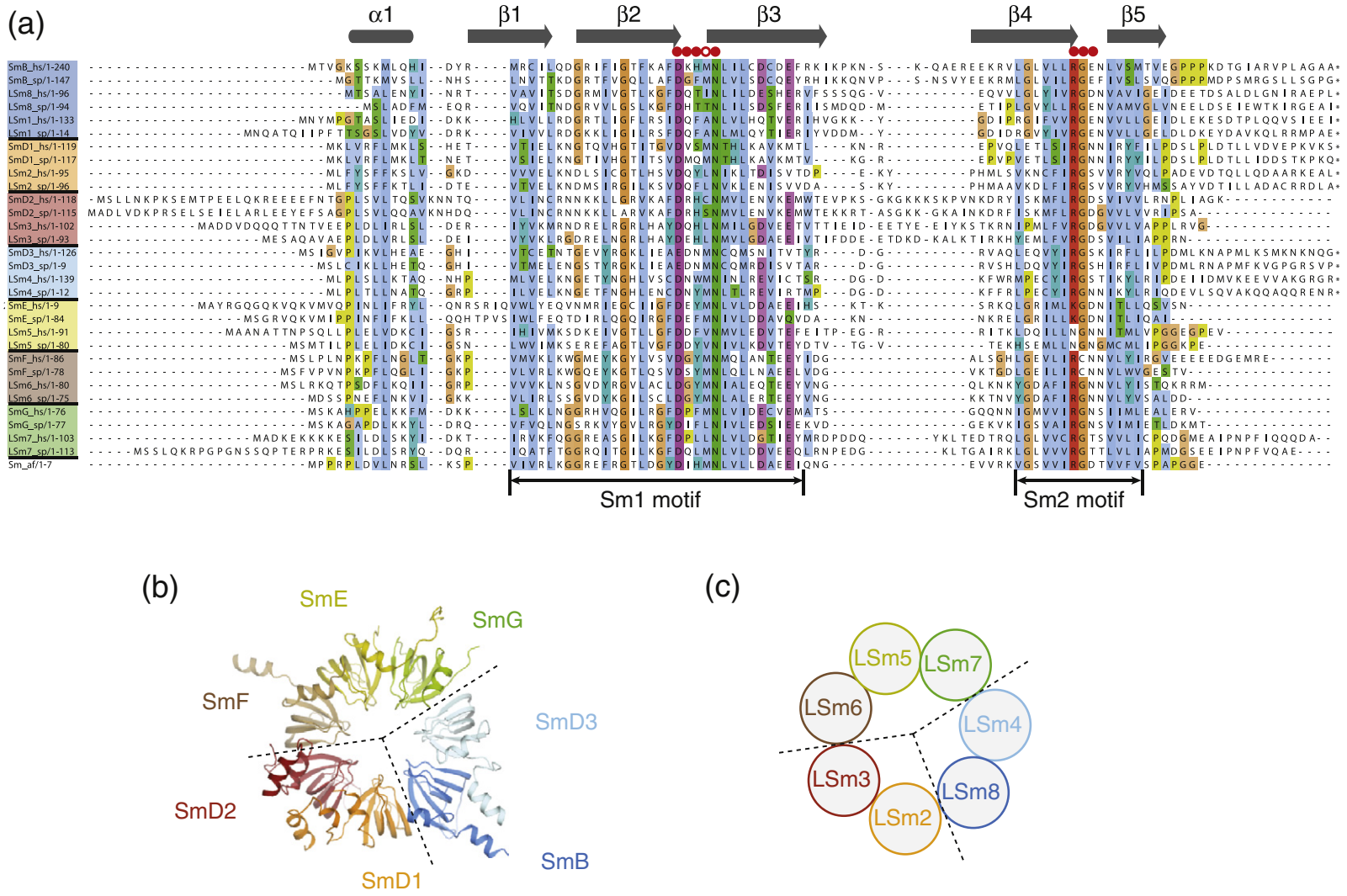


Fig. 1. LSm and Sm proteins. (a) Sequence alignment (calculated with T-Coffee[10]) of the human (hs), *S. pombe* (sp), and *A. fulgidus* (af) (L)Sm proteins. The residues are colored according to residue type; sequences ending with an asterisk are longer in the full-length protein and truncated in the figure for clarity. Related Sm and LSm proteins are grouped together.[11] Residues that are important for the interaction with RNA in the Sm and archaeal LSm complexes are indicated with red dots.[4,12,13] The open red dot indicates an interaction with a nucleotide that is mainly coordinated by a neighboring (L)Sm protein.[4] (b) Structure of the human Sm ring (PDB ID: 2Y9A[4]). The U4 RNA that passes through the hole in the middle of the Sm complex is removed for clarity. Broken lines indicate the Sm assembly intermediates. Figures displaying molecular structures were prepared with PyMOL (http://pymol.org/). (c) Predicted topology of the LSm2–8 (or LSm1–7) complex based on sequence conservation.

are characterized by the highly conserved Sm1 and Sm2 motifs that are 32 and 14 residues in length, respectively. Based on these signature sequences, additional proteins that contain an Sm fold have been identified and were termed LSm (*l*ike *Sm*) proteins[7–9] (Fig. 1a). Two different heptameric LSm complexes have been identified: a cytoplasmic LSm1–7 complex that is involved in mRNA degradation and a nuclear LSm2–8 complex that is involved in splicing.

The LSm2–8 complex contains a single copy of each of the seven LSm2–8 proteins and is based on electron microscopy data structurally similar to the Sm complex.[11,14,15] LSm2–8 is part of the U6 snRNP, where it specifically interacts with the 3′ polyU tract in the U6 snRNA.[7,9,14,16] The related LSm1 protein does not associate with LSm8 or any of the snRNPs[16] but is involved in the degradation of mRNA.[5] In that case, LSm1 associates specifically with LSm2–7 and multiple components of the mRNA degradation pathway.[17] The RNA recognition sequence for the LSm1–7 ring has not been established yet. It has, however, been proposed that the cytoplasmic LSm complex interacts with the oligoadenylated 3′ end of mRNA species.[18] Interestingly, overexpression of the LSm1 protein in yeast reduces the level of the LSm2–8 complex,[19] suggesting a competition between the assembly pathways of the two different LSm complexes.

Initial structural insights into the Sm and LSm protein complexes came from the high-resolution structures of two dimeric Sm complexes (SmD3B and SmD1D2).[20] The individual Sm proteins fold into a highly bent β-sheet preceded by a short α-helix. Sm proteins assemble into higher-molecular-mass complexes through the interaction of β-strand 4 of one monomer with β-strand 5 of the next monomer, resulting in the formation of an intermolecular antiparallel β-sheet.[20] This mode of interaction is reiterated multiple times to form hexameric,[21] heptameric (Fig. 1b),[4,12,21] or octameric[22] (L)Sm rings.

The heptameric Sm complex is unstable in the absence of RNA and requires the help of chaperones to form a stable ring. In that process, the SMN (*s*urvival of *m*otor *n*eurons) complex[23] recognizes three preformed Sm assembly intermediates (SmEFG, SmBD3, and SmD1D2) that are then positioned around the snRNA. In contrast, the LSm1–7 and LSm2–8 rings (Fig. 1c) are stable without RNA and are fully preformed before

Table 1. X-ray structure of the LSm657 complex

Data collection statistics	
Unit cell axes *a*, *b*, *c* (Å)	79.53, 63.10, 92.76
Unit cell angles α, β, γ (°)	90.00, 90.01, 90.00
Space group	$P2_1$
Wavelength (Å)	1.00
Resolution (Å) (highest-resolution bin)	46–2.5 (2.54–2.50)
Total reflections	146,718 (9867)
Unique reflections	60,656 (4062)
Redundancy	2.4 (2.4)
Completeness	96.9 (87.9)
$I/\sigma I$	18.9 (2.0)
R_{merge} (%)	4.0 (57.7)
Wilson *B*-factor (Å^2)	60.7
Refinement statistics	
R_{work} (%)	21.8 (35.5)
R_{free} (%)	26.9 (45.6)
RMSD bond length (Å)	0.006
RMSD bond angle (°)	1.03
Number of protein atoms	6750
Average *B*-factor (Å^2)	62.0
Number of solvent atoms	38
Average *B*-factor (Å^2)	62.8
Ramachandran plot	
Residues in the most favorable region (%)	86.4
Residues in the additionally allowed region (%)	13.6

Notes to Table 1

The LSm657 complex was purified using Ni-affinity chromatography, where the His-tag was present at the N-terminus of LSm7. The eluted protein complex was digested with tobacco etch virus protease to remove the affinity tag and was further purified by binding the complex to an ion-exchange matrix under low-salt conditions. Elution with a NaCl gradient separated the LSm657 complex from an LSm57 complex that was present due to the lower expression level of LSm6 compared to LSm5 and LSm7. Size-exclusion chromatography was performed as the final purification step in 50 mM sodium phosphate buffer (pH 7.5), 150 mM NaCl, and 1 mM DTT. Based on the size-exclusion data, the LSm657 complexes multimerize into hexamers, whereas the LSm57 complex does not assemble further.

For crystallization, the purified LSm657 complex was treated with trypsin to remove flexible regions. Based on the final structure, this resulted in the removal of the first 21 unstructured residues of LSm7. The trypsin was removed from the LSm complex using ion-exchange chromatography and size-exclusion chromatography, and the final protein concentration was adjusted to 10 mg/ml in 5 mM Tris (pH 7.5) and 150 mM NaCl. Crystals grew in 100 mM zinc acetate and 5% glycerol after 2 weeks upon sitting-drop vapor diffusion. X-ray diffraction data were collected at beamline X10SA of the Swiss Light Source (Villigen, Switzerland). Diffraction data were indexed, integrated, and scaled using XDS.[24] In the data set, there was a significant anomalous signal that extended to 3.1 Å due to bound Zn^{2+} from the crystallization solution. The structure was solved with the molecular replacement–single-wavelength anomalous dispersion protocol of Phaser in CCP4,[25] starting with a polyalanine search model derived from the homohexameric AF-SM2 ring, an archaeal LSm protein (PDB ID: 1LJO).[21] Two hexameric rings and 16 Zn^{2+} were located in the asymmetric unit. An electron density map at 2.5 Å resolution was generated from the molecular replacement–single-wavelength anomalous dispersion phases by density modification in Resolve[26] using phase extension, the 2-fold noncrystallographic symmetry between the two hexameric rings, and a prime-and-switch algorithm for eliminating model bias. The map revealed that the hexameric rings consist of dimers of LSm657 and was used for initial model building in Coot.[27] After rigid-body and simulated-annealing refinement with Phenix, the structure was rebuilt and refined iteratively with Refmac5 and Phenix.[28,29] The noncrystallographic symmetry relating the two copies of the ring was restricted to the secondary structure elements and was used as a constraint during refinement.

interacting with RNA.[15] Therefore, the RNA interaction and recognition properties of the heptameric LSm rings are thought to differ from those of the Sm ring. On the other hand, the high degree of similarity between the individual LSm and Sm proteins (Fig. 1a) suggests that the structure of the LSm ring is similar to the structure of the Sm ring (Fig. 1c). This also implies that there are assembly intermediates for the LSm ring similar to those that have been identified for the Sm proteins. Indeed, for the human LSm proteins, stable and soluble complexes of LSm23 (corresponding to SmD1D2), LSm48 (corresponding to SmD3B), and LSm657 (corresponding to SmEFG) have been purified.[15]

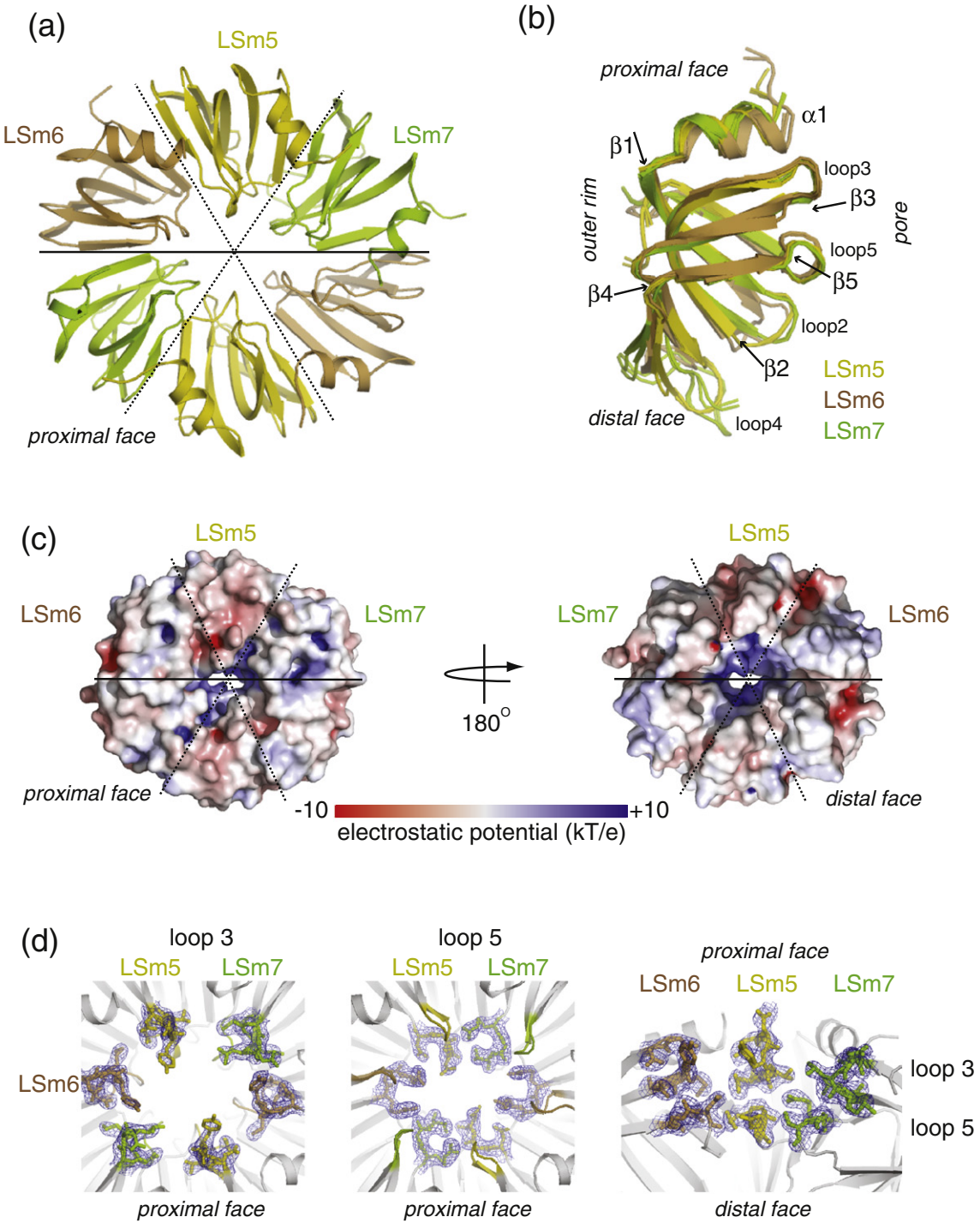


Fig. 2 (*legend on next page*)

Table 2. Structural differences between LSm and Sm proteins

	RMSD of Sm1 and Sm2 (C^{α} atoms)										
	LSm5	LSm6	LSm7	LSm3	SmE (LSm5)	SmF (LSm6)	SmG (LSm7)	SmB (LSm8 LSm1)	SmD1 (LSm2)	SmD2 (LSm3)	SmD3 (LSm4)
LSm5	***0.20***	**1.13**	**0.75**	0.99	0.64	0.78	0.83	0.82	0.81	0.81	0.78
LSm6		***0.58***	**1.34**	1.45	1.11	1.45	1.29	1.36	1.23	1.38	1.32
LSm7			***0.13***	0.97	0.69	0.76	0.86	0.61	0.91	0.71	0.86
LSm3				*0.06*	0.92	0.84	0.99	0.85	0.91	0.87	1.00
SmE					*0.33*	0.73	0.69	0.79	0.77	0.68	0.85
SmF						*0.33*	0.81	0.71	0.73	0.77	0.84
SmG							*0.33*	0.87	0.72	0.70	0.81
SmB								*0.32*	0.88	0.75	0.79
SmD1									*0.37*	0.84	0.67
SmD2										*0.32*	0.79
SmD3											*0.26*

Numbers indicate the RMSD (Å) of the C^{α} atoms in the Sm1 and Sm2 motifs. Coordinates were extracted from the 4 structures in the asymmetric unit of the LSm657 complex (LSm5, LSm6, and LSm7), from the 2 structures in the asymmetric unit of the *Saccharomyces cerevisiae* LSm3 octameric ring (LSm3, sc; PDB ID: 3BW1),[22] and from the 12 structures in the asymmetric unit of the human Sm complex (SmE–SmD3; PDB IDs: 2Y9A–2Y9D).[4] The diagonal italicized numbers indicate differences within the ensembles of the same protein; off-diagonal numbers indicate differences between the average structures of the different proteins. RMSDs and average structures were calculated using MOLMOL.[32] Numbers in boldface indicate the LSm structures solved in this study. Underlined entries indicate the Sm protein with the lowest RMSD to the LSm proteins. Note that the structural differences are small and do not show the relationship between the Sm and LSm proteins present at the sequence level (Fig. 1).

The LSm657 complex forms a hexameric ring

Here, we present the 2.5 Å crystal structure of the *Schizosaccharomyces pombe* LSm complex containing the LSm5, LSm6, and LSm7 proteins (Table 1). The asymmetric unit of the crystal contains two hexameric LSm657–657 rings (Fig. 2a). In the complex, LSm5 is in-between LSm7 and LSm6, in agreement with the order of the Sm proteins in the Sm ring, where SmE is located in-between SmF and SmG (Figs. 1b and 2a). This finding confirms the predicted topology of the native LSm2–8 and LSm1–7 rings that was derived from sequence conservation (Fig. 1c) and underlines the conservation between the Sm complex and the LSm complex. This conservation is further reflected by the fact that electron microscopy studies have shown that the SmEFG complex (the equivalent of the LSm657 complex) assembles into hexameric rings *in vivo*.[31]

The individual LSm proteins in the LSm657 complex multimerize into a hexameric ring in a canonical manner. β-Strand 4 of one monomer interacts with β-strand 5 of a neighboring monomer, thereby forming one continuous β-sheet through the ring. The LSm–LSm interfaces are stabilized by the N-terminal α-helix and through additional intermolecular contacts, including hydrophobic interactions and salt bridges, as previously observed for the Sm proteins.[4,12,20]

The LSm5, LSm6, and LSm7 proteins each adopt a canonical Sm fold that contains a short N-terminal α-helix, followed by five β-strands that assemble into a highly bent anti-parallel β-sheet (Fig. 2b). The Sm1 motif forms the core of the Sm fold and contains most of β-strands 1–3. The Sm2 motif folds around the Sm1 motif and mostly consists of β-strands 4 and 5, which engage in intermolecular contacts. Differences between the four copies of each LSm protein in the asymmetric unit are very small, and their Sm1 and Sm2 motifs superimpose with a low root-mean-square deviation (RMSD; 0.20 Å, 0.58 Å, and 0.13 Å, respectively; see Table 2, bold diagonal entries). Small displacements of the residues in loop 2 of

Fig. 2. Crystal structure of the LSm657 complex. (a) Crystal structure of the LSm657 ring viewed from the proximal side. The LSm proteins form a continuous β-sheet that includes intermolecular interactions between β-strand 4 of one monomer and β-strand 5 of a neighboring subunit. The interface between LSm6 and LSm7 (see the text below) is indicated by a continuous line. The other interfaces are indicated with broken lines. (b) Superposition of the four copies of LSm5 (yellow), LSm6 (brown), and LSm7 (green) present in the asymmetric unit of the crystal structure. All LSm proteins show a typical Sm fold: an N-terminal α-helix followed by a highly bent five-stranded β-sheet. The RMSD between multiple copies of the LSm proteins is low, as are the differences between LSm5, LSm6, and LSm7 (see Table 2, bold entries). Loops and secondary structure elements are indicated. (c) Electrostatic surface potential of the LSm657 ring. The orientation of the structure on the left is identical with the orientation shown in (a); the right panel displays the distal face of the ring. The potential was calculated with APBS.[30] Note the positive charge in the central pore of the ring. The approximate borders of the individual LSm proteins are indicated with lines. Note that the surface potential is not exactly symmetric due to small differences between two copies of the same LSm protein in the ring. (d) View into the putative RNA binding pockets of the LSm5, LSm6, and LSm7 proteins in the LSm657 complex. The $2F_o - F_c$ electron density map is shown at 1σ. Left and middle: View as in (a). For clarity, residues in loop 3 (left) and loop 5 (middle) are displayed separately. Right: View as in (b), displaying the inside of the pore of the ring.

LSm6 result in a larger RMSD between the four LSm6 copies. Differences between the LSm5, LSm6, and LSm7 proteins are also small, as reflected in the RMSDs well below 1.5 Å for the differences between the average LSm5, LSm6, and LSm7 structures (Table 2, bold off-diagonal entries). Similar RMSDs between the different Sm proteins within the heptameric Sm complex had been observed previously (Table 2).[4]

Differences in (L)Sm sequences are not reflected in the domain fold

Based on the sequence conservation between the individual Sm and LSm proteins (Fig. 1), we expected the LSm5, LSm6, and LSm7 proteins to possess the highest structural similarity with the SmE, SmF, and SmG proteins, respectively. The RMSDs between related Sm and LSm proteins were, however, not significantly lower than the RMSDs between more distant Sm and LSm proteins (Table 2 and Fig. 1). RMSDs between human Sm proteins, archaeal LSm proteins, and bacterial Hfq proteins were also observed to be below 1.5 Å.[33] These observations imply that the relationship between the Sm protein and the LSm protein that is present at the sequence level is not reflected in the three-dimensional structure of the core of the proteins. This suggests that the differences in protein sequence determine the specific assembly into high-molecular-mass complexes without disturbing the fold of the isolated domains. In that case, the amino acids that are unique to each of the seven classes of the Sm and LSm proteins (Fig. 1a) provide the complementary surfaces that determine the order of the proteins in the native rings (Fig. 1b and c). Comparison of the available interfaces within the LSm complex solved here, however, did not allow us to identify what residues determine this high specificity. A structure of the native LSm complexes might provide the additional data necessary to identify how the order of the LSm proteins in the native complex is established. In that regard, it is interesting to note that intermolecular hydrogen bonds and salt bridges present in the dimeric SmD1D2 and SmBD3 complexes[20] are not identical with those found between these subunits in the fully assembled Sm ring.[4] It is thus tempting to speculate that the Sm and LSm assembly intermediates are held together by a slightly different network of interactions than the proteins in the fully assembled complexes are.

Implications of the LSm657 structure for RNA binding

The central region of the LSm657 complex is positively charged, suggesting an important role for the pore in the interaction with RNA substrates (Fig. 2c). Indeed, the structure of the homoheptameric *Archaeoglobus fulgidus* Sm ring in complex with polyU RNA identified a nucleotide (uridine) binding pocket in this region.[13] In this structure, each of the Sm proteins interacts with a single base through conserved residues in loops 3 and 5 (Fig. 1a, red dots). The structures of the heptameric Sm ring (Fig. 1b) in complex with RNA[3,4,12,34] show that the RNA substrate is passed through the central Sm pore, where all Sm proteins interact with the consecutive bases in a slightly different manner.[4,12] The residues in the Sm complex that contact the RNA are largely conserved in the LSm proteins (Fig. 1a). This suggests that the LSm1–7 and LSm2–8 complexes use the same nucleotide binding pockets as the Sm complex. Nevertheless, several key amino acids in loops 3 and 5 are different in the structure of the LSm657 complex. Firstly, the LSm5 protein contains a conserved asparagine in loop 5 instead of a canonical arginine or lysine that hydrogen bonds to the base in the Sm ring. Secondly, the LSm7 protein contains a conserved leucine residue in loop 3 instead of a canonical aromatic residue that provides a stacking interaction with the base in Sm–RNA complexes. These differences prevent the prediction of the base preference for the LSm nucleotide binding pockets based on the known Sm–RNA structures. Extrapolation of RNA binding preferences from the Sm complexes to the LSm complexes is further complicated by the fact that the Sm proteins assemble around the snRNA to form a stable snRNP, whereas the LSm1–7 and LSm2–8 complexes are thought to interact more transiently with the free 3′ end of RNA.[35] Future studies that address the RNA recognition properties of the LSm complexes will be able to address these differences at a structural level.

Both LSm657 rings in the asymmetric unit of the crystal interact with eight Zn^{2+} in a very similar fashion (Supplementary Fig. 1). Four Zn^{2+} are bound on the surface of the hexameric ring and are localized through interaction with one or two protein side chains. In addition, one Zn^{2+} is tightly bound at every LSm5–LSm6 interface, and a cluster of three Zn^{2+} localizes to one of the two LSm6–LSm7 interfaces in each ring. Zinc binding to the LSm657 complex is most likely a direct consequence of the very high concentration of zinc acetate (100 mM) from the crystallization solution (Table 1).

NMR analysis of the LSm657 complex confirms the crystal structure

We have shown by X-ray crystallography that the LSm5, LSm6, and LSm7 proteins arrange into a hexameric LSm657–657 ring structure. To independently confirm this quaternary arrangement in solution, we decided to use high-resolution NMR spectroscopy. These studies were motivated by several considerations. Firstly, for the formation of

the native LSm1–7 or LSm2–8 complexes, two of the six intermolecular interfaces in the LSm657–657 complex need to dissociate to allow for the incorporation of the four additional LSm proteins. The circular structure could thus be disadvantageous for the assembly process of the native LSm complexes. Secondly, we wanted to assess if the presence of Zn^{2+} at the interfaces between the LSm proteins in the crystal structure (Fig. 2d) interfered with the quaternary structure of the complex. Our solution studies will address the multimerization state of the LSm657 complex in the absence of zinc. Finally, differences between the crystal structure and the solution state were previously observed for the LSm3 protein. In that case, the crystal structure suggested the assembly of the LSm3 rings into complexes of very high molecular mass, a feature that was not observed in solution.[22]

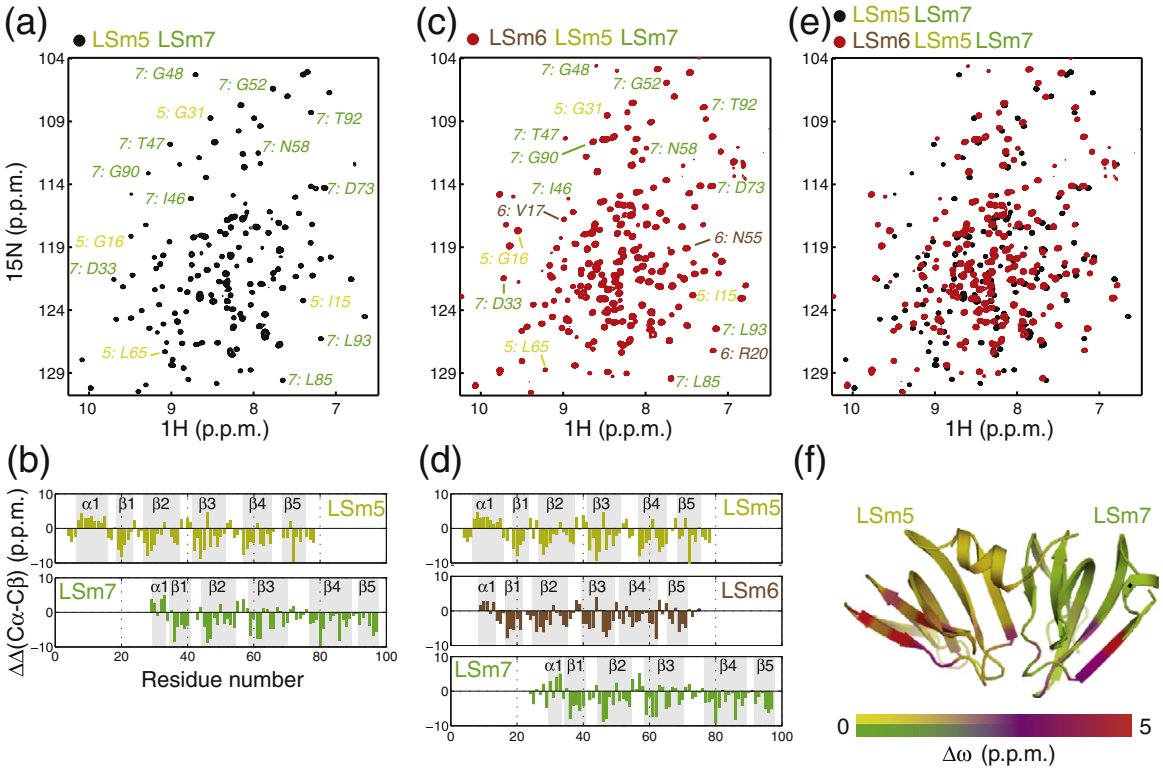


Fig. 3. NMR analysis of the LSm57 and LSm657 complexes. (a) Two-dimensional ^{1}H–^{15}N spectrum of LSm57; exemplary assignments are indicated. U-[^{13}C, ^{15}N, ^{2}H]-labeled NMR samples were prepared by overexpression of the LSm complex in a D_2O-based minimal medium containing $^{2}H^{13}C$-labeled glucose and ^{15}N-labeled ammonium chloride as the sole carbon and nitrogen sources, respectively. The overexpressed proteins were purified as described in the legend to Table 1. The NMR samples contained the full-length proteins and were not subjected to digestion with trypsin. The final protein concentration in the samples was between 0.8 mM and 1.5 mM in size-exclusion buffer supplemented with 5% D_2O and 0.03% sodium azide. No zinc acetate was added to the NMR samples. NMR experiments were recorded at 303 K/308 K (LSm57) or 308 K (LSm657) on Bruker AVIII-600 and AVIII-800 NMR spectrometers equipped with a room temperature probehead. The backbone resonances of the LSm57 complex were assigned using TROSY[36] versions of standard triple-resonance experiments (HNCACB/HNCOCACB). (b) Secondary chemical shifts of LSm57 (calculated as $\Delta\Delta C^{\alpha} - \Delta\Delta C^{\beta}$). Positive values correspond to helical regions, and negative values correspond to extended conformations. Note that the positive secondary chemical shift in the middle of some of the β-strands corresponds to the location where the β-strands are most highly bent. In general, the solution data (bars) correspond very well to the crystal structure (gray areas that represent the secondary structure elements). Residues 100–113 in LSm7 that are invisible in the NMR spectra are not displayed in the plot. (c) Two-dimensional ^{1}H–^{15}N spectrum of the 60-kDa LSm657 complex. Backbone resonances were assigned using TROSY[36] versions of HNCA/HNCOCA, HNCB/HNCOCB, HNCACO/HNCO, and ^{15}N-edited NOESY spectra (Supplementary Fig. 3). In addition, a diagonal free N–N-edited TROSY–NOESY–TROSY experiment was recorded to resolve spectral overlap[37] (Supplementary Fig. 3). (d) Secondary chemical shifts of the LSm657 complex (see (b)) agree very well with the secondary structure elements in the crystal structure. (e) Superposition of the LSm57 (a) and LSm657 (c) spectra indicating the large number of residues that experience chemical shift changes upon the interaction of LSm57 with LSm6. (f) Chemical shift differences between LSm57 dimer and LSm57 in the LSm657 hexamer. The largest differences (e.g., G90 in LSm7; a and c) are located at the outer β-strands of LSm5 and LSm7. These strands engage in interactions with LSm6 to form the hexameric ring observed in the crystal structure. Most other residues change their chemical shifts, likely as a result of a different intermolecular angle between LSm5 and LSm7 in the different complexes (see text).

From a technical point of view, it is not straightforward to study the hexameric LSm657 ring with a molecular mass of 60 kDa using high-resolution NMR spectroscopy. We thus decided to first dissect the complex into smaller building blocks that are easier to tackle spectroscopically. To that end, we made use of the fact that the 20-kDa dimeric LSm57 complex does not assemble into a higher oligomeric state (Fig. 1). This implies that β-strand 4 of LSm5 is unable to form a stable intermolecular β-sheet with β-strand 5 of LSm7 (see also the text above).Our ^{1}H–^{15}N NMR spectra of LSm57 are of high quality, and we could obtain a complete backbone assignment using a U-[^{13}C,^{15}N,^{2}H]- labeled sample (Fig. 3a). Two lines of evidence confirm that the dimeric LSm57 complex is folded similarly in solution as within the hexameric LSm657 complex in the crystal. Firstly, the secondary chemical shifts derived from the NMR data agree well with the secondary structure elements of the LSm57 complex (Fig. 3b). Secondly, almost all short amide–amide distances derived from the crystal structure are maintained in solution, as observed in ^{15}N-edited nuclear Overhauser enhancement spectroscopy (NOESY) spectra (Supplementary Fig. 2). The dimerization of LSm57 in solution is also similar to that observed in the crystal structure, as we observed the corresponding intermolecular contacts in our solution studies.

We next studied the 60-kDa hexameric LSm657 complex in solution using full protein deuteration and transverse relaxation optimized spectroscopy (TROSY) techniques (Fig. 3c). Comparison of the ^{1}H–^{15}N spectra of LSm57 and LSm657 shows that a transfer of assignments is not possible due to the large number of chemical shift changes (Fig. 3e and the text below). We therefore performed a complete resonance assignment of the LSm657 complex (Fig. 3c). As for the LSm57 dimer, the secondary chemical shifts of the LSm657 complex are in good agreement with the crystal structure of the LSm657 complex (Fig. 3d). In addition, ^{15}N-edited NOESY spectra recorded on the LSm657 sample confirmed short-range amide–amide distances derived from the crystal structure, including intermolecular cross-strand contacts (Supplementary Fig. 3). In cases of amide resonance overlap, we made use of the recently developed diagonal free TROSY ^{15}N (t_1)–NOESY–TROSY^{15}N (t_2) ^{1}H(t3) experiment,[37] which also unambiguously confirmed the presence of intermolecular contacts in solution. In summary, we conclude that the LSm657 complex forms a hexameric ring in solution that is structurally very similar to the crystal structure that we solved. This also confirms that the Zn^{2+} in the crystal does not induce an artificial formation of intermolecular interfaces.

In addition to the Sm fold, LSm7 contains a 25-residue-long N-terminal extension and a 15-residue-long C-terminal extension. In solution, these residues are likely disordered, as judged from the absence of corresponding resonances in the NMR spectra. This results from the elevated pH (7.5) and temperature (35 °C) during the measurements that lead to a rapid loss of the NMR signal from unprotected amide protons due to a fast exchange with bulk water. Further evidence for the unstructured nature of these residues comes from the observation that the N-terminal residues of LSm7 were readily cleaved during the trypsin treatment of the protein complex used for crystallization and from the lack of electron density for the C-terminal LSm7 residues. Interestingly, regions in the Sm proteins that lie outside the Sm fold also seem disordered in the assembly intermediates SmBD1 and SmD1D2.[20] However, in the context of the RNA-bound Sm complex, these regions adopt a folded state that then directly interacts with the RNA.[4]

The angle between LSm proteins is adaptable in different LSm complexes

Upon incorporation of the dimeric LSm57 into the hexameric LSm657 complex, two new intermolecular interfaces are formed: the free β-strand 4 in LSm5 engages in interactions with β-strand 5 of LSm6, and the free β-strand 5 in LSm7 engages in interactions with β-strand 4 of LSm6. As such, we expected chemical shift differences between the NMR spectra of the LSm57 complex and the NMR spectra of the LSm657 complex that were limited to these residues (Fig. 3e). Indeed, resonances corresponding to these residues change their chemical shift significantly (Fig. 3f). However, we also observed smaller but significant chemical shift changes for many other resonances remote from the newly formed interface (Fig. 3e and f). We interpret these additional shifts as resulting from the adjustment of the angles between the individual LSm proteins. In the free LSm57 dimer, the LSm–LSm angle is predicted to be close to 51°, which is optimal for the formation of a heptameric ring.[4,20] In the crystal structure of the hexameric LSm657 ring, we observe, however, that the angle is restrained to 60°. Interestingly, a similar change in the angle between (L)Sm proteins has been observed for the *A. fulgidus* Sm2 protein. This archaeal Sm complex can arrange into a hexameric ring in the free state;[21] however, it forms a heptameric ring in the presence of RNA,[38] which shows the plasticity of the (L)Sm proteins that allows for assembly into differently sized complexes.

LSm657 is a functional assembly intermediate

At least two intermolecular interfaces within the hexameric LSm657–657 complex need to open to allow for the formation of the heptameric LSm1–7 or LSm2–8 complex (see the text above). Previous

attempts to reconstitute the LSm657 particles into LSm1–7 or LSm2–8 rings were performed in the presence of high levels of urea, a condition that is very remote from that found in a cellular environment.[15] To test if additional LSm proteins can be incorporated into the LSm657 complex under physiological conditions, we performed pull-down experiments with purified His-tagged LSm657 and untagged LSm23 complexes (Fig. 4a, left). While untagged LSm23 alone (Fig. 4a, lane 1) showed no affinity for the Ni resin (Fig. 4a, lane 2), His-tagged LSm657 (Fig. 4a, lane 3) interacted with the affinity resin (Fig. 4a, lane 4). When we applied an incubated mixture of untagged LSm23 and His-tagged LSm657 (Fig. 4, lane 5) to the affinity resin, LSm23 was copurified with LSm657 (Fig. 4, lane 6). This clearly shows that LSm23 interacts strongly with LSm657 and that a stable complex containing five different

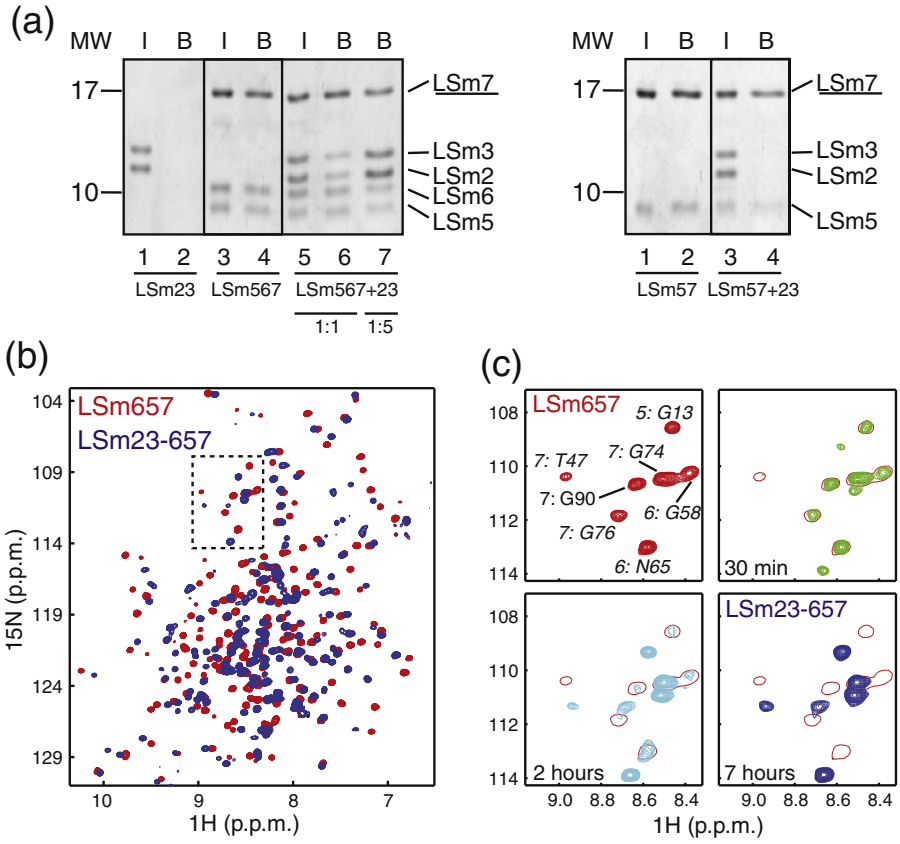


Fig. 4. LSm657 can further assemble into LSm23–657. (a, left) Pull-down experiments of LSm657 and LSm23. Untagged LSm23 (lane 1) does not interact with the affinity resin (lane 2). His-tagged LSm657 (lane 3) does bind to the affinity resin (lane 4). Mixing untagged LSm23 with His-tagged LSm657 in a 1:1 ratio (lane 5), followed by incubation at 37 °C for 1 h, results in the selective purification of LSm23 with LSm657 due to the (partial) formation of an LSm23–657 complex (lane 6). Interestingly, incubation of LSm657 with a 5-fold molar excess of LSm23 resulted in the formation of an LSm23–657 complex that was about 2-fold enriched in LSm23. This suggests that the LSm23–657 complex contains two copies of LSm23 and is thus heptameric. The purification of LSm23 was analogous to the purification of LSm657. LSm7 that binds to the affinity matrix is underlined. Proteins were analyzed by SDS-PAGE and visualized with Coomassie blue. Molecular mass is indicated on the left. Lanes marked "I" refer to the input, and lanes marked "B" refer to the complexes eluted from the affinity resin with imidazole. (a, right) Pull-down experiments of LSm57 and LSm23. His-tagged LSm57 (lane 1) interacts with the affinity resin (lane 2). Mixing untagged LSm23 with His-tagged LSm57 in a 1:1 ratio (lane 3), followed by incubation at 37 °C for 1 h, did not result in the formation of an LSm23–57 complex (lane 4). This indicates that LSm57 is not an efficient assembly intermediate of the LSm1–7 or LSm2–8 complexes. (b) ^{1}H–^{15}N TROSY spectrum of ^{2}H–^{15}N-labeled LSm657 before (red) and 7 h after (blue) the addition of an excess of unlabeled LSm23. Chemical shift changes are due to the formation of the LSm23–657 complex. (c) The formation of the LSm23–657 complex is slow, corresponding to a process with a high-energy barrier, as expected of the reorganization of the LSm ring. Shown are the spectra of LSm657 (red contours) recorded before the addition of LSm23 (top left) and 30 min, 2 h, and 7 h after LSm23 was added. The LSm657 resonances are shown as single contours in the three latter spectra. Thirty minutes after the addition of LSm23 to LSm657 (green contours), the LSm23–657 resonances appear. Some of the LSm657 resonances are still visible after 2 h (cyan contours). After 7 h (blue contours), LSm657 has been completely converted into LSm23–657. The region displayed corresponds to the boxed region in (b).

LSm proteins is formed. Interestingly, the addition of an excess of LSm23 compared to LSm657 resulted in the formation of an LSm23–657 complex that was enriched in LSm23 (Fig. 4, lane 7). This suggests that the LSm657 complex can incorporate two copies of LSm23 to form a heptameric LSm$(23)_2$–657 complex. This is in agreement with known assembly states of (L)Sm proteins that have been reported to be either hexameric, heptameric, or octameric. It is interesting to note that the LSm23–657 complex corresponds to the SmD1D2–EFG complex,[39] which is the smallest Sm assembly intermediate for the Sm complex that has been shown to interact with RNA.

In a control experiment, we show that LSm57 (Fig. 4a, right, lane 4) is not able to efficiently form a stable complex with LSm23. This indicates that LSm57, as opposed to LSm657, is not a functional building block of the native LSm ring; this correlates well with the prediction that LSm3 interacts with LSm6 in the native LSm ring (Fig. 1c).

To observe the assembly reaction of LSm657 and LSm23 at atomic resolution, we added NMR-inactive LSm23 to U-[^{15}N,^{2}H]-labeled LSm657 (Fig. 4b). As in the pull-down experiments, we clearly observe that LSm23 interacts with LSm657, since we found significant chemical shift changes in most of the resonances of LSm657. Interestingly, the changes in the NMR spectrum did not occur instantaneously but were induced over a time period of several hours (Fig. 4c). This indicates that the interaction between LSm23 and LSm657 is not a simple binding event between the two complexes. The slow change in the spectra rather reports on an event with a high energetic barrier, as is expected of the remodeling of the LSm657 complex into an LSm23–657 complex. Although the process we observe here in the NMR tube is relatively slow, it is reasonable to believe that the assembly is significantly faster in a cellular environment due to the presence of assisting protein complexes and/or LSm14, LSm48, or RNA.

The differences between the NMR spectra of LSm657 and the NMR spectra of LSm23–657 are reminiscent of the changes that we observed in the NMR spectra of LSm57 upon assembly into the hexameric LSm657 rings (Figs. 3e and 4b). There, the spectra of LSm57 changed due to the direct interaction with LSm6 and due to changes in the intermolecular orientation of the LSm5 and LSm7 proteins (see the text above). The differences between the NMR spectra of LSm657 and the NMR spectra of the LSm23–657 complex could be explained similarly as resulting directly from the interaction with LSm23 and indirectly from a change in the intermolecular angle between LSm5, LSm6, and LSm7, which are no longer assembled in a hexameric ring but in a heptameric LSm$(23)_2$–657 ring.

In order to identify residues that experience the largest chemical shift changes and are thus at the newly formed LSm657–LSm23 interface, we would require chemical shift assignments of the LSm23–657 complex. Unfortunately, this information cannot be obtained based on the assignments of the LSm657 complex (Fig. 3c), as complex formation is slow on the chemical shift timescale (see the text above; Fig. 4c), preventing the transfer of assignments from LSm657 to LSm23–567. However, based on the order of the LSm proteins in the native LSm complexes (Fig. 1c), it is tempting to speculate that the LSm6–LSm7 interface dissociates to accommodate LSm23. This notion is supported by the analysis of our LSm657 structure using the PISA (Protein Interfaces, Surfaces, and Assemblies) server (European Bioinformatics Institute‡), which indicates that the LSm67 interface contains fewer hydrogen bonds and salt bridges than the LSm56 and LSm57 interfaces.

In summary, our data provide the first insights into the structure of heteromeric LSm complexes. We solved the high-resolution crystal structure of the LSm657 complex and show that it forms a hexameric ring, where the order of the proteins is in agreement with the order of the Sm proteins in the Sm complex. Using high-resolution NMR spectroscopy, we show that the structure and assembly state of the complex in solution are very similar to the crystal structure that we solved. In addition, we used pull-down experiments and NMR spectroscopy to show that the hexameric LSm657 ring can act as an assembly intermediate towards the native LSm complexes by incorporating LSm23. Interestingly, it was shown previously that SmEFG, which corresponds to LSm657, is a functional assembly building block of the spliceosomal Sm complex. Our data thus indicate that not only the structure but also the assembly process of the LSm and the Sm rings are conserved. Finally, our results form the basis of further studies addressing the assembly, structure, and RNA recognition properties of the cytoplasmic LSm1–7 complex and the nuclear LSm2–8 complex.

Accession numbers

The coordinates for the LSm657 complex have been deposited in the Protein Data Bank (PDB) under accession code 3SWN. The chemical shifts of the LSm57 and LSm657 complexes have been deposited in the Biological Magnetic Resonance Bank under accession codes 17781 and 17782, respectively.

Acknowledgements

We thank Silke Wiesner and Oliver Weichenrieder for discussion and Markus Hartmann for help

‡ http://www.ebi.ac.uk/pdbe/prot_int/pistart.html

with the recording of the diffraction data. We thank Wolfgang Bermel (Bruker Biospin GmbH, Rheinstetten, Germany) and Vincent Truffault for help with the implementation of the TROSY–NOESY–TROSY experiment. This study was supported by the Max Planck Society and by a Marie Curie Reintegration Grant (FP7/2007-2013, grant agreement no. 239164) to R.S.

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2011.09.051

References

1. Lerner, M. R. & Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA*, **76**, 5495–5499.
2. Wahl, M. C., Will, C. L. & Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
3. Stark, H., Dube, P., Luhrmann, R. & Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539–542.
4. Leung, A. K., Nagai, K. & Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature*, **473**, 536–539.
5. Boeck, R., Lapeyre, B., Brown, C. E. & Sachs, A. B. (1998). Capped mRNA degradation intermediates accumulate in the yeast spb8-2 mutant. *Mol. Cell. Biol.* **18**, 5062–5072.
6. Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B. & Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J.* **1**, 1259–1265.
7. Seraphin, B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14**, 2089–2098.
8. Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H. & Luhrmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.* **14**, 2076–2088.
9. Cooper, M., Johnston, L. H. & Beggs, J. D. (1995). Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.* **14**, 2066–2075.
10. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
11. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.* **18**, 3451–3462.
12. Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. & Wahl, M. C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184.
13. Toro, I., Thore, S., Mayer, C., Basquin, J., Seraphin, B. & Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20**, 2293–2303.
14. Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. & Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3′-end of U6 snRNA, thereby facilitating U4/U6 duplex formation *in vitro*. *EMBO J.* **18**, 5789–5802.
15. Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K. & Kambach, C. (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *J. Biol. Chem.* **280**, 16066–16075.
16. Mayes, A. E., Verdone, L., Legrain, P. & Beggs, J. D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J.* **18**, 4321–4331.
17. Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D. & Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. *Nature*, **404**, 515–518.
18. Chowdhury, A., Mukhopadhyay, J. & Tharun, S. (2007). The decapping activator LSm1p–7p–Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA*, **13**, 998–1016.
19. Luhtala, N. & Parker, R. (2009). LSM1 over-expression in *Saccharomyces cerevisiae* depletes U6 snRNA levels. *Nucleic Acids Res.* **37**, 5529–5536.
20. Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A. *et al.* (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.
21. Toro, I., Basquin, J., Teo-Dreher, H. & Suck, D. (2002). Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. *J. Mol. Biol.* **320**, 129–142.
22. Naidoo, N., Harrop, S. J., Sobti, M., Haynes, P. A., Szymczyna, B. R., Williamson, J. R. *et al.* (2008). Crystal structure of LSm3 octamer from *Saccharomyces cerevisiae*: implications for LSm ring organisation and recruitment. *J. Mol. Biol.* **377**, 1357–1371.
23. Chari, A., Golas, M. M., Klingenhager, M., Neuenkirchen, N., Sander, B., Englbrecht, C. *et al.* (2008). An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell*, **135**, 497–509.
24. Kabsch, W. (2010). XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 125–132.
25. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
26. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 965–972.
27. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.

28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
29. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221.
30. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
31. Plessel, G., Luhrmann, R. & Kastner, B. (1997). Electron microscopy of assembly intermediates of the snRNP core: morphological similarities between the RNA-free (E.F.G.) protein heteromer and the intact snRNP core. *J. Mol. Biol.* **265**, 87–94.
32. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–5, 29–32.
33. Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.* **31**, 4091–4098.
34. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J. & Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, **458**, 475–480.
35. Chan, S. P., Kao, D. I., Tsai, W. Y. & Cheng, S. C. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, **302**, 279–282.
36. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. (1997). Attenuated T_2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA*, **94**, 12366–12371.
37. Diercks, T., Truffault, V., Coles, M. & Millet, O. (2010). Diagonal-free 3D/4D HN,HN-TROSY–NOESY–TROSY. *J. Am. Chem. Soc.* **132**, 2138–2139.
38. Achsel, T., Stark, H. & Luhrmann, R. (2001). The Sm domain is an ancient RNA-binding motif with oligo (U) specificity. *Proc. Natl Acad. Sci. USA*, **98**, 3685–3689.
39. Raker, V. A., Plessel, G. & Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.* **15**, 2256–2269.

Supplementary Materials

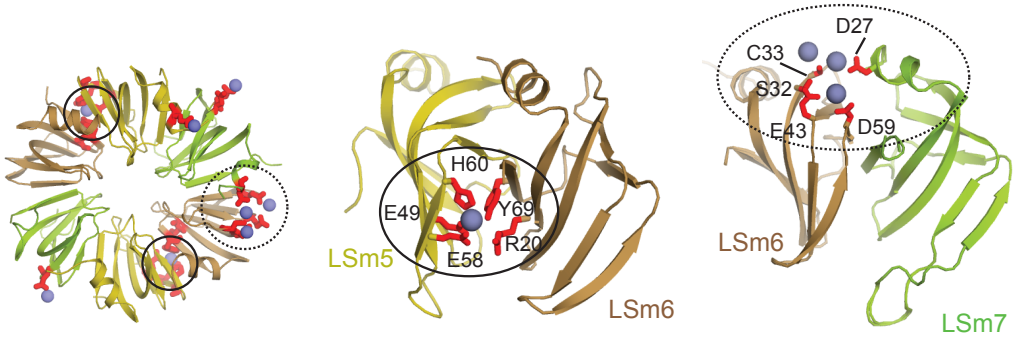
Structure of the LSm657 complex: an assembly intermediate of the LSm1-7 and LSm2-8 rings.

Markus Mund[1], Ancilla Neu[1], Janina Ullmann[1], Ursula Neu[2] and Remco Sprangers[1]

[1] Max Planck Institute for Developmental Biology, Spemannstrasse 35, D-72076 Tuebingen, Germany

[2] Interfaculty Institute of Biochemistry, University of Tuebingen, Hoppe-Seyler-Str. 4, D-72076 Tuebingen, Germany

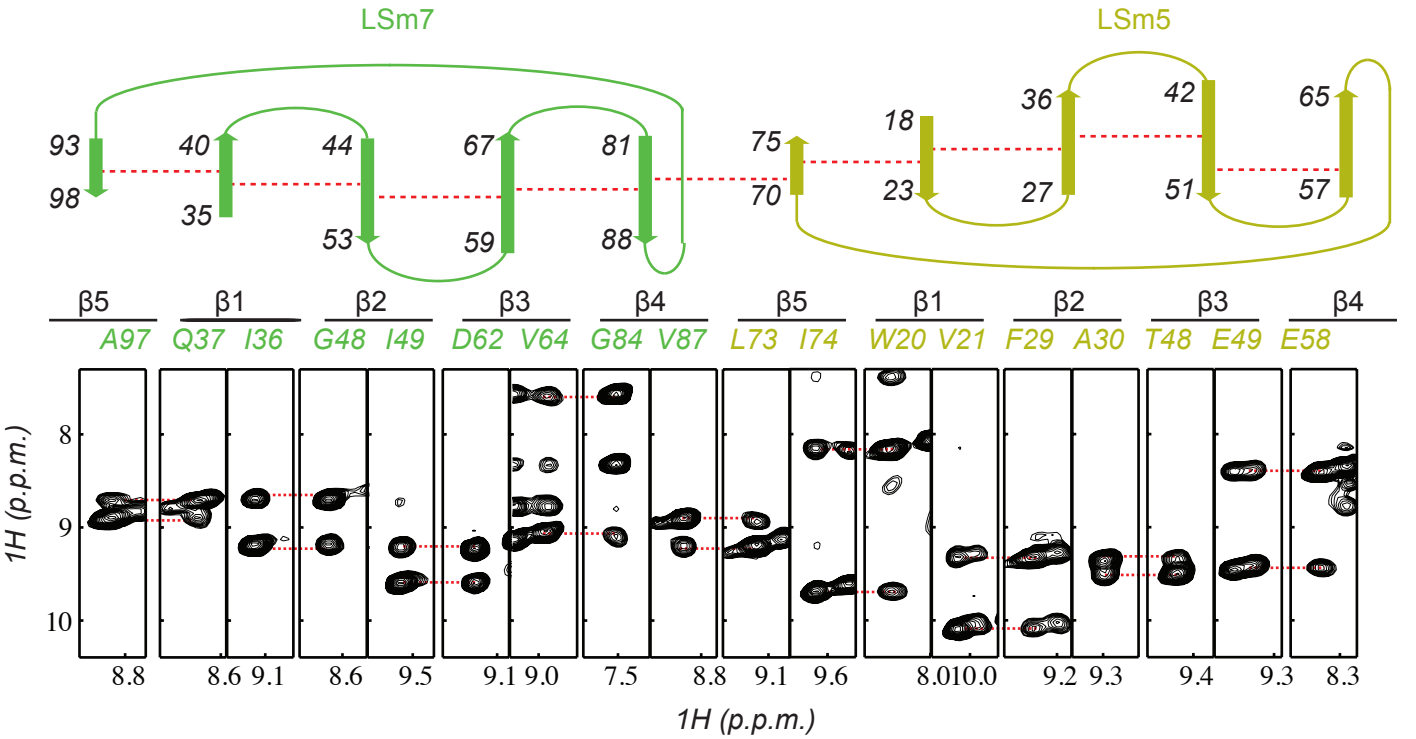
Supplementary Figure 1



Supplementary Figure 1: The LSm657 complex coordinates zinc in the crystal structure

View of the LSm657 ring in the same orientation as in Figure 2A. The location of the Zn^{2+} ions in the crystal structure (left) is indicated. One Zn^{2+} ion is tightly bound between LSm5 and LSm7 (middle). A cluster of Zn^{2+} ions is located at the LSm6:LSm7 interface (right). The occupancy of some zinc sites is lower than 1 and therefore not all of the LSm6:LSm7 interfaces necessarily contain 3 Zn^{2+} ions. Side chains that coordinate the Zn^{2+} ions are labeled and shown as red sticks.

Supplementary Figure 2

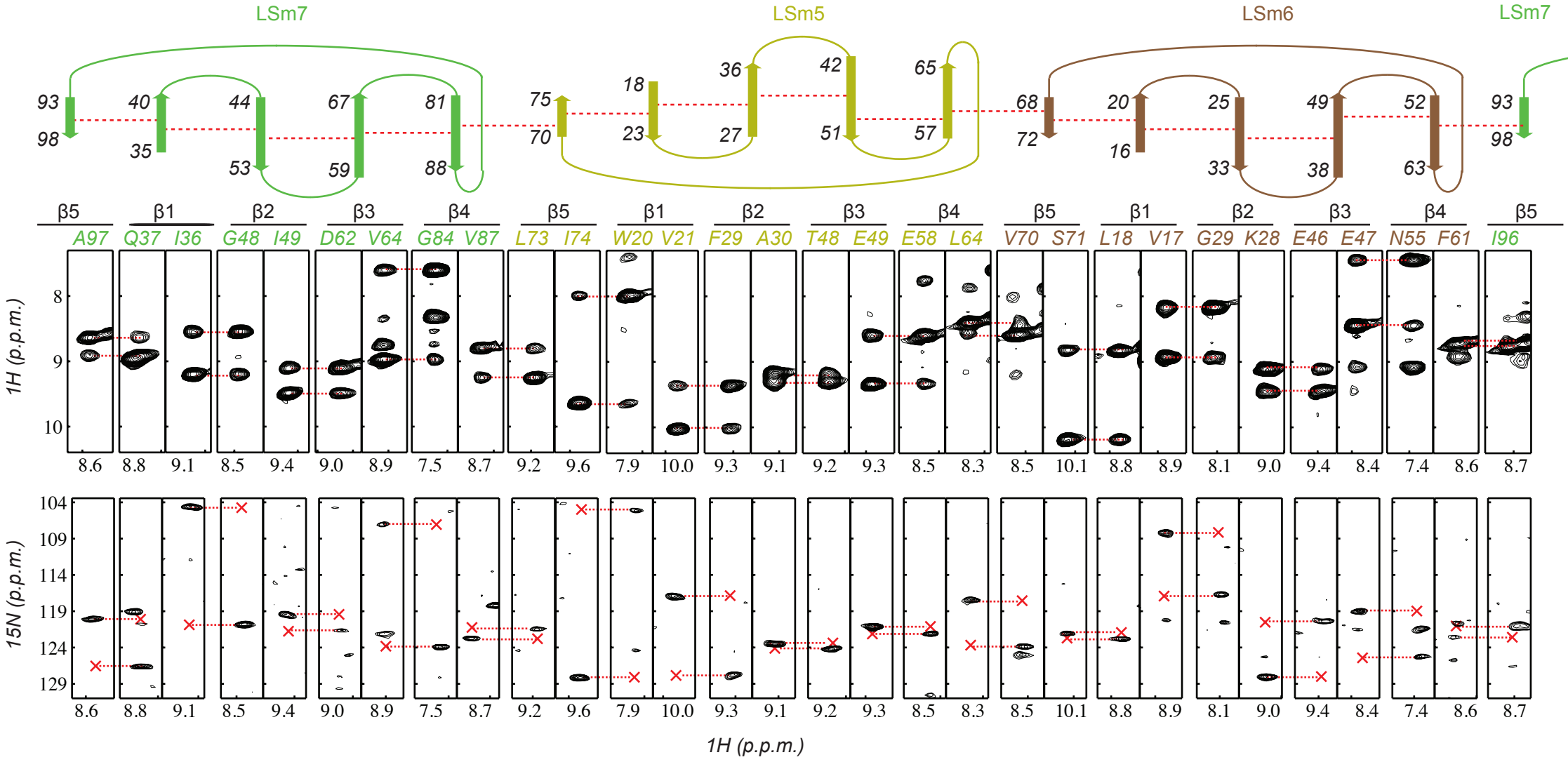


Supplementary Figure 2: The LSm57 structure in solution

Top: Schematic representation of the topology of the LSm57 complex, as derived from the crystal structure of the LSm657 complex. The N-terminal α-helix is not shown for clarity. Residue numbers at the beginning and end of the strands are indicated. The red dotted lines correspond to short amide-amide distances in the crystal structure.

Bottom: Strips derived from a $^{1}H(t_1)$-NOE-TROSY $^{15}N(t_2)$ ^{1}H (t_3) experiment that display cross strand NOEs (intra- or intermolecular) that confirm the presence of the contacts indicated with the red dotted lines (top) in solution.

Supplementary Figure 3



Supplementary Figure 3: The LSm657 structure in solution

Top: Schematic representation of the topology of the LSm657 complex, as derived from the crystal structure. The red dotted lines correspond to short amide-amide distances in the structure. (see also legend of Supplementary Figure 2)

Middle: Strips derived from an $^{1}H(t_1)$-NOE-TROSY $^{15}N(t_2)$ ^{1}H (t_3) experiment that display cross strand NOEs (intra- or intermolecular) that confirm the presence of the contacts indicated with the red dotted lines (top) in solution.

Bottom: Strips derived from a diagonal free TROSY ^{15}N (t1)-NOESY- TROSY ^{15}N (t2) ^{1}H(t3) experiment [1], displaying the "circular walk" through the LSm657 hexameric ring. The crosses at the end/ beginning of the red dotted lines indicate the location of the diagonal.

References

1. Diercks, T., Truffault, V., Coles, M. & Millet, O. (2010). Diagonal-free 3D/4D HN,HN-TROSY-NOESY-TROSY. *Journal of the American Chemical Society* **132**, 2138-9.

Protein–Protein Interactions Hot Paper

DOI: 10.1002/anie.201402885

In Vitro Reconstitution of a Cellular Phase-Transition Process that Involves the mRNA Decapping Machinery**

*Simon A. Fromm, Julia Kamenz, Erik R. Nöldeke, Ancilla Neu, Georg Zocher, and Remco Sprangers**

Abstract: *In eukaryotic cells, components of the 5′ to 3′ mRNA degradation machinery can undergo a rapid phase transition. The resulting cytoplasmic foci are referred to as processing bodies (P-bodies). The molecular details of the self-aggregation process are, however, largely undetermined. Herein, we use a bottom-up approach that combines NMR spectroscopy, isothermal titration calorimetry, X-ray crystallography, and fluorescence microscopy to probe if mRNA degradation factors can undergo phase transitions in vitro. We show that the Schizosaccharomyces pombe Dcp2 mRNA decapping enzyme, its prime activator Dcp1, and the scaffolding proteins Edc3 and Pdc1 are sufficient to reconstitute a phase-separation process. Intermolecular interactions between the Edc3 LSm domain and at least 10 helical leucine-rich motifs in Dcp2 and Pdc1 build the core of the interaction network. We show that blocking of these interactions interferes with the clustering behavior, both in vitro and in vivo.*

Processing bodies (P-bodies) are cytoplasmic micrometer-scale ribonucleoprotein (RNP) foci that were initially identified in mammalian cells[1] and that have now been observed in distant branches of eukaryotes. The list of proteins that are found in these mRNP granules has grown significantly[2] and comprises general mRNA decay factors including the Dcp2 decapping enzyme and the enhancers of decapping Dcp1, Edc3, Pat1, and LSm1–7.[1c,3] The proposed roles of the clustering of the decapping enzyme within a confined cellular space includes enhanced substrate binding,[4] buffering of the concentration of translating mRNAs,[5] regulation of the concentrations of free cytoplasmic proteins,[6] and a means to respond to cellular stress.[7]

The details of the interactions that underlie the self-assembly process of proteins and RNA into P-bodies are largely undetermined. Genetic studies have addressed the importance of individual proteins[8] and revealed two major characteristics. 1) the details of the intermolecular interactions vary among species. For example, Edc3 and the Q/N rich C-terminal region of LSm4 can act as a scaffold for P-body assembly in *S. cerevisiae*,[3c,9] whereas the Pdc1 protein has been shown to be important in *S. pombe*.[10] 2) the P-body assembly process is highly redundant and deletion of a single protein does not abolish the aggregation process completely.[8] This redundancy complicates genetic approaches that aim to identify the interactions that regulate the self-assembly process.

Recently, it was shown that the purified Nck and N-WASP proteins can undergo a phase transition in vitro.[11] These experiments are in line with the idea that the formation of cellular granules relies on a network of multivalent weak interactions between different components.[6,12] Herein, we ask whether it is possible to reconstitute a phase-transition process with proteins that have been shown to localize to P-bodies. To that end, we focus on the purified *S. pombe* Dcp2 mRNA decapping enzyme, its prime activator Dcp1 and the scaffolding proteins Edc3 and Pdc1 and show that these proteins can spontaneously self-assemble into oil-like droplets.

The proteins Edc3 and Dcp2 (Figure 1 A) have been shown to interact through contacts between the Edc3 LSm domain and short helical leucine-rich motifs (HLMs) in Dcp2.[13] Herein, we use NMR spectroscopy titration experiments and show that Dcp2 contains at least seven different HLMs that are recognized specifically by the Edc3 LSm domain (Figure 1 B). Interestingly, the extent of the observed chemical shift changes varies, indicating that the HLM affinities vary considerably. To quantify these, we used isothermal titration calorimetric (ITC) assays (Figure 1 C,D) and find that the Edc3:HLM affinities range from (2.5 ± 0.3) μM to the mM range.

A prerequisite for protein phase separation is the capability to associate into indefinite soluble assemblies. In general, this condition is fulfilled for multivalent proteins with

[*] S. A. Fromm, E. R. Nöldeke, A. Neu, Dr. R. Sprangers
Max Planck Institute for Developmental Biology
Spemannstrasse 35, 72076 Tübingen (Germany)
E-mail: remco.sprangers@tuebingen.mpg.de
J. Kamenz
Friedrich Miescher Laboratory of the Max Planck Society
Tübingen (Germany)
G. Zocher
Interfaculty Institute of Biochemistry, University of Tübingen
Tübingen (Germany)

[**] We acknowledge Janina Ullmann for excellent technical support, Dr. Michael Rosen (University of Texas Southwestern) for insights at the initial phase of the project, Andrew Renault for providing access to the fluorescent microscope and Silke Wiesner and Silke Hauf for valuable discussions. J.K. thanks the Boehringer Ingelheim Fonds for a Ph.D. stipend. This work was supported by the Max Planck Society. This project has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013), ERC grant agreement no. 616052.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201402885.

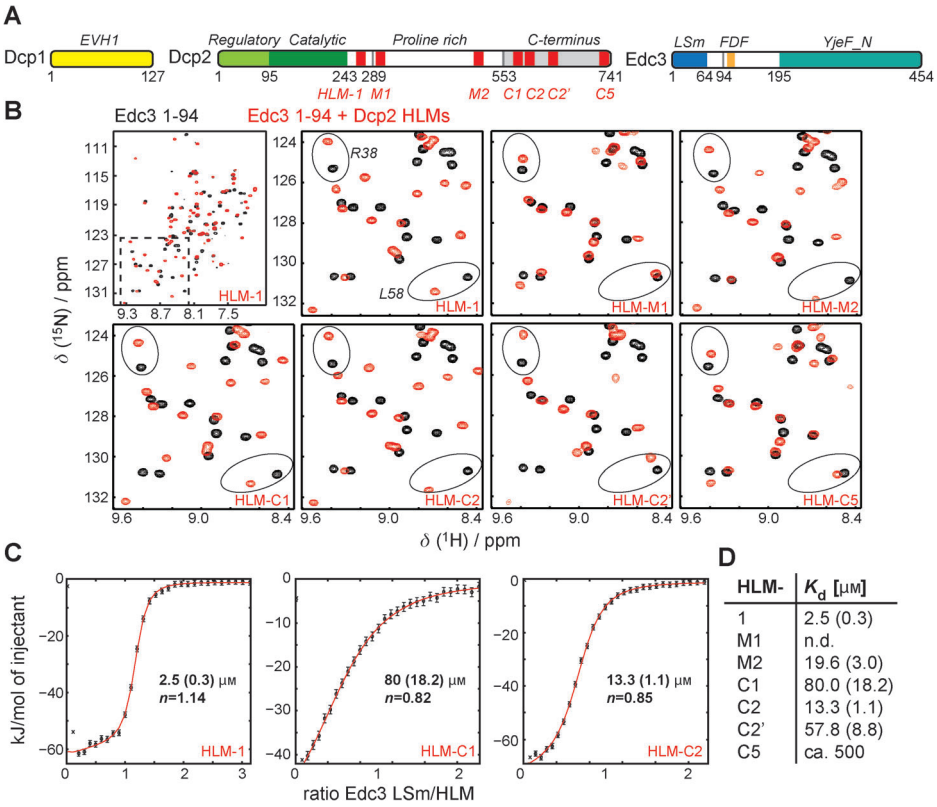


Figure 1. The seven HLMs in Dcp2 bind to the Edc3 LSm domain with a wide range of affinities. A) Schematic representation of *S. pombe* Dcp1 (EVH1 domain, yellow), Dcp2 (regulatory light green, catalytic dark green, HLMs red, proline-rich mid domain white, C-terminus gray) and Edc3 (LSm dark blue, FDF motif orange, YjeF N light blue). All used protein constructs are listed in Supporting Information Table S1 A. B) ^{1}H–^{15}N correlation spectra of the binding between the monomeric Edc3 LSm domain and the seven isolated Dcp2 HLMs. Black: free ^{15}N labeled Edc3 LSm domain, Red: the Edc3 LSm domain in the presence of a fivefold excess of the individual Dcp2 HLMs. The boxed region in the top left panel is shown in all other spectra. Residues R38 and L58 are highlighted with ovals to indicate that the HLMs induce chemical shift perturbations to a different extent. C) ITC graphs for the binding of the Edc3 LSm domain to the Dcp2 HLM-1, HLM-C1 and HLM-C2 sequences. The best fit is drawn with a red line and the extracted K_d values including the error (standard deviation) are indicated. Deviations from $n = 1.0$, where n refers to the stoichiometry, result from small inaccuracies in the protein-concentration determination. D) Overview of the determined K_d values (error in parenthesis) for the Dcp2 HLM:Edc3 interactions (n.d.: not determinable with ITC).

independent binding sites. Theoretically, Edc3 and Dcp2 can form such an indefinite and highly branched network of interactions because Edc3 is a dimer (and thus contains two LSm domains) and Dcp2 contains seven Edc3 docking sites (Figure S1 A in the Supporting Information). In addition, the binding affinities in the Edc3:Dcp2 system range from low μM to mM which has been suggested to be beneficial for the generation of phase transitions.[6] To address if the Edc3:Dcp2 network is suitable to induce a phase separation in vitro, we mixed the recombinantly expressed and purified far C-terminal region of Dcp2 that contains four HLMs (Dcp2 residues 553–741; Table S1 A) with full-length Edc3. Interestingly, a clear phase separation can be observed using a bright-field microscope (Figure 2 A, left). Both Edc3 and Dcp2 are significantly enriched within the droplet-like structures as judged from the fluorescent signal that resulted from the Oregon green labeled proteins (Figure 2 A, and Figures S1 B, S1 C). Our data thus shows that the network of interactions between Edc3 and Dcp2 is sufficient to induce phase separations.

To obtain insights into the conditions that result in phase separation we varied the concentrations of Edc3 and the part of Dcp2 that contains 4 HLMs (Figure 2 B, left). We observed that phase separation occurs when a number of conditions are met. First, the absolute concentration of the individual proteins needs to be sufficiently high. In case a certain threshold concentration is not exceeded, phase separation does not take place and the proteins remain homogeneously distributed in solution. Second, the molar ratio of the two proteins needs to be within specific limits. In case the excess of either of the two proteins is too large, phase separation is abolished. This occurrence can be explained by the fact that a large excess of Edc3 would result in a situation where all HLMs are bound to a dimeric Edc3 protein, which would result in a loss of intertwining between different Dcp2 chains. A large excess of Dcp2 would, on the other hand, result in a situation where there are insufficient dimeric Edc3 proteins available to link the Dcp2 chains.

In a second set of experiments, we determined if the number of intermolecular interactions between components influences the phase-separation process. As full length Dcp2 that comprises seven HLMs was not expressed stably in our hands, we designed a version of Dcp2 that contains the complete unfolded C-terminal region but that lacks a part of the proline-rich region. This Dcp2 C-term ΔMid protein (Dcp2 residues 242–289 + 553–741; Supporting Information Table S1 A) contains five HLMs and undergoes phase transitions at concentrations that are significantly lower than those required for the Dcp2 construct that contains four HLMs. As an example: in the presence of 150 μM Edc3 phase separation takes place when the Dcp2 C-term ΔMid (five HLMs) concentration is 25 μM (125 μM modular HLM concentration), whereas double the concentration of the Dcp2-fragment that contains four HLMs (50 μM; 200 μM modular HLM concentration) is required. In agreement with previous observations,[11] we can thus conclude that the valency and affinities of the interacting partners determines whether phase separation occurs at specific concentrations. Importantly, as controls we performed experiments using either

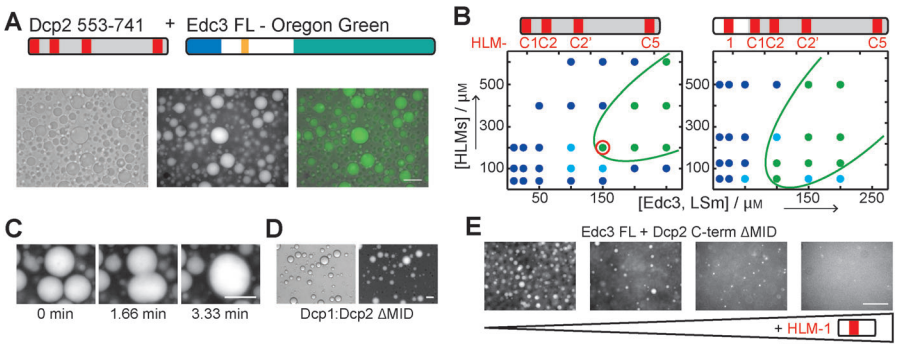


Figure 2. In vitro phase transitions of purified Edc3 and Dcp2. A) Phase transition of 50 µM Dcp2 553-741 and 150 µM Edc3 (doped 1:100 with Edc3-OregonGreen). Left: bright field (BF) channel; middle: Oregon Green (OG) channel; right: merge. B) Phase diagrams of phase transition of Edc3 together with Dcp2 553–741 (that contains 4 HLM sequences; left) or Dcp2 242–741ΔMID (that contains 5 HLM sequences). Given concentrations are modular concentrations, for example, 50 µM of Dcp2 553–741 is 200 µM HLMs, because 4 HLMs are in the Dcp2 553–741 construct. Modular LSm concentration is identical with total Edc3 concentration as an Edc3 monomer has one LSm domain. Occurrence of phase transition at a given condition is color coded; blue = no phase transition; light blue = beginning phase transition; green = clear phase transition. Green lines indicate progression of the phase boundaries. The red encircled condition is shown in (A) and (C). C) The in vitro droplets are highly dynamic and fuse over time. The time-scale is indicated below the OG channel pictures. D) Droplets formed by 25 µM Dcp1:Dcp2ΔMID and 100 µM Edc3 (doped 1:100 with Edc3-OG). Left: BF channel, Right: OG channel. E) Droplets formed by 25 µM Dcp2 242–741ΔMID and 100 µM Edc3 disappear by addition of increasing amounts of Dcp2 242–291 (containing HLM-1) (5 µM, 10 µM, 25 µM, 50 µM). Scale bars: 50 µm.

a monomeric version of Edc3 or a Dcp2 sequence that only contains a single HLM (Figure S1 D,E). In none of these control experiments we observed phase separation (Figure S1 D,E), as the valency of one of the components is reduced to one.

Cellular foci, including P-bodies, have been shown to be highly dynamic.[14] To determine if the in vitro Edc3:Dcp2 phase separations that we prepared from purified components display a similar behavior we monitored the droplets over a longer time. We observed a number of fusion events that confirm the liquid-like behavior of the in vitro droplets (Figure 2 C). This underscores that the in vitro system is similar to the in vivo situation regarding this aspect.

In vivo, the foci that contain Dcp2 and Edc3 also comprise numerous additional proteins. To test if it is possible to recruit additional proteins within the reconstituted in vitro droplets, we extended the Dcp2 C-term ΔMid construct such that it also includes the Dcp2 catalytic and the regulatory domain (Dcp2 ΔMid, residues 1–289 + 553–741; Table S1 A) that has been shown to interact with Dcp1.[15] We then mixed purified Dcp1:Dcp2 complexes that harbored five HLM sequences with Edc3 and observed a clear phase separation, indicating that the Edc3:Dcp2 droplets can recruit additional factors (Figure 2 D). It should be noted that Dcp1 contains the binding site for Xrn1,[16] thereby providing a mechanism to recruit the exonuclease to the Dcp1:Dcp2:Edc3 assemblies. In addition, the Edc3 protein contains an FDF repeat, providing a means to recruit the helicase DDX6/Dhh1.[17] The catalytic domain of Dcp2[18] and the helicase can then provide binding sites for mRNA substrates.

The Edc3:HLM interactions form the core of the in vitro droplets we assemble. Interference with this interaction is thus expected to result in the loss of phase separation. To test this hypothesis, we supplemented the Dcp2:Edc3 droplet conditions with the HLM-1 sequence (Figure 1 D). With increasing amounts of the HLM-1 Dcp2 peptide, we observe a disappearance of the phase separation in vitro (Figure 2 E). To assess the consequences of the specific interference with the Edc3:HLM interaction for P-body formation in vivo, we overexpressed the HLM-1 peptide in *S. pombe*. To observe the recruitment of Edc3 to P-bodies we replaced the *edc3+* gene at its endogenous locus by the *edc3+-mCherry* fusion construct. To simultaneously assess P-body integrity we additionally fused Dcp2 or Lsm7 to GFP at their endogenous loci. In cells that did not express the peptide that interferes with the Edc3:HLM interaction we observed P-bodies that contain Edc3-mCherry and Dcp2-GFP or Lsm7-GFP, respectively (Figure 3 A,B, top rows). On the other hand, in cells that overexpress the HLM peptide the Edc3-mCherry protein is no longer recruited into P-bodies (Figure 3 A,B, bottom

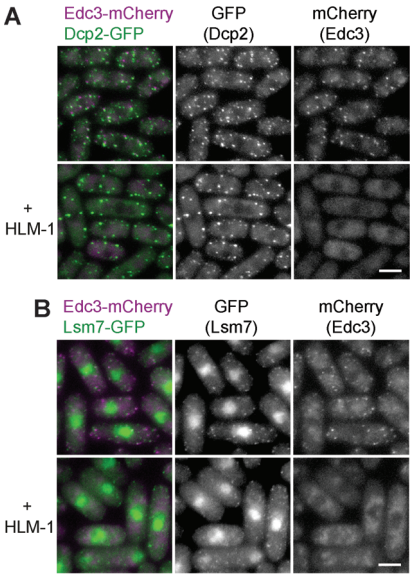


Figure 3. Effect of HLM-1 expression on Edc3 localization in vivo. A,B) Fluorescent micrographs of *S. pombe* expressing mCherry and GFP tagged versions of Edc3 and Dcp2 (A) or Lsm7 (B), respectively. Without HLM-1 overexpression Edc3 and Dcp2 are enriched into P-bodies (top row). Upon overexpression (*adh1* promoter) of HLM-1 Edc3 no longer localizes to P-bodies but is diffusely spread in the cell (bottom row). Dcp2-GFP (A) and Lsm7 (B) still localize to P-bodies upon overexpression of HLM-1, indicating that interference with the Edc3:HLM interactions does not generally disturb P-body formation. Scale bar: 5 µm.

rows). This lack of recruitment is most likely a consequence of the Edc3 LSm domains that are now saturated with the overexpressed mono-valent peptide. These in vivo observations are in agreement with our in vitro phase separation. Importantly, the overexpression of the HLM competition peptide does not result in the general loss of P-bodies as both Dcp2-GFP and Lsm7-GFP are still found in distinct cellular foci (Figure 3A,B, middle panels, bottom rows). These observations stress the fact that in vivo P-body formation is a highly redundant process, where the loss of one P-body component does not result in a general loss of these foci.

To obtain insights into the redundancy of P-body formation, we turned to the Pdc1 protein that was recently implicated in P-body integrity in *S. pombe*.[10] Pdc1 is related to metazoan Edc4 (Ge-1, Hedls)[19] and both proteins contain an N-terminal WD-40 domain and a central coiled-coil region (Figure 4A). Close inspection of the protein sequence at the N-terminal region of the *S. pombe* Pdc1 protein indicates the presence of several HLMs (Figure 4A). Using NMR titration experiments (Figure 4B) we confirmed that the N-terminal region of Pdc1 contains at least three HLM sequences that can interact with the Edc3 LSm domain. The extent of the induced chemical shift perturbations in the Edc3 LSm domain varied between the Pdc1 HLMs, as was the case for the seven Dcp2 HLMs (Figure 1B, 4B). Using ITC experiments, we quantified the associated affinities and found that the three Pdc1 HLM sequences interact with the Edc3 LSm domain between 150 μM and lower mM range (Figure 4C,D). Interestingly, the Pdc1 protein contains a central coiled-coil region, through which the protein can oligomerize, resulting in an increased number of HLMs in the biological unit of the protein. Above, we have shown that a large number of HLMs is favorable for the phase separation process (Figure 2B). To test whether Edc3 and Pdc1 are able to induce phase transitions in vitro we initially tried to purify the Pdc1 protein from *E. coli*. Unfortunately, full-length Pdc1 was, in our hands, not stable enough to allow for in vitro phase separation experiments. Consequently, we engineered a Pdc1 protein where the coiled-coil region was replaced with glutathione S-transferase (GST). GST is a dimer in solution and thus mimics the oligomerization effect of the Pdc1 coiled-coil region. We then used this designed dimeric Pdc1 protein (that has an HLM valency of six) and the dimeric Edc3 protein (that has an LSm valency of two) to test whether Pdc1 and Edc3 can undergo phase transitions in vitro. Interestingly, we observe that Pdc1 and Edc3 are able to engage in an indefinite network of intermolecular interaction that results in the formation of an oil-like droplet phase (Figure 4E), as we observed for Dcp2 and Edc3 (Figure 2A). We can thus conclude from our in vitro experiments that Pdc1 and Dcp2 are redundant proteins with regard to the potential of inducing phase separation when mixed together with Edc3.

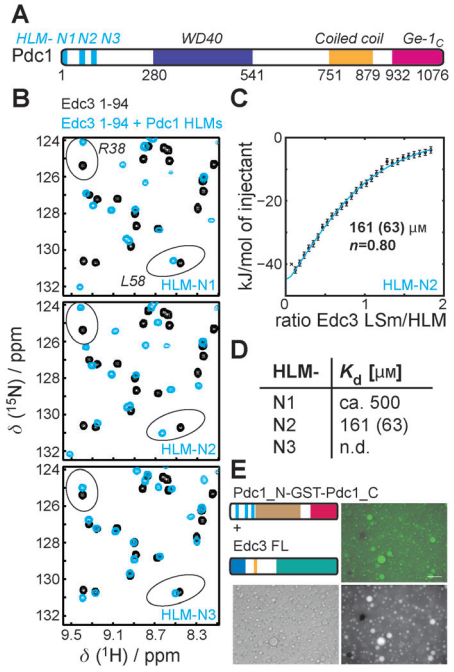


Figure 4. Pdc1 contains three HLMs at the N-terminus that bind to Edc3 with different affinities. A) Domain organization of the Pdc1 protein (N-terminal HLMs cyan, WD40 repeats purple, coiled-coil orange, Ge-1_C-like domain (see below) pink). B) ^1H–^{15}N correlation spectra of the monomeric ^{15}N labeled Edc3 LSm domain in the absence (black) and presence (cyan) of the individual Pdc1 HLMs. The boxed region in the top left panel of Figure 1B is shown, ovals highlight two specific residues (see above). C) ITC graph of the Edc3 LSm domain binding to HLM-N2 from Pdc1. The best fit is drawn with a cyan line and the extracted K_d value is indicated. D) Overview of all K_d values (error in parentheses) determined for binding of Pdc1 HLMs to Edc3 LSm (n.d., not determinable with ITC). E) In vitro phase separation of designed dimeric Pdc1 construct that mimics full length Pdc1 (50 μM) together with Edc3 (150 μM; doped 1:100 with Edc3-OregonGreen). Scale bar, 50 μm.

Sequence alignments of the C-terminal region of the *S. pombe* Pdc1 protein with the C-terminal regions of the human, *D. melanogaster* and *A. thaliana* Edc4 proteins suggest the presence of a Ge-1_C domain[20] in the Pdc1 protein, although the sequence identity (17%) is very low (Figure S2A). To confirm the presence of a Ge-1_C domain in Pdc1, we solved the crystal structure of residues 932 to 1076 to a resolution of 1.35 Å (PDB code: 4Q2S; Table S2). The structure displays a closely packed helical bundle, where the three N-terminal helices make an approximate 90-degree angle with the five C-terminal helices (Figure 5A). The structure of the domain is very similar to the known structure of the Ge-1_C region of the *D. melanogaster* Ge-1 protein (Figure S2B).

Previously, it was shown that the C-terminal 290 residues of Pdc1 interact with Dcp2.[10] To probe if this interaction is direct, we performed NMR chemical shift titration experiments with the ^{15}N labeled Pdc1 Ge-1_C domain and an unlabeled Dcp1:Dcp2 decapping complex (Dcp2 residues 1–95). We observe that resonance signals from residues that are located in the C-terminal helices of the Pdc1 Ge-1_C domain are significantly perturbed (Figure 5B). This result indicates that Pdc1 exploits a surface at the far C-terminus of the protein to directly contact the Dcp1:Dcp2 decapping complex (Figure 5A). Pdc1 can thus interact with Edc3 through a number of HLMs at its far N-terminal region and with the decapping complex through the Ge-1_c domain at its far C-terminal region. These results underscore the scaffolding

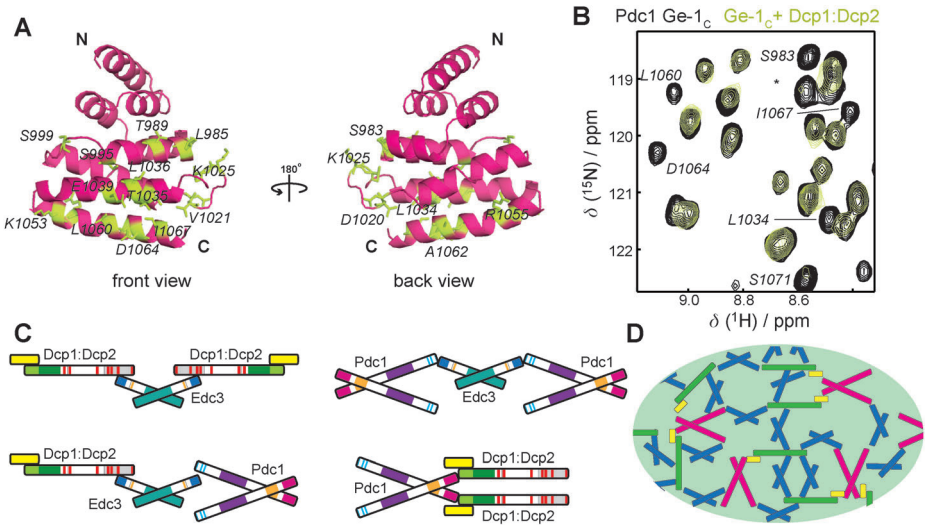


Figure 5. A) Ribbon diagram of the crystal structure of the Pdc1 Ge-1_C domain. Residues that interact with the decapping complex are highlighted in olive. B) NMR spectra of the Pdc1 Ge-1_C domain in the absence (black) and presence (olive) of the Dcp1:Dcp2 decapping complex that contains only the Dcp2 regulatory region. Resonances that experience chemical shift changes are labeled with the residue number. "*" refers to a resonance that is not assigned. C) The building blocks of the network of intermolecular interactions that leads to phase separation: One Edc3 dimer can interact with two decapping enzymes (left top), with two Pdc1 proteins (right top) or with one decapping complex and one Pdc1 dimer (bottom left). In addition one Pdc1 dimer can interact with two decapping complexes (bottom right). D) Schematic representation of how the building blocks displayed in (C) can be extended into an indefinite network of interactions. Edc3 blue, Dcp1 yellow, Dcp2 green, Pdc1 pink. These intermolecular interactions can result in phase separation (see Figures 2 and 4).

function of Pdc1 and explain its importance for processing body formation. Addition of Pdc1 to Edc3 and Dcp2 (Figure 2B) is thus expected to increase the strength of the interaction network, which is advantageous for the phase transitions process.

We have shown that it is possible to reproduce the in vivo phase separation behavior of the mRNA degradation machinery in a well-defined in vitro setting. Importantly, this bottom-up approach provides insights into the clustering behavior that could not have been obtained using genetic approaches owing to the highly redundant nature of the clustering process (Figure 5C,D).

Additional mechanisms that promote the self-aggregation process of the mRNA degradation machinery will be exploited in the cell. These additional interactions will be able to lower the critical concentration required for phase separation to biologically relevant levels. In that regard, it is worth mentioning that it was recently shown that low complexity (LC) regions in the RNA binding protein FUS are able to induce phase transitions.[6,12,21] Mechanistically, the LC regions aggregate through a mechanism that involves the formation of amyloid-like fibers, a feature that we did not detect in our current studies. This suggests that both processes that lead to phase separation are of a fundamentally different nature. Interestingly, short LC regions are also present in Dcp2 and Pdc1 and in other *S. pombe* processing body proteins, including LSm4, Pat1p, Sum2 (Scd6), Ste13 (Dhh1), and Exo2 (Xrn1).[22] Future experiments will be able to shed light on how these LC regions, the HLM:Edc3 interactions and other specific intermolecular contacts (Figure 5C) modulate the phase-transition process that underlies P-body formation (Figure 5D). It has been suggested that biological systems might have evolved such that specific proteins are close to the phase separation conditions.[6] Minor modifications in the valency and the affinity of the interacting partners by post-translational modifications or small changes in protein levels can then result in the sudden appearance or disappearance of cellular phase separation.[23] Interestingly, the HLM rich region of Pdc1 contains a large number of phosphorylation sites[24] that could potentially interfere with Edc3 binding and reduce the tightness of the intermolecular interaction network.

Our results form a starting point for future in vitro studies that address how the activity of the cellular enzymes can be influenced by local cellular molecular crowding. In addition, we anticipate that in vitro approaches similar to the one described herein can be exploited to address how other cellular granules are formed and how specific proteins are targeted to a specific class of granules.

Received: February 28, 2014
Published online: May 26, 2014

Keywords: NMR spectroscopy · phase transitions · processing bodies · protein–protein interactions · self-assembly

[1] a) T. Eystathioy, E. K. Chan, S. A. Tenenbaum, J. D. Keene, K. Griffith, M. J. Fritzler, *Mol. Biol. Cell* **2002**, *13*, 1338–1351; b) V. I. Bashkirov, H. Scherthan, J. A. Solinger, J. M. Buerstedde, W. D. Heyer, *J. Cell Biol.* **1997**, *136*, 761–773; c) E. van Dijk, N. Cougot, S. Meyer, S. Babajko, E. Wahle, B. Seraphin, *EMBO J.* **2002**, *21*, 6915–6924.
[2] a) A. Eulalio, I. Behm-Ansmant, E. Izaurralde, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 9–22; b) M. Kulkarni, S. Ozgur, G. Stoecklin, *Biochem. Soc. Trans.* **2010**, *38*, 242–251.
[3] a) D. Ingelfinger, D. J. Arndt-Jovin, R. Luhrmann, T. Achsel, *RNA* **2002**, *8*, 1489–1501; b) U. Sheth, R. Parker, *Science* **2003**, *300*, 805–808; c) C. J. Decker, D. Teixeira, R. Parker, *J. Cell Biol.* **2007**, *179*, 437–449.
[4] V. Balagopal, R. Parker, *Curr. Opin. Cell Biol.* **2009**, *21*, 403–408.
[5] J. Coller, R. Parker, *Annu. Rev. Biochem.* **2004**, *73*, 861–890.
[6] A. A. Hyman, K. Simons, *Science* **2012**, *337*, 1047–1049.
[7] D. Teixeira, U. Sheth, M. A. Valencia-Sanchez, M. Brengues, R. Parker, *RNA* **2005**, *11*, 371–382.
[8] D. Teixeira, R. Parker, *Mol. Biol. Cell* **2007**, *18*, 2274–2287.

[9] M. A. Reijns, R. D. Alexander, M. P. Spiller, J. D. Beggs, *J. Cell Sci.* **2008**, *121*, 2463 – 2472.
[10] C. Y. Wang, W. L. Chen, S. W. Wang, *Mol. Cell. Biol.* **2013**, *33*, 1244 – 1253.
[11] P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon, M. K. Rosen, *Nature* **2012**, *483*, 336 – 340.
[12] S. C. Weber, C. P. Brangwynne, *Cell* **2012**, *149*, 1188 – 1191.
[13] a) C. Gaudon, P. Chambon, R. Losson, *EMBO J.* **1999**, *18*, 2229 – 2240; b) S. A. Fromm, V. Truffault, J. Kamenz, J. E. Braun, N. A. Hoffmann, E. Izaurralde, R. Sprangers, *EMBO J.* **2012**, *31*, 279 – 290.
[14] a) C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, A. A. Hyman, *Science* **2009**, *324*, 1729 – 1732; b) N. Kedersha, G. Stoecklin, M. Ayodele, P. Yacono, J. Lykke-Andersen, M. J. Fritzler, D. Scheuner, R. J. Kaufman, D. E. Golan, P. Anderson, *J. Cell Biol.* **2005**, *169*, 871 – 884; c) A. Aizer, P. Kafri, A. Kalo, Y. Shav-Tal, *PLoS One* **2013**, *8*, e49783; d) C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4334 – 4339.
[15] M. She, C. J. Decker, D. I. Svergun, A. Round, N. Chen, D. Muhlrad, R. Parker, H. Song, *Mol. Cell* **2008**, *29*, 337 – 349.
[16] J. E. Braun, V. Truffault, A. Boland, E. Huntzinger, C. T. Chang, G. Haas, O. Weichenrieder, M. Coles, E. Izaurralde, *Nat. Struct. Mol. Biol.* **2012**, *19*, 1324 – 1331.
[17] F. Tritschler, J. E. Braun, A. Eulalio, V. Truffault, E. Izaurralde, O. Weichenrieder, *Mol. Cell* **2009**, *33*, 661 – 668.
[18] M. V. Deshmukh, B. N. Jones, D. U. Quang-Dang, J. Flinders, S. N. Floor, C. Kim, J. Jemielity, M. Kalek, E. Darzynkiewicz, J. D. Gross, *Mol. Cell* **2008**, *29*, 324 – 336.
[19] a) M. Fenger-Grøn, C. Fillman, B. Norrild, J. Lykke-Andersen, *Mol. Cell* **2005**, *20*, 905 – 915; b) J. H. Yu, W. H. Yang, T. Gulick, K. D. Bloch, D. B. Bloch, *RNA* **2005**, *11*, 1795 – 1802.
[20] M. Jinek, A. Eulalio, A. Lingel, S. Helms, E. Conti, E. Izaurralde, *RNA* **2008**, *14*, 1991 – 1998.
[21] a) T. W. Han, M. Kato, S. Xie, L. C. Wu, H. Mirzaei, J. Pei, M. Chen, Y. Xie, J. Allen, G. Xiao, S. L. McKnight, *Cell* **2012**, *149*, 768 – 779; b) M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, S. L. McKnight, *Cell* **2012**, *149*, 753 – 767.
[22] J. C. Wootton, S. Federhen, *Methods Enzymol.* **1996**, *266*, 554 – 571.
[23] C. F. Lee, C. P. Brangwynne, J. Gharakhani, A. A. Hyman, F. Julicher, *Phys. Rev. Lett.* **2013**, *111*, 088101.
[24] A. Koch, K. Krug, S. Pengelley, B. Macek, S. Hauf, *Sci. Signaling* **2011**, *4*, rs6.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

In Vitro Reconstitution of a Cellular Phase-Transition Process that Involves the mRNA Decapping Machinery**

*Simon A. Fromm, Julia Kamenz, Erik R. Nöldeke, Ancilla Neu, Georg Zocher, and Remco Sprangers**

anie_201402885_sm_miscellaneous_information.pdf

Experimental section

Molecular Cloning and Protein Purification.

All plasmids used for protein overexpression in *E. coli* strain BL21 (DE3) Codon Plus RIL (Stratagene) were generated using standard cloning techniques (Supplementary Table S1A). Transformed cells were grown at 37 °C to an OD_{595} of 0.6-0.8 in LB and protein production was induced with 1 mM IPTG at 20 °C. Cells containing protein for NMR, ITC and X-ray experiments were lysed in buffer A (50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 10 mM imidazole, 1 mM DTT) supplemented with 0.1 % Triton X-100, 0.5 mg/ml Lysozyme and 0.2 U/ml DNase I. The soluble fraction of the lysate was applied to Ni^{2+}-NTA resin equilibrated in buffer A. The retained protein fraction was eluted with buffer A containing 300 mM imidazole. After removal of the affinity tag with TEV protease, the protein was purified to homogeneity using size exclusion chromatography in SEC buffer (25 mM HEPES, pH 7.3, 125 mM NaCl, 1 mM DTT). HLM sequences were produced as MBP fusion constructs and were not digested with TEV protease.

NMR spectroscopy

Isotopically labeled proteins were grown in (H_2O or 70% D_2O based) M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source and/ or ^{1}H ^{13}C glucose as the sole carbon source. NMR spectra were recorded at 30 °C on Bruker AVIII-600 or AVIII-800 spectrometers. Titration experiments were carried out with 0.1 mM ^{15}N labeled protein (Edc3 LSm or Pdc1 Ge-1_C) and 0.5 mM unlabeled protein (MBP-HLMs or Dcp1:$Dcp2_{regulatory}$; containing full length Dcp1 and residues 1-95 of Dcp2). The backbone resonances of the Pdc1 Ge-1_C domain were determined using HNCACB and HN(CO)CACB pulse-sequences. Spectra were processed with the NMRPipe/NMRDraw software suite [1], figures showing NMR spectra were prepared using NMRView (onemoonscientific.com) and figures showing molecular structures were made using Pymol (pymol.org).

ITC

ITC data were recorded at 30 °C with a TA Instruments NanoITC Low Volume calorimeter, where the cell contained the individual HLM sequences into which the Edc3 LSm domain was titrated. ITC data were fitted using in house written scripts using an independent binding model.

X-ray crystallography

Residues 923 to 1076 of the Pdc1 protein crystallized in 0.2 M ammonium acetate, 0.1 M HEPES pH 7.5, 25% (w/v) PEG 3350. Data collection at 100K was performed at beamlines X10SA and X06DA at the Swiss Light Source, Villigen, Switzerland. Three data sets were recorded to determine the structure of the Ge-1_C domain of the Pdc1 protein. A high-resolution data set resulted from data acquisition at 1.0 Å wavelength diffracting to 1.35 Å resolution. Two additional crystals of Pdc1 served for data recording at 1.7 Å and 2.07 Å wavelength to establish phases using the anomalous signal of proteinaceous sulfur. All data sets were reduced using the XDS/XSCALE package [2], are of spacegroup P21212 and contain one protein chain in the asymmetric unit. Due to non-isomorphism initial structure determination was performed using both long-wavelength data sets. The sulfur sites were determined by SHELXD [3] using the data set recorded at 2.07 Å. Phase refinement as implemented in SHARP/AUTOSHARP [4] was performed including the data sets recorded at 1.7 Å and 2.07 Å. The electron density map resulting from density modification was interpretable and allowed manual model building using COOT [5]. A crude model consisting of 75 residues was transferred to the high-resolution dataset using PHASER [6]. After initial simulated annealing, refinement proceeded in alternating rounds of manual model rebuilding in COOT and restrained coordinate and anisotropic B-factor refinement in REFMAC [7]. Solvent molecules were added with COOT:find_waters. The structure was validated using Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC. Ramachandran statistics are as follows: 96.9% in favored regions and 3.1% in additional allowed regions.

Phase separation and fluorescent microscopy.

Proteins for phase separation microscopy experiments were purified in a buffer that contained 25 mM phosphate, pH7.3, 200 mM NaCl and 1 mM DTT. Proteins were fluorescently labeled with Oregon Green in the same buffer, however, in the absence of DTT. To optimize the labeling efficiency Oregon Green maleimide (dissolved in dimethylformamide) was added in 5-10 fold excess. The coupling reaction was incubated for 2 h at room temperature and contained less than 1 % DMF v/v. The reaction was stopped by addition of DTT to a final concentration of 4 mM. Uncoupled dye was removed from the protein fraction using a desalting PD-10 column and the coupling efficiency was determined via UV/Vis spectroscopy. A C-terminal cysteine was added to proteins that contained no accessible natural cysteine. Phase separation experiments were carried out in special 96-well microscopy plates (MatTek Corp.). The final sample volume per well was 50-80 µl. Images were

recorded at room temperature on a Zeiss Axio Observer.Z1 coupled to an AxioCam MRM camera after mixing of the different components. All images were processed with ImageJ (NIH, USA).

***S. pombe* strains and imaging.**

S. pombe strains used in this study are listed in Supplementary Table S1B. PCR-based gene targeting was used to replace genes by gene fusions at their endogenous loci[8]. For overexpression of the HLM-1 the coding sequence of amino acids 242 to 291 of the *dcp2+* gene fused to the 13myc epitope[9] were cloned into a pDual vector[10] under the control of the constitutive *adh1* promoter[11] and integrated into the *leu1* locus. Cell growth and imaging conditions for the visualization of P-bodies have been described previously[12].

SUPPORTING FIGURES

Figure S1

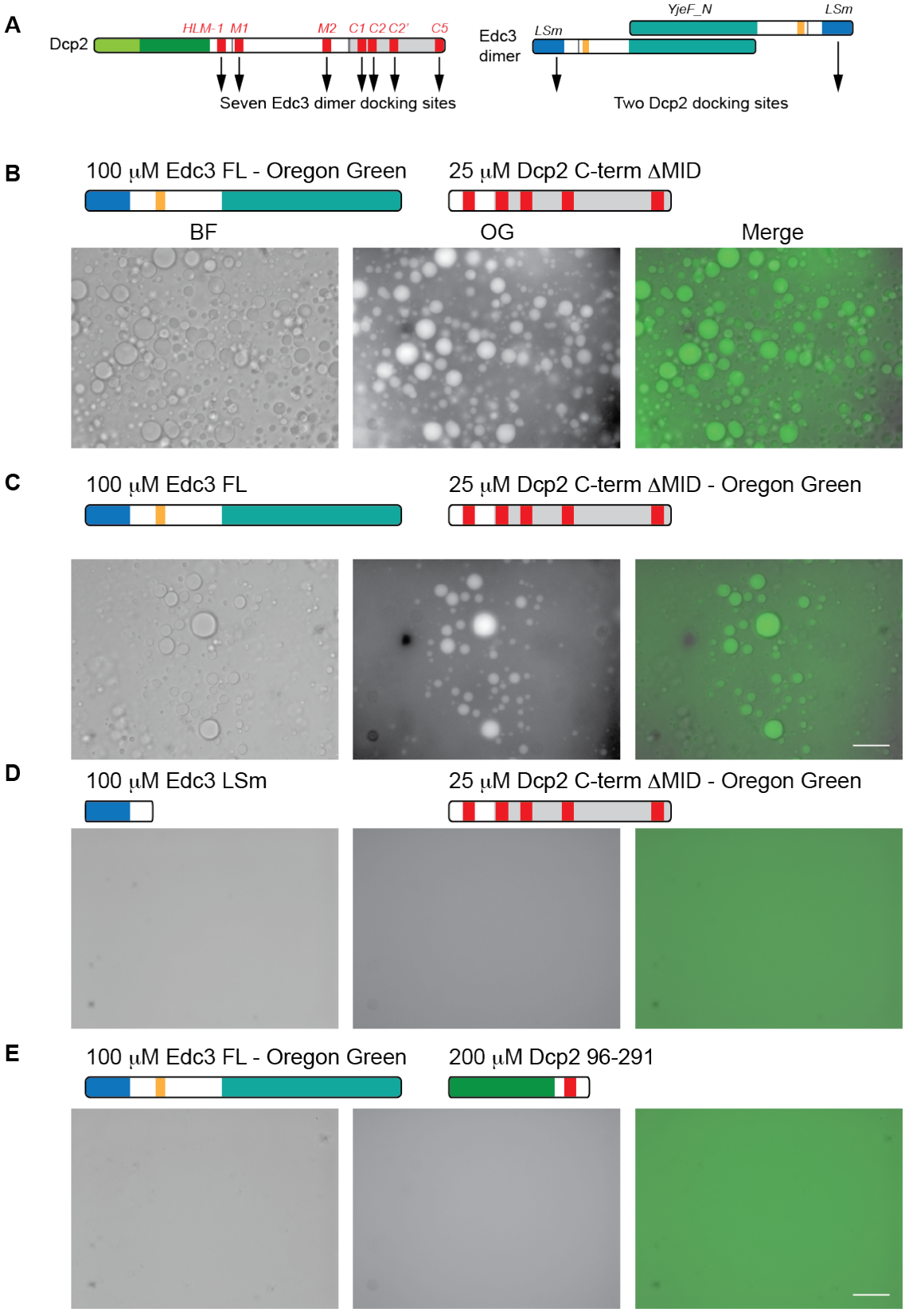
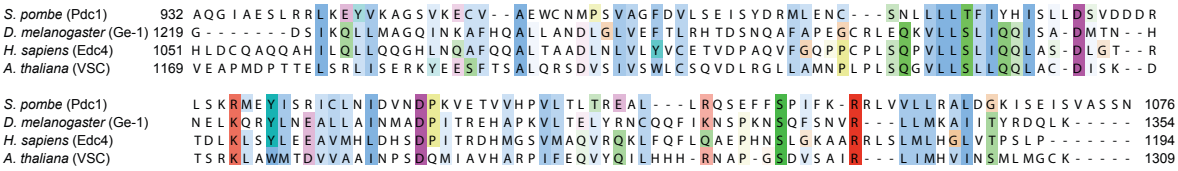


Figure S1 Phase separation of the Dcp2 and Edc3 proteins. (A): The Dcp2 protein has seven docking sites (HLM sequences) for the LSm domain of Edc3. The dimeric Edc3 protein has two LSm domains that can interact with HLM sequences. (B, C): Both the Edc3 and the Dcp2 protein are enriched in the droplet phase. In (B), the Edc3 protein is fluorescently labeled with Oregon green, in (C) the Dcp2 protein is fluorescently labeled with Oregon green. (D): A monomeric version of the Edc3 protein is not able to undergo phase transitions with Dcp2. (E) A Dcp2 protein that contains a single HLM is not able to undergo phase transitions with the dimeric Edc3 protein. The proteins that are fluorescently labeled are indicated. The scale bars correspond to 50 μm.

Figure S2

A



B

S. pombe Pdc1 Ge-1_C

D. melanogaster Ge-1 Ge-1_C (PDB: 2VXG)

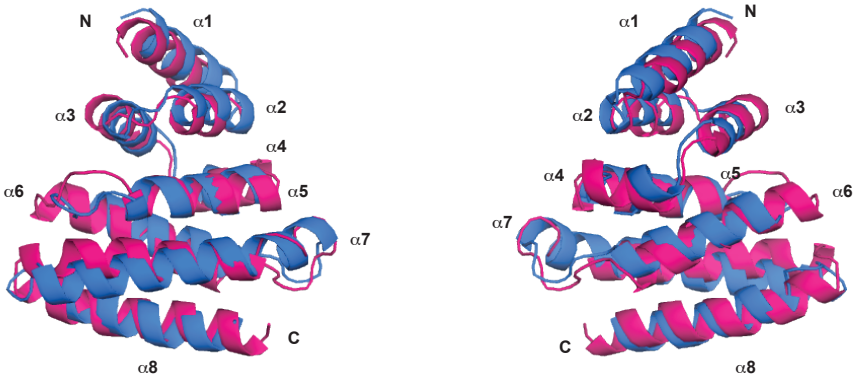


Figure S2 (A): Sequence alignment of the *S. pombe*, *D. melanogaster*, *H. sapiens* and *A. thaliana* Edc4 homologs. The residue numbers are indicated and the alignment is colored according to residue type, where the color intensity correlates with sequence conservation. (B): Overlay of the structures of the Pdc1 Ge-1_C domain and the *D. melanogaster* Ge-1_C domain of the Ge-1 protein [13]. The view is identical to the view in Figure 5 of the main paper.

SUPPORTING TABLES

Protein/Protein complex	Residues	Purification/Solubility-Tag	Used in Figure	Internal reference
Edc3	1-94	N-His_6-TEV	1B, 1C, 1D, 4B, 4C, 4D, S1D	194
Dcp2	242-291 (HLM-1)	N-His_6-MBP-TEV	1B, 1C, 1D, 2E	226
Dcp2	292-318 (HLM-M1)	N-His_6-MBP-TEV	1B, 1D	882
Dcp2	482-504 (HLM-M2)	N-His_6-MBP-TEV	1B, 1D	895
Dcp2	553-576 (HLM-C1)	N-His_6-MBP-TEV	1B, 1C, 1D	302
Dcp2	577-604 (HLM-C2)	N-His_6-MBP-TEV	1B, 1C, 1D	880
Dcp2	605-640 (HLM-C2')	N-His_6-MBP-TEV	1B, 1D	881
Dcp2	709-741 (HLM-C5)	N-His_6-MBP-TEV	1B, 1D	329
Edc3	1-454	N-His_6-TEV	2A-E, 4E, S1B, S1C, S1E	180
Dcp2	553-741	N-His_6-TEV	2A-C	68
Dcp2 Dcp2 C-term ΔMid	242-741Δ290-552	N-His_6-TEV	2B, 2E, S1B, S1D	752
Dcp2 C742 Dcp2 C-term ΔMid	242-741Δ290-552 (cysteine at C-terminus)	N-His_6-TEV	S1C	764
Dcp1:Dcp2 C742 Dcp1:Dcp2 ΔMid	1-127 (Dcp1), 1-741Δ290-552 (cysteine at C-terminus)	N-His_6-TEV	2D	769
Control	MBP	N-His_6-MBP-TEV	1B, 4B	189
Pdc1	1-1076	N-His_6-TEV	text	829
Pdc1-GST-Pdc1	1-105 (Pdc1)-GST-880-1076 (Pdc1)	N-His_6-TEV	4E	928
Pdc1	932-1076	N-His_6-TEV	5A, 5B, 5C	892
Dcp1:Dcp2	1-127 (Dcp1), 1-95 (Dcp2)	N-His_6-TEV (Dcp1), none (Dcp2)	5B, 5C	50
Dcp2 C292	96-291 (cysteine at C-terminus)	N-His_6-TEV	S1E	767
Pdc1	1-25 (HLM-N1)	N-His_6-MBP-TEV	4B, 4D	884
Pdc1	49-73 (HLM-N2)	N-His_6-MBP-TEV	4B, 4C, 4D	885
Pdc1	80-104 (HLM-N3)	N-His_6-MBP-TEV	4B, 4D	886

Table S1A: List of proteins used in this study, including construct boundaries and affinity/ solubility tags.

Figure 3A		
RS001	*h+*	*leu1 ade6-M216 dcp2+-GFP<<kanR edc3+-mCherry<<natR*
RS035	*h+*	*leu1 ade6-M216 dcp2+-GFP<<kanR edc3+-mCherry<<natR leu1+<<Padh1-dcp2-242-292-13myc*

Figure 3B		
RS037	*h+*	*leu1 ade6-M210 lsm7+-GFP<<kanR edc3+-mCherry<<natR*
RS039	*h+*	*leu1 ade6-M210 lsm7+-GFP<<kanR edc3+-mCherry<<natR leu1+<<Padh1-dcp2-242-292-13myc*

Table S1B: List of *S. pombe* strains used in this study.

	Pdc1 Ge-1_C domain
Data collection	
Space group	P 21 21 2
Cell dimensions	
a, *b*, *c* (Å)	47.33 77.84 36.51
a, b, g (°)	90 90 90
Resolution (Å)	40-1.35
R_{meas}	5.9(102.5)
R_{merge}	5.4(94.2)
I / sI	17.19(2.03)
Completeness (%)	99.7(100.0)
Redundancy	6.3(6.4)
Refinement	
Resolution (Å)	1.35
No. reflections	30287
R_{work} / R_{free}	18.5/21.2
No. atoms	
Protein	1092
Water	70
B-factors	
Protein	20.1
Water	29.9
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.43

Values in parentheses are for highest-resolution shell.

Table S2: X-ray data collection and refinement statistics.

SUPPORTING REFERENCES

[1] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *Journal of biomolecular NMR* **1995**, *6*, 277-293.

[2] W. Kabsch, *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 125-132.

[3] G. M. Sheldrick, *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 479-485.

[4] G. Bricogne, C. Vonrhein, C. Flensburg, M. Schiltz, W. Paciorek, *Acta crystallographica. Section D, Biological crystallography* **2003**, *59*, 2023-2030.

[5] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta crystallographica. Section D, Biological crystallography* **2010**, *66*, 486-501.

[6] A. J. Mccoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J Appl Crystallogr* **2007**, *40*, 658-674.

[7] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta crystallographica. Section D, Biological crystallography* **2011**, *67*, 355-367.

[8] J. Bahler, J. Q. Wu, M. S. Longtine, N. G. Shah, A. McKenzie, 3rd, A. B. Steever, A. Wach, P. Philippsen, J. R. Pringle, *Yeast* **1998**, *14*, 943-951.

[9] M. S. Longtine, A. McKenzie, 3rd, D. J. Demarini, N. G. Shah, A. Wach, A. Brachat, P. Philippsen, J. R. Pringle, *Yeast* **1998**, *14*, 953-961.

[10] A. Matsuyama, A. Shirai, Y. Yashiroda, A. Kamata, S. Horinouchi, M. Yoshida, *Yeast* **2004**, *21*, 1289-1305.

[11] M. McLeod, M. Stein, D. Beach, *The EMBO journal* **1987**, *6*, 729-736.

[12] S. A. Fromm, V. Truffault, J. Kamenz, J. E. Braun, N. A. Hoffmann, E. Izaurralde, R. Sprangers, *The EMBO journal* **2012**, *31*, 279-290.

[13] M. Jinek, A. Eulalio, A. Lingel, S. Helms, E. Conti, E. Izaurralde, *Rna* **2008**, *14*, 1991-1998.

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,

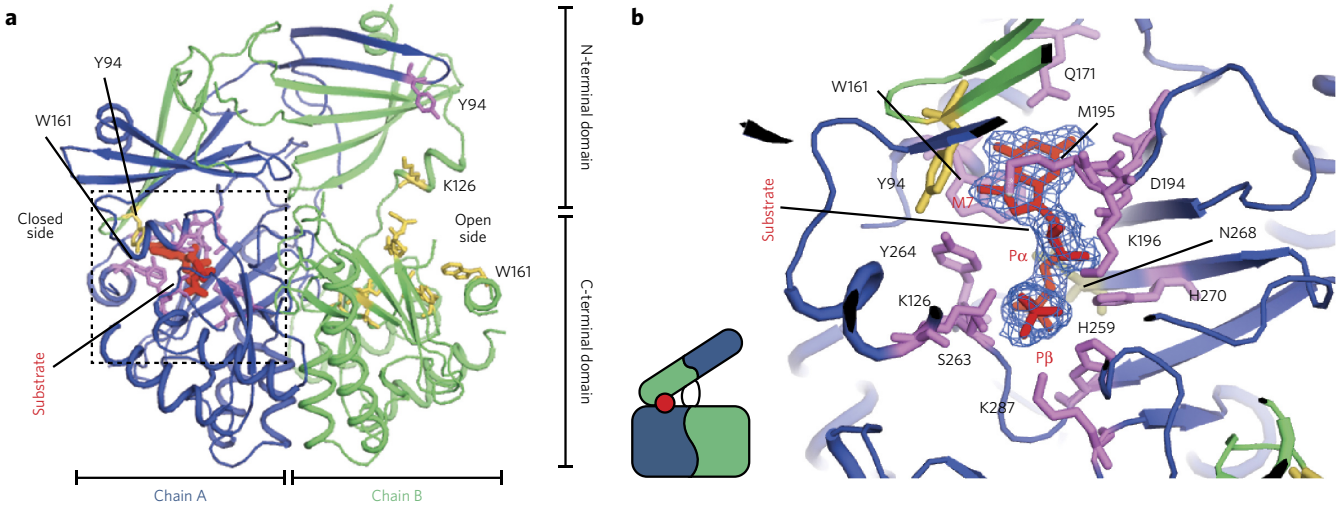


Figure 1 | Structure of the yeast DcpS–substrate complex. (**a**) 2.3 Å resolution crystal structure of the Dcs1p enzyme from *S. cerevisiae* in complex with m^{7}GDP. The dimeric enzyme adopts an asymmetric conformation with one closed and one open binding site. The protein chains are colored blue with 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

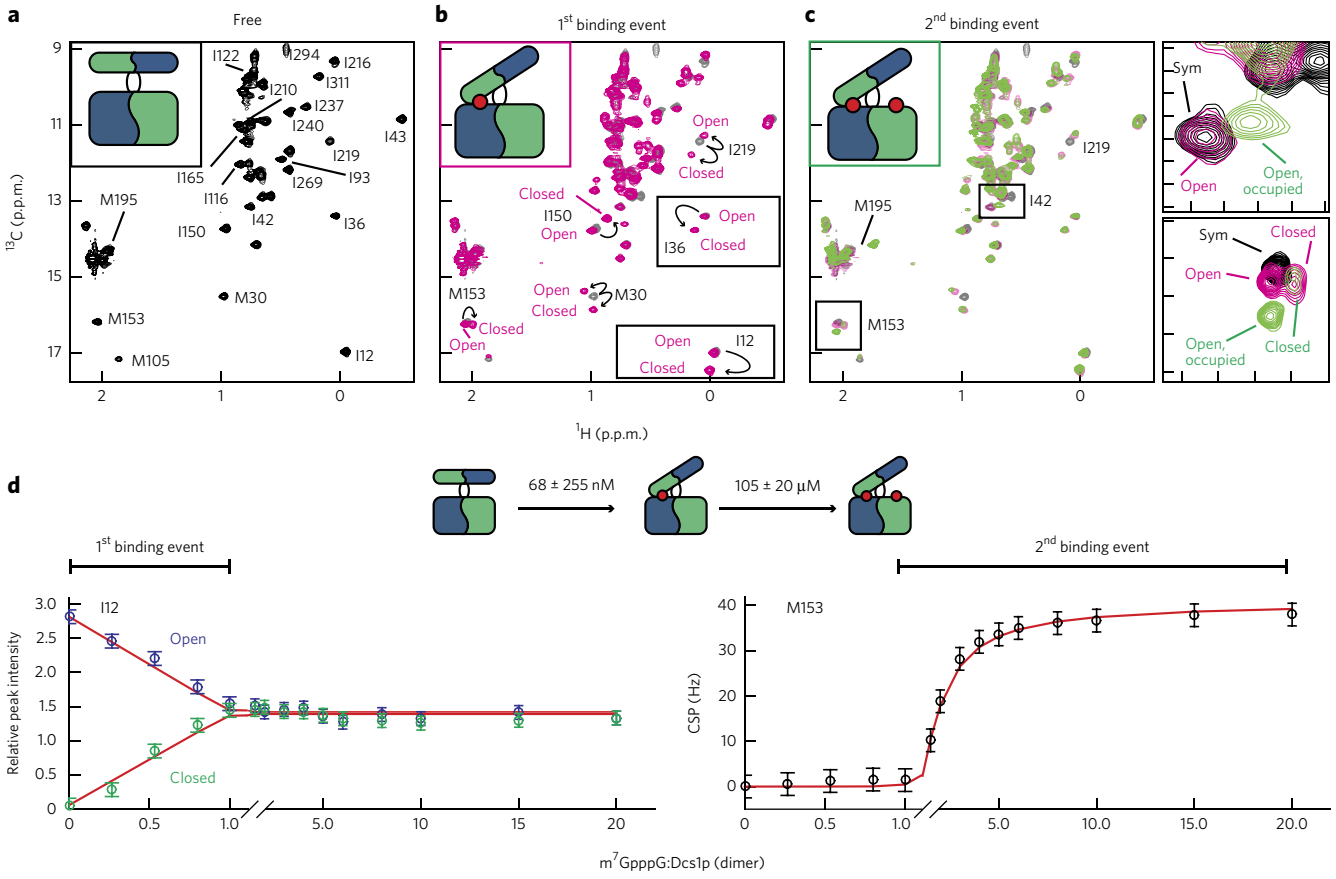


Figure 2 | The scavenger decapping enzyme binds to two substrate molecules in a sequential manner. (**a**–**c**) Methyl-TROSY NMR spectra of the yeast enzyme (0.25 mM dimer concentration) without (**a**) or with (**b**,**c**) substrate. Cartoons indicate the states of the complexes. (**a**) Ligand-free, 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.

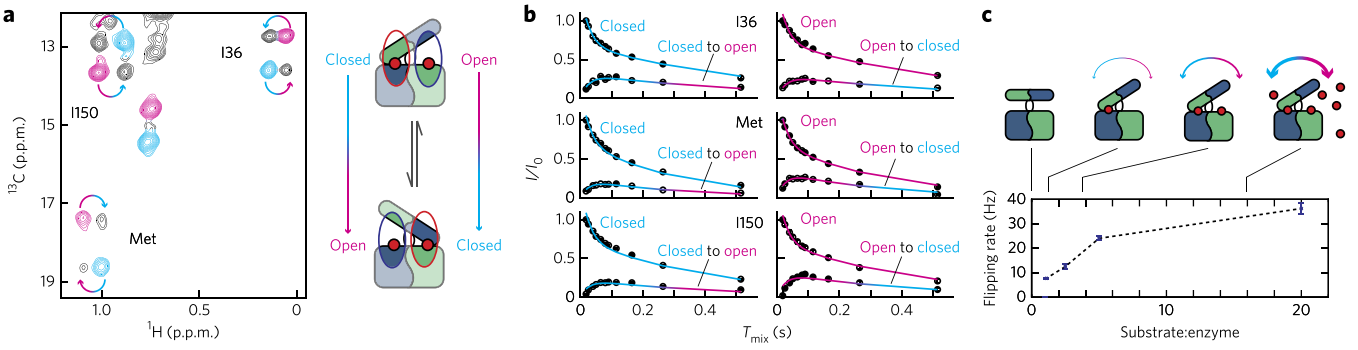


Figure 3 | Quantification of domain-flipping motions in the scavenger decapping enzyme. (**a**) Longitudinal exchange experiment that directly reports on motions (the open binding site closes and the closed binding site opens) in the enzyme. The cyan- and pink-colored resonances result from the closed 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.

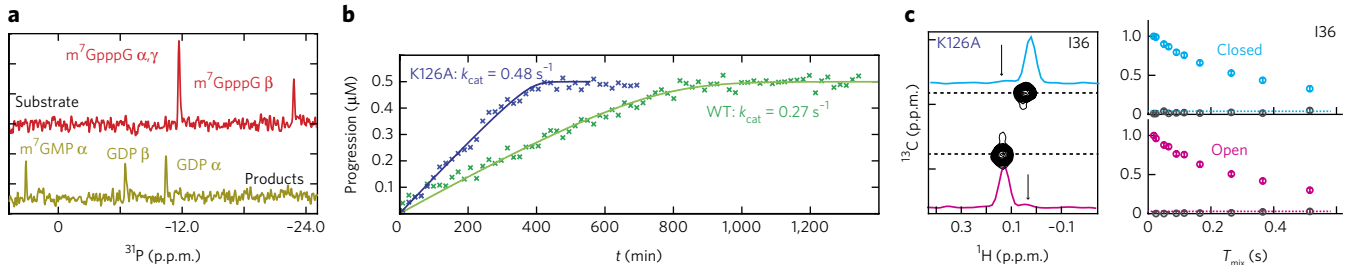


Figure 4 | Turnover rates of the enzyme are much slower than the flipping rates. (**a**) ^{31}P NMR spectra that report on the Dcs1p-mediated degradation (50 nM enzyme) of the (0.5 mM) m^{7}GpppG substrate (top, red) into m^{7}GMP and GDP products (bottom, yellow). NMR spectra (18 min each) were recorded 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].

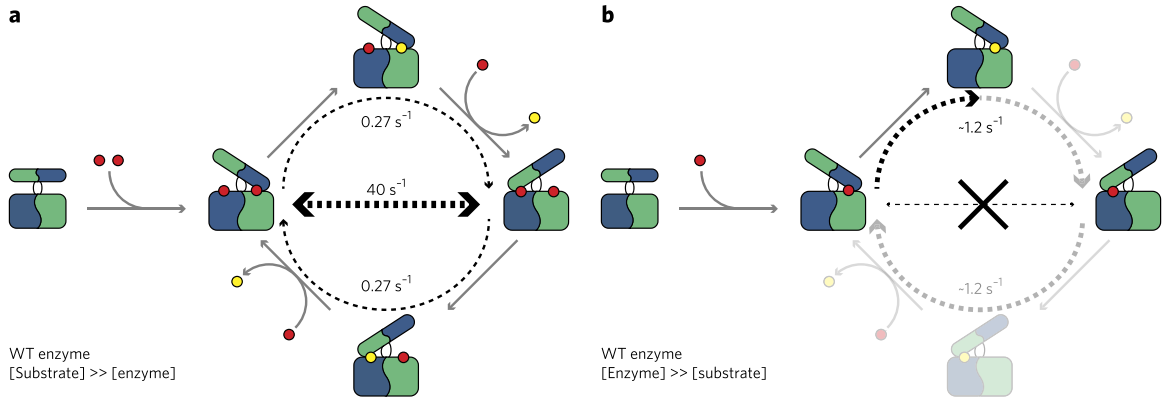


Figure 5 | Cartoon representation of the substrate inhibition mechanism of the scavenger decapping enzyme. (**a**) Under conditions of substrate excess, both binding sites in the enzyme are occupied. These result in fast, nonproductive flipping motions (horizontal dashed arrow) and low turnover rates (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

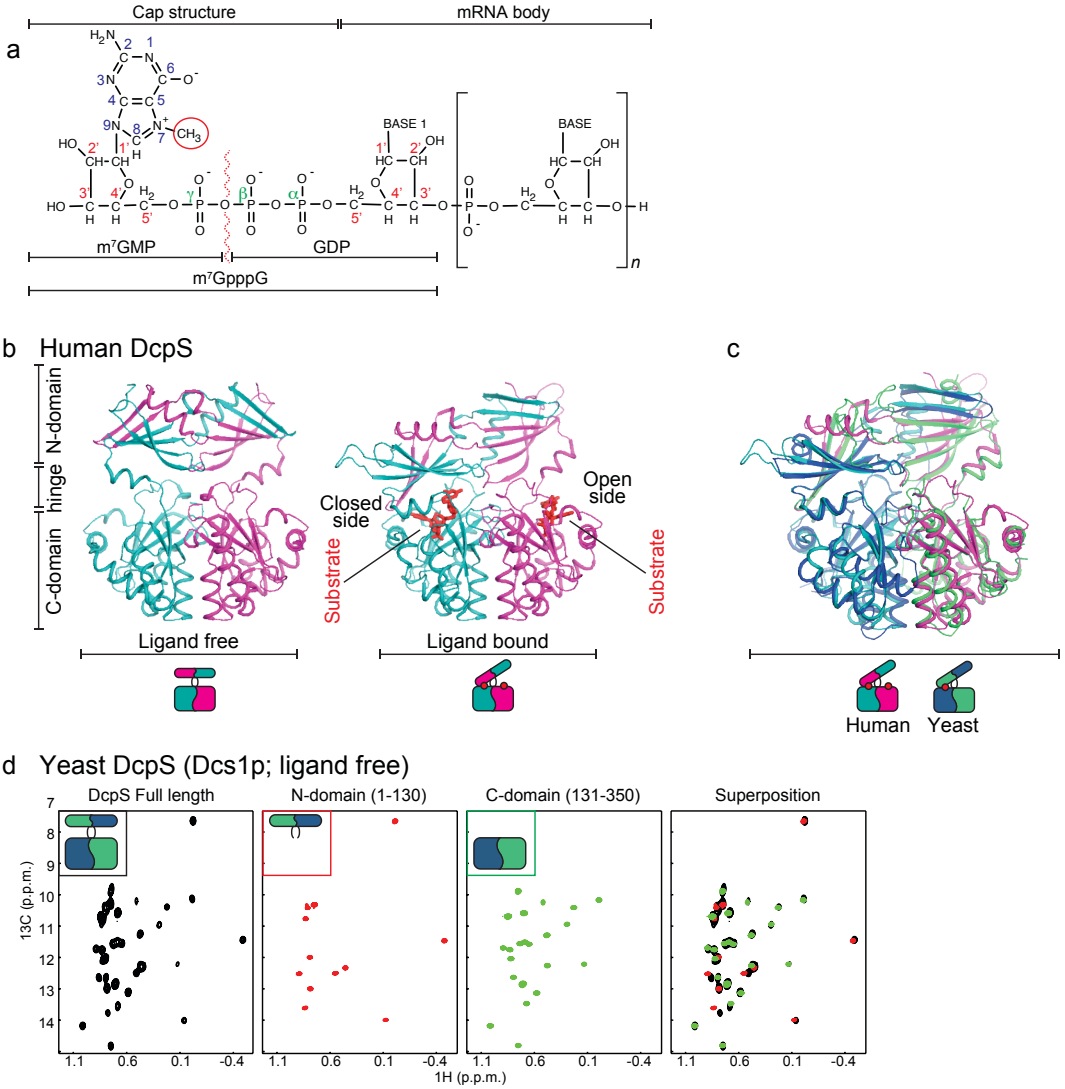
[3] Current address: The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

[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.

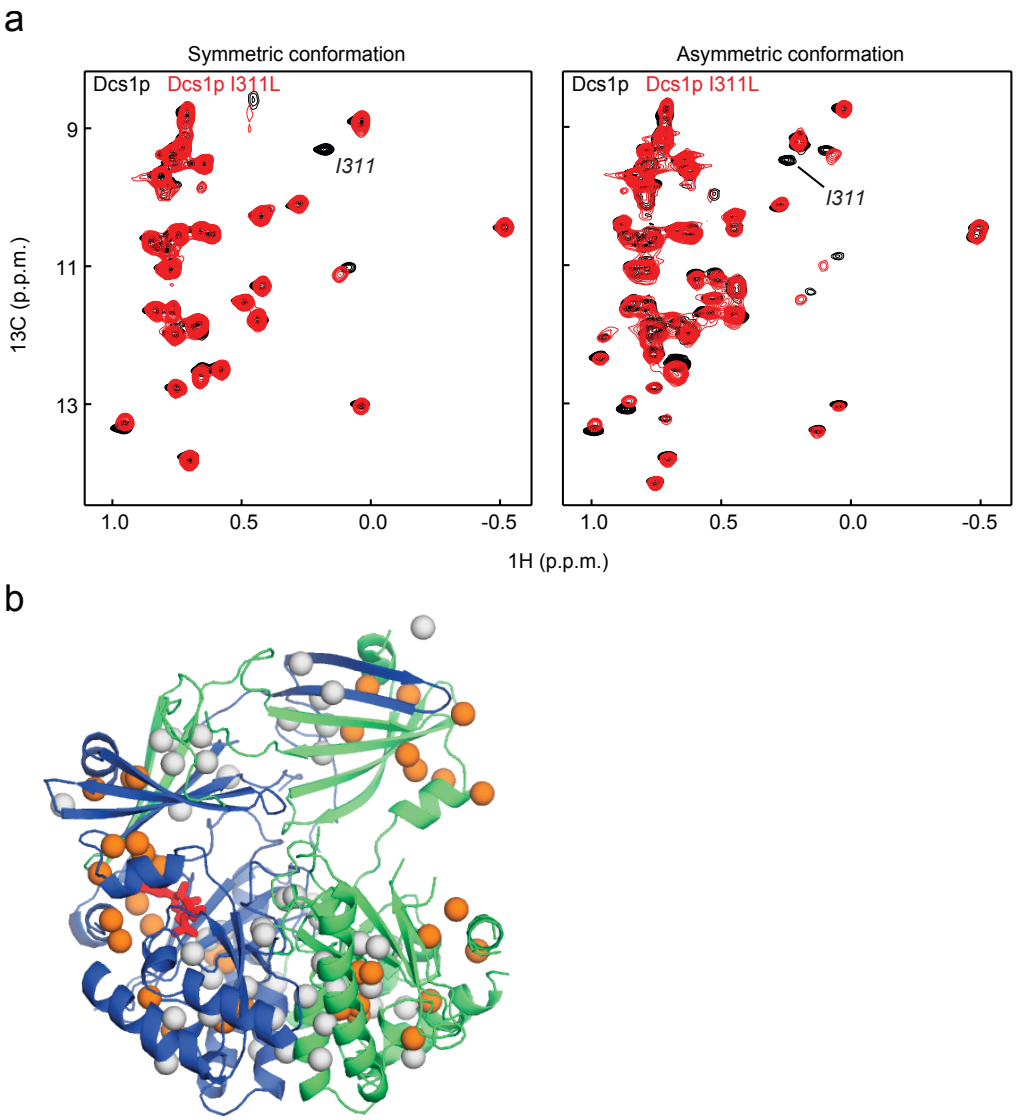
Nature Chemical Biology: doi:10.1038/nchembio.1866

(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

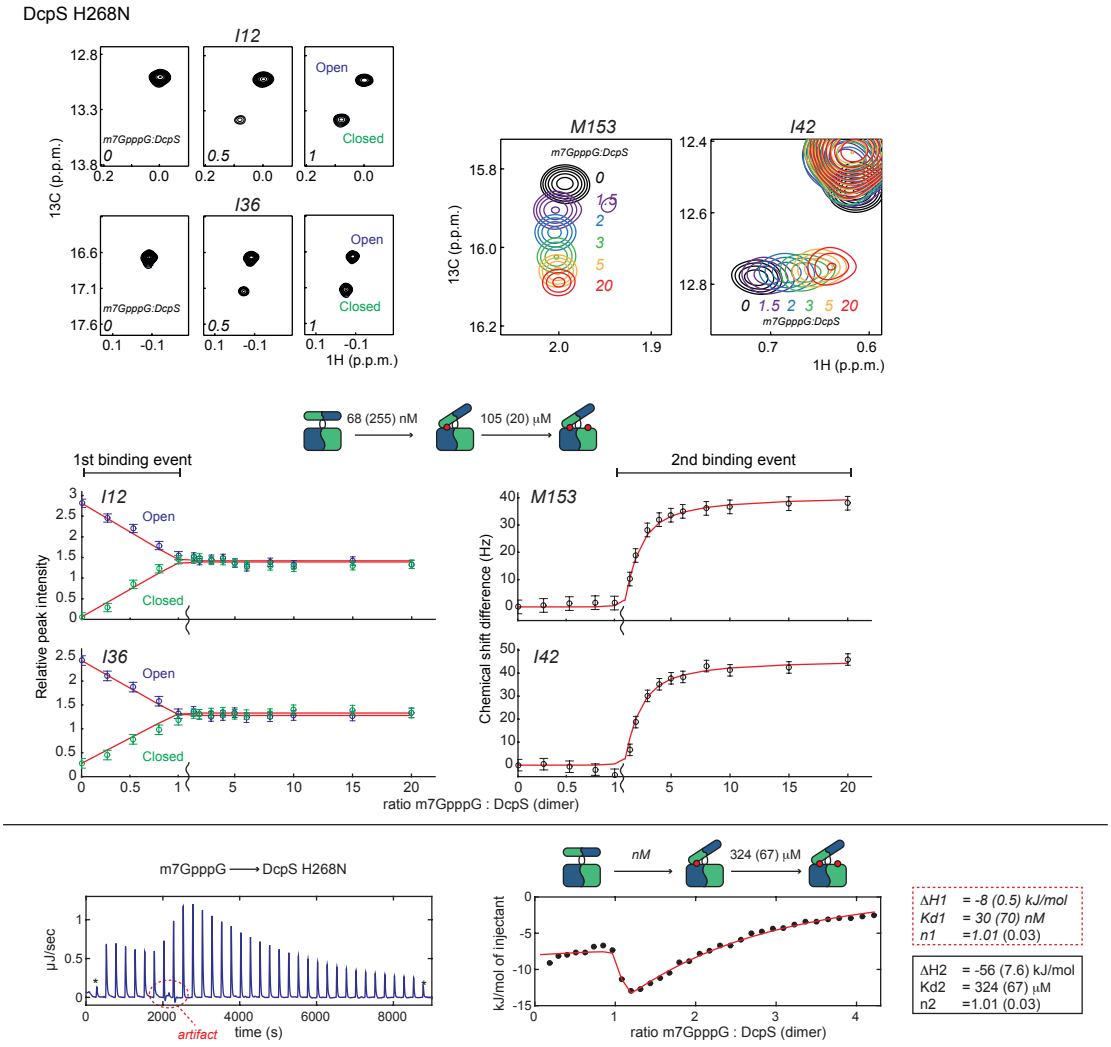


Supplementary Figure 2 | Assignment of methyl groups in the yeast scavenger decapping enyme.

(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:

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

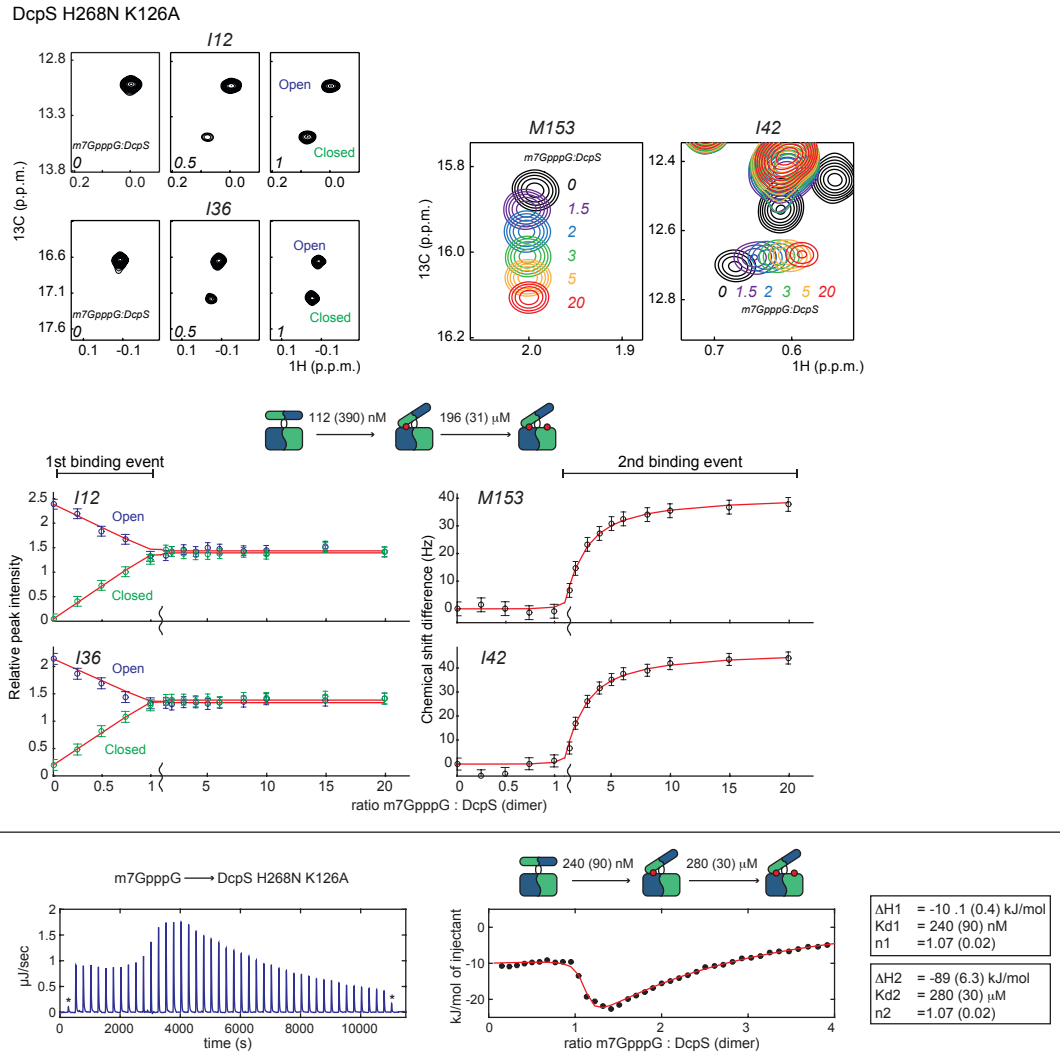
where V_0 is the volume of the ITC cell and ΔH_1 and ΔH_2 are molar heats of ligand binding for the first and second binding steps respectively. In the fitting procedure, the parameters ΔH_1, ΔH_2, K_a^1, K_a^2 and n (the total number of binding events per step) are optimized to minimize the difference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

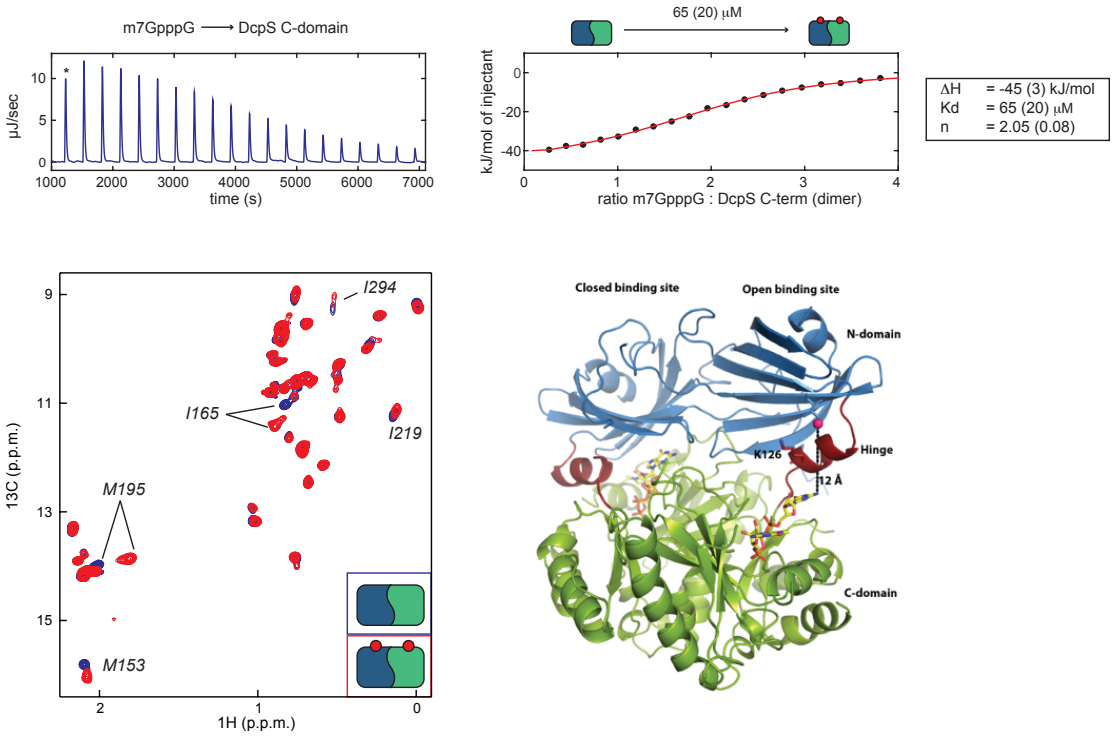
Bottom: ITC thermogram for the interaction of Dcs1p (H268N, K126A) with m^{7}GpppG that displays a sequential interaction mode. The red line corresponds to the fit of the data; the extracted binding constants are indicated next to the graph. Please note that the third event that we observed in the ITC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

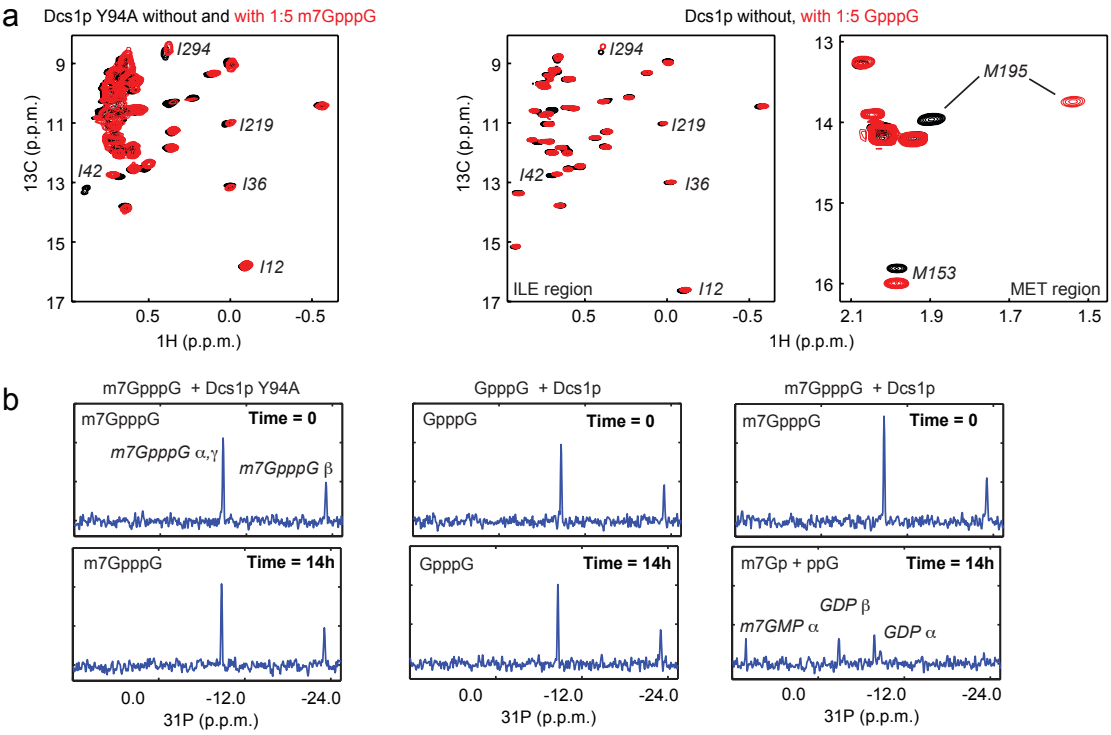
Bottom left: Methyl TROSY NMR spectrum of the C-terminal domain of Dcs1p in the absence (red) and presence (blue) of a 5-fold excess of m^{7}GpppG. The same residues experience chemical shift changes in the isolated domain and in the full-length protein (Figure 2c), indicating that the m^{7}GpppG substrate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

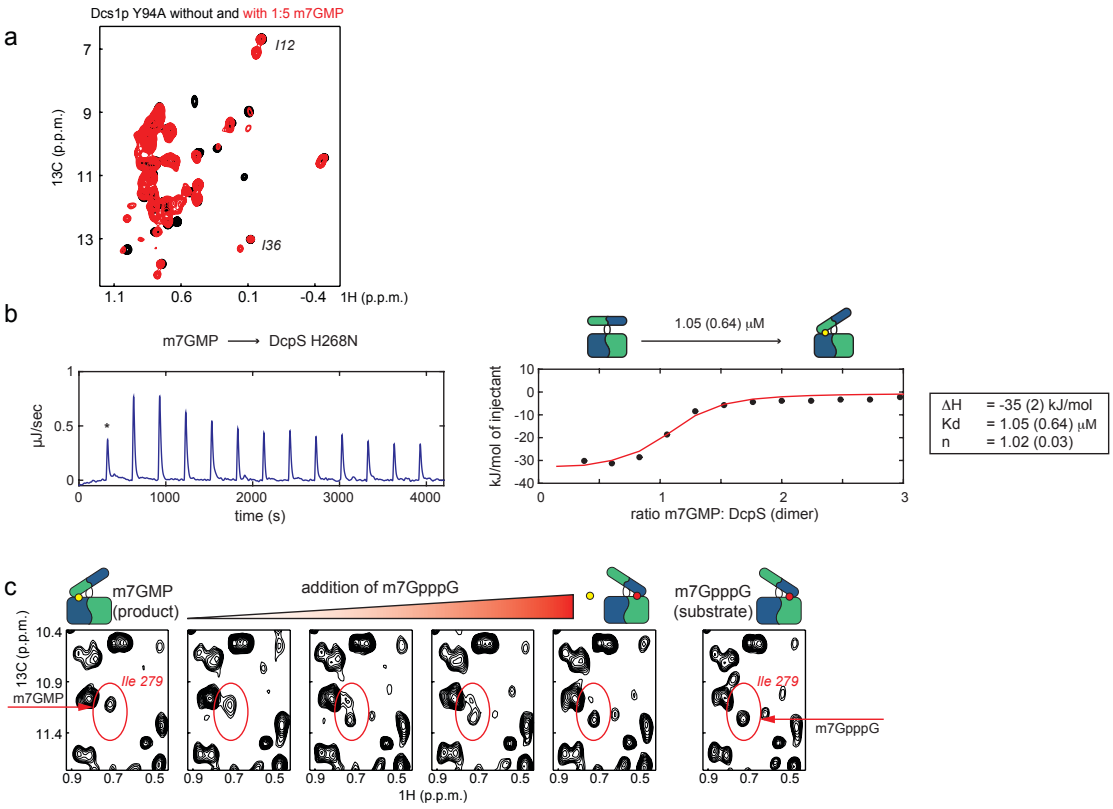
Nature Chemical Biology: doi:10.1038/nchembio.1866



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.

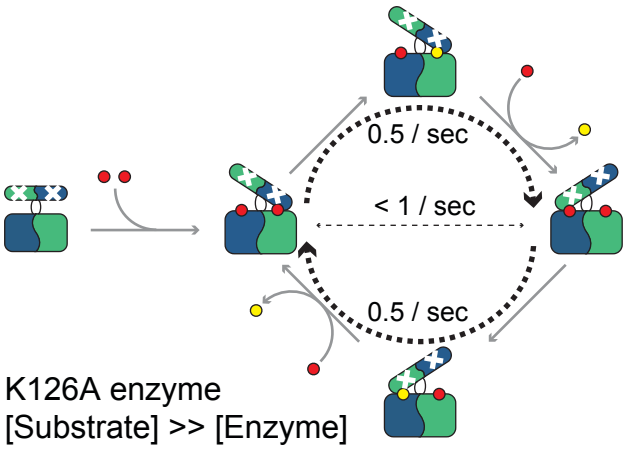
(b) DcpS binds the m^{7}GMP product with an affinity of 1.05 ± 0.64 µM in the first binding event. A second binding event is not observed, indicating that the product has no affinity for the open binding site. Using NMR spectroscopy and ITC we have not been able to detect any interaction of the GDP product 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.

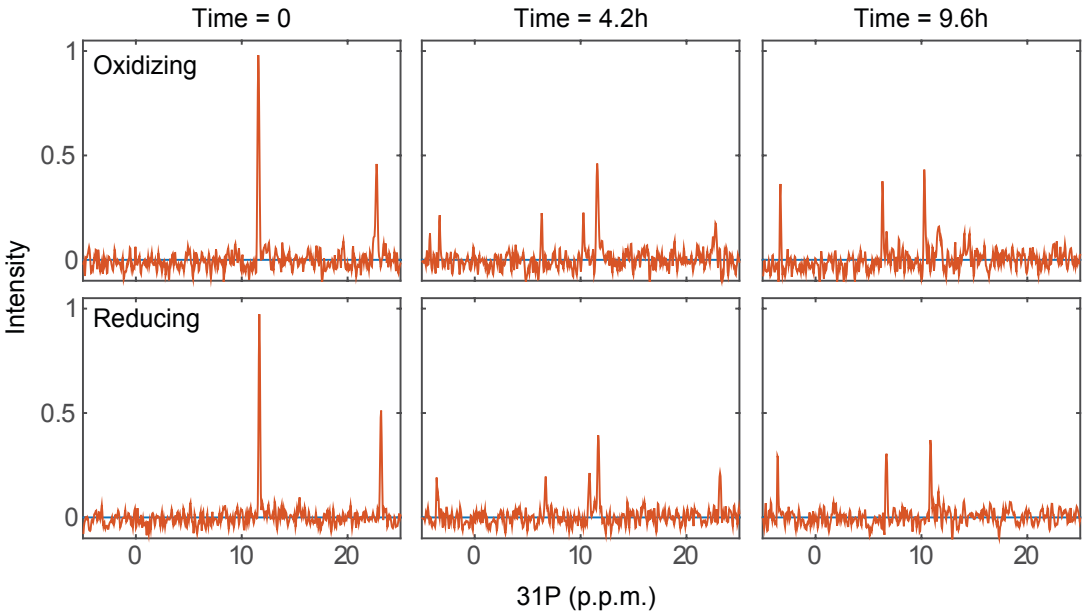
Nature Chemical Biology: doi:10.1038/nchembio.1866



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