NMR Studies of RNA Degradation by the Archaeal Exosome Complex

NMR–Studien zum RNA Abbau durch den Archaealen Exosomenkomplex

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Miloš Cvetković

aus Belgrad, Serbien

Tübingen

2017

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

Tag der mündlichen Qualifikation:	29.05.2017.
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Remco Sprangers
2. Berichterstatter:	Prof. Dr. Thilo Stehle

Table of Contents:

Summary

The exosome complex belongs to an evolutionarily conserved group of barrel–shaped molecular machines that have significant roles in RNA degradation and processing. This group comprises the major enzymes responsible for the 3′–5′ RNA degradation pathway, which are present in bacterial, archaeal and eukaryotic organisms.

The archaeal exosome complex (260 kDa) consists of a hexameric core and a trimeric cap. The active sites with phosphorolytic exoribonuclease activity are located within the internal chamber of the hexameric core. To enter the catalytic chamber, the substrate RNA molecule needs to pass through the opening on top of the assembly, which is encircled by the trimeric cap structure. The cap structure interacts with RNA, which provides substrate specificity and improves the RNA degradation efficiency. The underlying molecular mechanism, however, remains elusive. Here, we resorted to the novel approaches in NMR spectroscopy to address how the exosome cap structure is able to modulate the exosomal function. To that end, we use the archaeal exosome complex from the thermophile organism *Sulfolobus solfataricus* that is particularly amenable to the NMR studies as it has full three-fold symmetry and is stable at higher temperature.

During the PhD project, a long RNA binding surface was identified on the Rrp4–cap structure. It was shown that the RNA interacts with the exosome complex through four contact points that function independently from one another. Based on precise affinity measurements, we managed to deconstruct this multivalent exosome–RNA interaction and quantified the contributions of the individual contact points. This showed that the interaction contribution ("strength") increases from the periphery of the cap structure towards the active sites, which provides a mechanism that favours base-by-base substrate ratcheting towards the active sites. In addition to this actively degraded RNA, two more "waiting" RNA molecules can be bound on the Rrp4–cap structure. Interestingly, the interaction energy between the RNA and the exosome cap is significantly reduced after the 3′ end of the "waiting" RNA is translocated into the exosome chamber, thereby preventing friction between the substrate and the exosome cap during processive degradation. Furthermore, this project also emphasises the capabilities and advantages of methyl TROSY NMR spectroscopy for the studies of large macromolecular assemblies.

Zusammenfassung

Der Exosomenkomplex gehört zu einer evolutionär konservierten Gruppe von fassförmigen molekularen Maschinen, die bei RNA-Abbau und RNA-Verarbeitung signifikante Rollen aufweisen. Diese Gruppe umfasst die wichtigsten Enzyme, die für den 3'–5'-RNA-Abbaupfad verantwortlich sind, die in bakteriellen, archaealen und eukaryotischen Organismen vorhanden sind.

Der archaeale Exosomenkomplex (260 kDa) besteht aus einem hexameren Kern und einer trimären Kappe. Die aktiven Stellen mit phosphorolytischer Exoribonuklease-Aktivität befinden sich innerhalb der inneren Kammer des hexameren Kerns. Um in die katalytische Kammer einzutreten, muss das Substrat-RNA-Molekül durch die Öffnung auf der Oberseite der Anordnung hindurchgehen, die von der trimeren Kappenstruktur umschlossen ist. Die Kappenstruktur wechselwirkt mit RNA, die die Substratspezifität liefert und die RNA-Degradationseffizienz verbessert. Der zugrundeliegende molekulare Mechanismus ist jedoch noch unklar. Hier haben wir auf neue Ansätze in der NMR-Spektroskopie zurückgegriffen, um herauszufinden, wie die Exosomkappen-Struktur in der Lage ist, die exosomale Funktion zu modulieren. Zu diesem Zweck verwenden wir den archäischen Exosomenkomplex aus dem thermophilen Organismus *Sulfolobus solfataricus*, der besonders für NMR-Studien geeignet ist, da er eine volle Dreifachsymmetrie aufweist und bei höherer Temperatur stabil ist.

Während des PhD-Projektes wurde eine lange RNA-Bindeoberfläche auf der Rrp4-Kappen-Struktur identifiziert. Es wurde gezeigt, dass die RNA mit dem Exosomenkomplex über vier Kontaktpunkte, die unabhängig voneinander funktionieren, interagiert. Basierend auf präzisen Affinitätsmessungen haben wir diese multivalente Exosomen-RNA-Interaktion analysiert und die Beiträge der einzelnen Kontaktpunkte quantifiziert. Dies zeigte, dass der Wechselwirkungsbeitrag ("Stärke") von der Peripherie der Kappenstruktur zu den aktiven Stellen hin zunimmt, was einen Mechanismus zur Verfügung stellt, der ein Basen-für-Basen Fortschreiten des Substrats in Richtung der aktiven Stellen begünstigt. Zusätzlich zu dieser aktiv abgebauten RNA können an der Rrp4-Kappen-Struktur zwei weitere "Warte"-RNA-Moleküle gebunden werden. Interessanterweise wird die Wechselwirkungsenergie zwischen der RNA und der Exosomkappe signifikant reduziert, nachdem das 3'-Ende der "Warte"-RNA in die Exosomenkammer transloziert ist, wodurch Reibung zwischen dem Substrat und der Exosomenkappe während des Abbauvorgangs verhindert wird. Darüber hinaus unterstreicht dieses Projekt auch die Eignung und Vorteile der Methyl-TROSY-NMR-Spektroskopie für die Untersuchung großer makromolekularer Komplexe.

List of Publications

Peer-reviewed articles:

Paper 1:

Cvetkovic MA, Wurm JP, Audin MJ, Schütz S, Sprangers R. *The Rrp4–exosome complex recruits and channels substrate RNA by a unique mechanism*, Nat Chem Biol. 2017 Mar 13; 13(5):522-528

Paper 2:

Audin MJ, Wurm JP, **Cvetkovic MA**, Sprangers R. *The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation*, Nucleic Acids Res. 2016 Apr 7; 44(6):2962-73

Book Chapter:

Cvetkovic MA, Sprangers R. *Methyl TROSY spectroscopy to study large biomolecular complexes;* a chapter of the book: *Modern Magnetic Resonance*, 2nd edition, 2017; Springer International Publishing; ISBN: 978-3-319-28275-6

Abbreviations:

a/eIF2(γ)	Archaeal/eukaryotic translation initiation factor 2, γ subunit
β-CASP	Metallo-β-lactamase-assocoated CPSF Artemis SNM1/PSO2
Ccr4	Carbon catabolite repressor 4
CPSF	Cleavage and polyadenylation specificity factor
CR3	Cystein–rich repeat
crRNA	CRISPR ribonucleic acid
CSD1, CSD2	Cold-shock domains
Csl4	Cep1 synthetic lethal 4
CSP	Chemical shift perturbation
Dcp1, Dcp2	Decapping proteins
DcpS	Scavenger decapping enzyme
DNA	Deoxyribonucleic acid
eIF4e	Eukaryotic translation initiation factor 4E
Exo	Exosome
ITC	Isothermal titration calorimetry
KH	K-homology domain
Lsm1-7	Like-Sm proteins
miRNA	Micro ribonucleic acid
mRNA	Messenger ribonucleic acid
Mtr3	mRNA transporter protein 3
ncRNA	Noncoding ribonucleic acid
NDP	Nucleoside diphosphate
NMR	Nuclear Magnetic Resonance
Not	Negative on TATA
NTD	N-terminal domain
PABP	Poly(A)-binding protein
Pan2–Pan3	Poly(A) nuclease proteins
Pi	Inorganic phosphate
PIN	PiLT N terminus domain
piRNA	PIWI-interacting ribonucleic acid
PNPase	Polynucleotide phosphorylase

p.p.m.	Parts per million
RNA	Ribonucleic acid
RNase	Ribonuclease
RNB	Ribonuclease II–like domain
RNP	Ribonucleoprotein
RppH	RNA 5′ pyrophosphohydrolase
rRNA	Ribosomal ribonucleic acid
Rrp	Ribosomal RNA processing protein
siRNA	Small interfering ribonucleic acid
Ski complex	Superkiller complex
snRNA	Small nuclear ribonucleic acid
snoRNA	Small nucleolar ribonucleic acid
SPR	Surface plasmon resonance
TRAMP complex	Trf4/Air2/Mtr4 polyadenylation complex
tRNA	Transfer ribonucleic acid
TROSY	Transverse relaxation-optimised spectroscopy
UTR	Untranslated region

1. Introduction

1.1. Biomolecular Complexes

Biochemical processes in the cell used to be viewed as simple enzymatic reactions between individual molecules. With the developments in biochemistry, cell and structural biology over the past decades, our picture of the cell grew more and more intricate. We are now aware that the majority of pivotal cellular processes is performed by macromolecular assemblies *(Alberts 1998)* that consist of several polypeptide chains and, in certain cases, polynucleotides. Some of these processes are transcription, translation, RNA splicing, rRNA modifications, DNA replication, protein degradation and photosynthesis.

Components of a biomolecular complex can have different functions and they all act in a highly coordinated manner in order to carry out specific biological processes. Because of the coordinated action of their different parts, biomolecular complexes are often called molecular machines. Association of proteins is a form of their quaternary structure and illustrates a fundamental principle of modularity in nature. Modularity refers to a concept that a system can be viewed as a collection of smaller units, modules, each with a discrete function *(Hartwell, Hopfield et al. 1999)*. Modularity is present in the cell on several levels: domains are functional modules of a respective multidomain protein, protein subunits are functional modules of a biomolecular complex and biomolecular complexes are functional modules of certain cellular networks or biochemical pathways *(Pereira-Leal, Levy et al. 2006)*. Assemblies of identical (homo-oligomers) or different subunits (hetero-oligomers) provide many advantages over single proteins *(Marsh and Teichmann 2015; Pieters, van Eldijk et al. 2016)*. Subunits can build large structural elements, inside or outside of the cell, that could not be feasible by simply increasing the size of a single polypeptide chain; assembling of several subunits into a complex can substantially improve their stability; having distinct elements in a complex increases possibilities for regulation, including allostery; physical gathering of subunits enables linking their different functions; biomolecular complexes can adopt highly specific and elaborate shapes that enable their proper functions, like in the case of compartmentalization when certain reactions are physically secluded from the surroundings; interfaces between subunits can form

active or binding sites necessary for the function of the complex. Furthermore, from the evolutionary perspective, modularity provides means for acquiring new functional or structural features through different combinations of available elements *(Pereira-Leal, Levy et al. 2006; Marsh and Teichmann 2015)*.

1.2. RNA Degradation

RNA is a highly diverse group of essential molecules in the cell. It can be divided into mRNA that code for proteins and a long list of noncoding RNA (ncRNA) species with various, but equally important functions *(Cech and Steitz 2014)*. Some ncRNAs are crucial parts of the protein synthesis machinery, like rRNA and tRNA; others, like snRNA and snoRNA, are involved in processing and modification of other RNA species *(Matera, Terns et al. 2007)*; some ncRNAs are used as scaffolds for different protein binding partners, like 7SK RNA *(Nguyen, Kiss et al. 2001; Diribarne and Bensaude 2009)*; some ncRNAs have role in maintaining chromosome stability as part of telomerases or in protecting the genome against retrotransposons (piRNA) and foreign nucleic acids (crRNA); many ncRNAs are significantly involved in regulation of gene expression on different levels, for example miRNA, siRNA *(Carthew and Sontheimer 2009)* and various RNA elements like riboswitches *(Tucker and Breaker 2005)*. Moreover, outside of the cell, RNA can function as storage of genetic information as in the case of many viruses. All RNA molecules are normally bound to their protein partners, forming ribonucleoprotein (RNP) complexes.

RNA degradation is a carefully controlled process involved in general RNA turnover, RNA quality control and RNA processing *(Houseley and Tollervey 2009)*. Alteration of decay rates of certain mRNAs, as a result of external (environmental) or internal (intracellular) signals, represents one of the principal mechanisms of gene expression regulation. In that way the cell can adequately react to environmental changes or enter new stage of development at the right moment. mRNA quality control is a surveillance mechanism that ensures degradation of aberrant transcripts that could otherwise lead to production of malfunctional or even dangerous proteins. RNA degradation also takes part in processing of different RNA species through trimming or cleaving its longer precursor transcripts.

RNA molecules feature directionality, with a distinct 5′ end with a phosphate group and a 3′ end with a hydroxyl group. Degradation can start from either end of the polynucleotide chain or with a cut inside the chain. These three reactions are performed by three different sets of specialized enzymes *(Houseley and Tollervey 2009; Labno, Tomecki et al. 2016)*: 5′-3′ exoribonucleases (from the 5′ end), 3′-5′ exoribonucleases (from the 3′ end) and endonucleases (incision inside the chain). The best-known 5′-3′ exoribonucleases are cytoplasmic Xrn1 and nuclear Rat1 (Fig. 1, right panel). Major cellular components responsible for the 3′-5′ RNA degradation are barrel-shaped protein assemblies RNase PH, PNPase and the exosome complex. Since most RNA molecules are accompanied, *in vivo*, by proteins that protect both of their ends from degradation, exoribonucleolytic degradation usually consists of several steps and involves auxiliary proteins. Many different endoribonucleases are known, like members of the RNase III family that play an important role in the processing of different RNA precursors and that are essential for the RNA interference *(Nicholson 2014)*, or endoribonucleases RNase E and RNase Y that participate in RNA processing and degradation in bacteria *(Mohanty and Kushner 2016)*.

1.2.1. mRNA degradation in Eukarya

mRNA turnover is the most studied and the best understood process of RNA degradation, especially in eukaryotic organisms *(Parker 2012; Labno, Tomecki et al. 2016)*. Special protective features of the eukaryotic mRNA are the 7-methylguanosine cap structure linked via triphosphate to the 5′ end of the molecule (m7Gppp), and the poly(A) tale at the 3′ end. Poly(A)-binding proteins (PABP) are attached to the long poly(A) tail, while another protein, eIF4e, is bound to the cap (Fig.1, right panel). They both contribute to the stability of the mRNA by preventing degradation and they also promote translation. The critical step for the subsequent mRNA degradation is, normally, shortening of the poly(A) tail. This process is highly regulated and can be performed by several deadenylation complexes, predominantly by the Ccr4-Not complex and by the Pan2-Pan3 complex. After initial trimming of the poly(A) tail, the protective PABP detaches and mRNA degradation can proceed down two different pathways: 5′-3′ or 3′-5′ decay (Fig. 1, right panel). In the 5′-3′ degradation pathway, Lsm1-7/Pat1 binds the short poly(A) tail and afterwards recruits decapping complex Dcp1/Dcp2 through a so-far unknown mechanism. After decapping, degradation from the 5′ end of the mRNA is performed by the exoribonuclease Xrn1. In the 3′-5′

degradation pathway, trimming of the poly(A) tail and release of PABPs are followed by the recruitment of the mRNA to the exosome complex that carries out degradation from the 3′ end. A short RNA segment with the cap remains after this process and it is then decapped by the DcpS, scavenger decapping enzyme *(Neu, Neu et al. 2015)*.

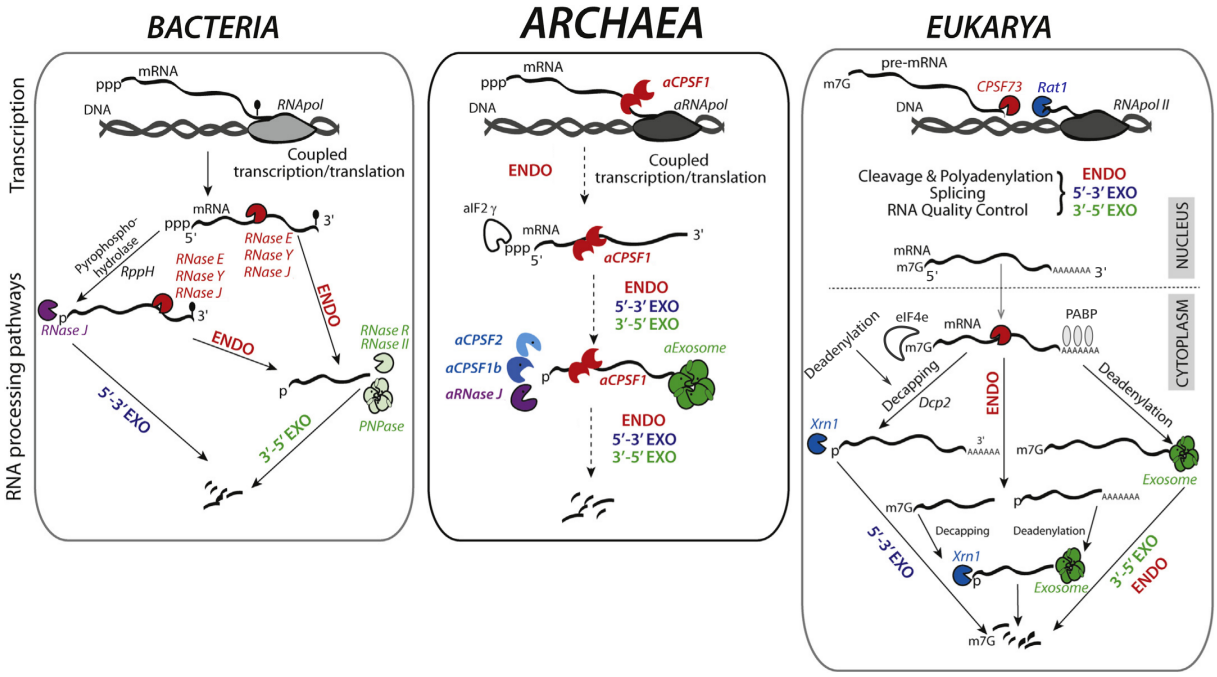


Fig. 1. Simplified overview of mRNA processing pathways in the three domains of life. Enzymes with endoribonucleolytic activities are coloured in red, those with 5′-3′ exoribonucleolytic activities in blue and purple, and those with 3′-5′ exoribonucleolytic activities in green. RNAse J is represented in both purple and red because it harbours both activities. Adapted from *Clouet-d'Orval, Phung et al. 2015*.

1.2.2. mRNA degradation in Bacteria

Bacterial mRNA has a triphosphorylated 5′ end and, often, a 3′-end stem-loop, which both function as protective structural features. For several mRNA molecules, it has been shown that highly structured 5′ UTRs serve as elements that increase stability *(Mohanty and Kushner 2016)*. 3′-end A-rich polynucleotide tails can be found in bacterial mRNAs, but they do not have protective role. On the contrary, they are usually signals for degradation or elements that facilitate degradation *(Mohanty and Kushner 2011)*. mRNA degradation in bacteria starts either with the removal of the pyrophosphate from the 5′ end by the enzyme RppH, or with the endonucleolytic cleavage by the enzymes RNase E, RNase Y or RNase J (Fig. 1, left panel) *(Hui, Foley et*

al. 2014). After the removal of the pyrophosphate from the 5′ end, in gram-positive bacteria like *Bacillus subtilis*, the enzyme RNase J that has both endo- and exoribonucleolytic activity, carries out degradation in the 5′-3′ direction. In gram-negative bacteria like *Escherichia coli*, the 5′-monophosphorylated mRNA is more prone to cleavage by the endoribonuclease RNase E. In the second degradation pathway, the initial endonucleolytic cleavage forms two RNA fragments. One fragment has its 3′ end devoid of the protective stem-loop and it is, therefore, susceptible to the 3′-5′ exoribonucleolytic activities of the RNase R, RNase II and PNPase. Second RNA fragment is released with a monophosphorylated 5′ end, and can thus be further degraded either exoribonucleolytically by the RNase J (in *Bacillus subtilis*) or endoribonucleolytically by the RNase E (in *Escherichia coli*) (Fig. 1, left panel).

1.2.3. mRNA degradation in Archaea

Degradation of mRNA in archaeal organisms is the least understood (Fig. 1, central panel). β-CASP family of ribonucleases, where the bacterial RNase J also belongs to, plays a pivotal role in this process *(Clouet-d'Orval, Phung et al. 2015)*. Similarly to bacteria, archaeal mRNA molecules have a triphosphorylated 5′ end, but lack any known 3′-end protective structure *(Evguenieva-Hackenberg and Klug 2009)*. 3′-end A-rich polynucleotide tails are present in mRNA molecules of several archaeal subgroups and they stimulate exoribonucleolytic degradation *(Portnoy and Schuster 2006; Mohanty and Kushner 2011)*. On the other hand, the translation initiation factor a/eIF2(γ) that is attached to the 5′ triphosphate, acts as a protection against degradation from that side, which resembles arrangement at the 5′ end of the eukaryotic mRNA *(Hasenohrl, Lombo et al. 2008)*. Archaeal RNase J (aRNase J) harbours only 5′-3′ exoribonucleolytic activity, the same as two other members of the β-CASP family found in Archaea, aCPSF1b and aCPSF2 *(Clouet-d'Orval, Phung et al. 2015)*. For the fourth member of the family, aCPSF1, endonucleolytic activity has been demonstrated as well *(Phung, Rinaldi et al. 2013)*. No enzyme capable to remove the 5′-terminal pyrophosphate has yet been discovered in archaeal organisms, which prevents us to draw clear parallels with the 5′-3′ mRNA degradation pathway in bacteria *(Clouet-d'Orval, Phung et al. 2015)*. Endonucleolytic cleavage is probably one of the initial steps that allows efficient exoribonucleolytic mRNA degradation from the released unprotected 5′ monophosphate end (Fig. 1, central panel). There might be also an unknown mechanism that directly transforms primary transcripts with protected 5′

triphosphorylated ends into secondary transcripts with 5′ monophosphorylated ends that are susceptible to archaeal 5′-3′ exoribonucleases *(Clouet-d'Orval, Phung et al. 2015)*. mRNA degradation from the 3′ end is carried out straightforwardly by a highly efficient molecular machine called the exosome complex.

1.3. Barrel-shaped RNA Degrading Complexes

Bacterial, archaeal and eukaryotic organisms all contain specialized barrel-shaped complexes that are responsible for the 3′-5′ RNA degradation *(Januszyk and Lima 2011)*: RNase PH, PNPase and the exosome complex. Evolutionary conservation of the general architecture clearly indicates its significance for the function that these ubiquitous molecular machines have (Fig. 2). Their respective subunits are evolutionarily related and they oligomerize in a ring arrangement. All these barrel-shaped assemblies feature three-fold symmetry, with the exception of eukaryotic exosome where the symmetry is disrupted. The top part of the complex is in all cases, except for the RNase PH, characterized by the exposure of RNA recognition sites which facilitate entrance of the RNA substrate through the top opening and thus provide functional distinction between the opening at the top of the complex and the one at the bottom. Entrance of the RNA into the central chamber is necessary for degradation to take place, as the active sites are confined there *(Januszyk and Lima 2011)*. The reaction is processive, which means that it is carried out in successive rounds without release of the substrate. RNA degradation in these complexes is phosphorolytic, i.e. inorganic phosphate is the attacking group in the reaction. An exception is the eukaryotic exosome complex where the major, barrel-shaped part of the assembly, does not contain ribonucleolytic activity. An additional subunit, Rrp44 (Fig. 2, in red), is hence attached to the assembly and it conducts RNA degradation hydrolytically.

1.3.1. RNase PH

The RNase PH domain, characterized by a layered βαβα fold, represents the basic building block for all RNA degrading barrel-shaped complexes *(Harlow, Kadziola et al. 2004; Januszyk and Lima 2011)* (Fig. 3). A homohexameric RNase PH complex can be found in bacteria and some archaea and it is involved in tRNA processing. In this barrel-shaped complex, RNase PH protomers are arranged as three identical

homodimers, where two individual subunits are in an inverted, head-to-tail, orientation *(Ishii, Nureki et al. 2003; Harlow, Kadziola et al. 2004)*. In Fig. 2 protomers are marked with two different colours for clarity. Each RNase PH protomer has one phosphorolytic active site and they face the interior of the barrel. Due to orientation of the protomers within the complex, three active sites are situated in the lower part of the barrel and the other three in the upper part *(Januszyk and Lima 2011)*. The molecular weight of the bacterial RNase PH is around 170 kDa *(Ishii, Nureki et al. 2003; Harlow, Kadziola et al. 2004)*.

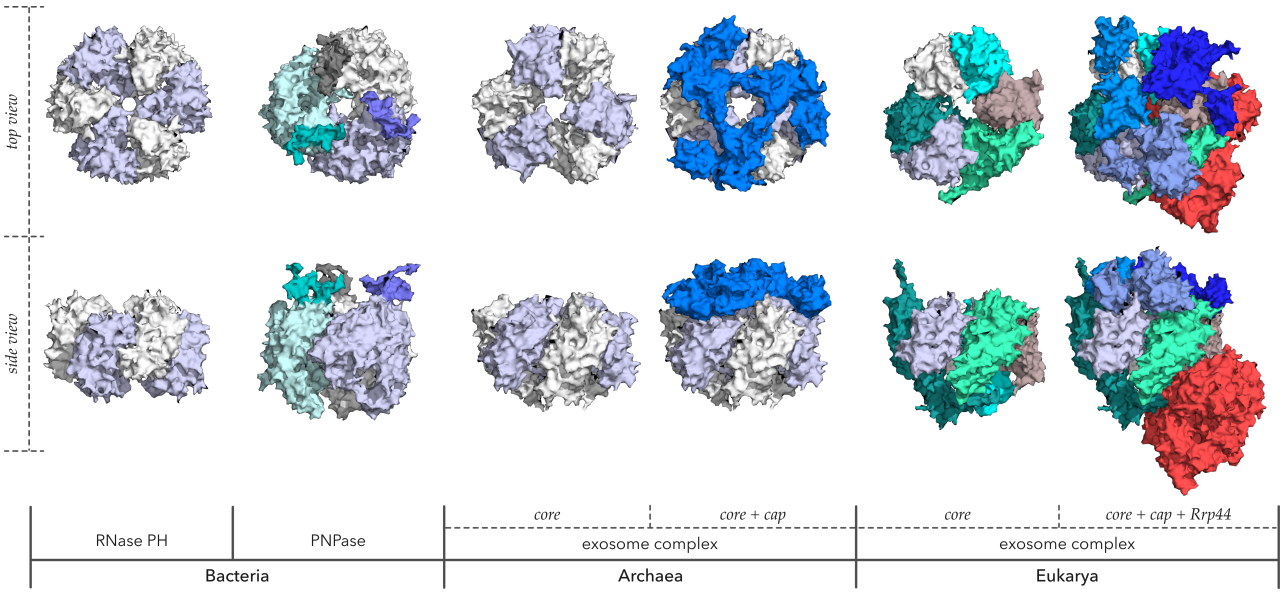


Fig. 2. Barrel-shaped RNA degrading complexes in bacterial, archaeal and eukaryotic organisms. Structures used for the comparison: *Aquifex aeolicus* RNase PH (PDB: 1UDN), *Streptomyces antibioticus* PNPase (1E3P), *Sulfolobus solfataricus* archaeal exosome complex (2JEA) and *Saccharomyces cerevisiae* eukaryotic exosome complex (4IFD).

1.3.2. PNPase

Polynucleotide phosphorylase (PNPase) is a complex that can be found in bacteria as well as in mitochondria and chloroplasts of some eukaryotic organisms. It consists of three PNPase polypeptides and each of them contains an RNase PH 1 domain, an alpha domain, an RNase PH 2 domain, a KH domain and an S1 domain *(Januszyk and Lima 2011)* (Fig. 3). In Fig. 2, each polypeptide is marked with different colour for clarity. RNA-binding domains are marked with a darker colour. The molecular weight of the assembled complex is around 250 kDa *(Stickney, Hankins et al.*

2005; Lin, Wang et al. 2012). The domains RNase PH 1 and RNase PH 2 are oriented head-to-tail, thus resembling the homodimer element of the RNase PH complex discussed previously. The active site is functional only in the RNase PH 2 domain, while the RNase PH 1 domain lost activity during evolution *(Symmons, Jones et al. 2000)*. The KH and S1 RNA binding domains are located at the top of the barrel, surrounding the opening where RNA substrate enters the chamber of the complex. Besides phosphorolytic RNA degradation, three functional active sites of the PNPase can also perform the reverse reaction and the enzyme can add an A-rich polynucleotide tail to the 3'-end of RNA, thereby targeting it for degradation *(Mohanty and Kushner 2000)*. PNPase is part of the bacterial RNA degradosome, a large assembly that connects all major enzymes involved in RNA degradation like RNase E, enolase and helicase RhIB *(Carpousis, Van Houwe et al. 1994; Bandyra, Bouvier et al. 2013)*.

1.3.3. Archaeal exosome complex

The archaeal exosome complex is structurally similar to the PNPase, as it contains all domains present in the PNPase polypeptide (Fig. 3). It consists of a hexameric barrel-shaped core and a trimeric cap tightly bound to the top of the core, with total molecular weight of around 260 kDa *(Lorentzen, Dziembowski et al. 2007)*. The hexameric core is a trimer of the Rrp41-Rrp42 heterodimers *(Lorentzen, Walter et al. 2005)*. These two proteins belong to the RNase PH family and can be considered orthologs of the two domains in PNPase: Rrp41 is an ortholog of the RNase PH 2 domain, while Rrp42 is an ortholog of the RNase PH 1 domain *(Januszyk and Lima 2011)* (Fig. 3). Accordingly, the active site is functional only in the Rrp41 subunit. The trimeric cap structure of the archaeal exosome consists of the Rrp4 and/or Csl4 subunit. They feature conserved RNA-binding domains: S1 and KH domains in the Rrp4; S1 and Zn ribbon domains in the Csl4 *(Evguenieva-Hackenberg, Walter et al. 2003)*. The presence of the RNA-binding domains around the upper opening of the complex, where RNA substrate enters into the catalytic chamber, is another parallel with the PNPase. The cap subunits contribute to more efficient RNA degradation *(Walter, Klein et al. 2006)* and they modulate substrate specificity of the enzyme *(Roppelt, Klug et al. 2010)*. The exosome complex with the Rrp4 cap displays substantial preference for the RNA substrates with poly(A) tail, whereas the complex with the Csl4 cap does not show any noticeable substrate preference *(Roppelt, Klug et al. 2010)*. Another protein, DnaG, can be part of the archaeal exosome complex *in vivo*,

but only when the complex contains Csl4 cap subunit *(Walter, Klein et al. 2006; Hou, Klug et al. 2013)*. As a poly(A)-binding protein, DnaG is able to provide the Csl4–exosome complex with a preference for poly(A) tail *(Hou, Klug et al. 2013)*.

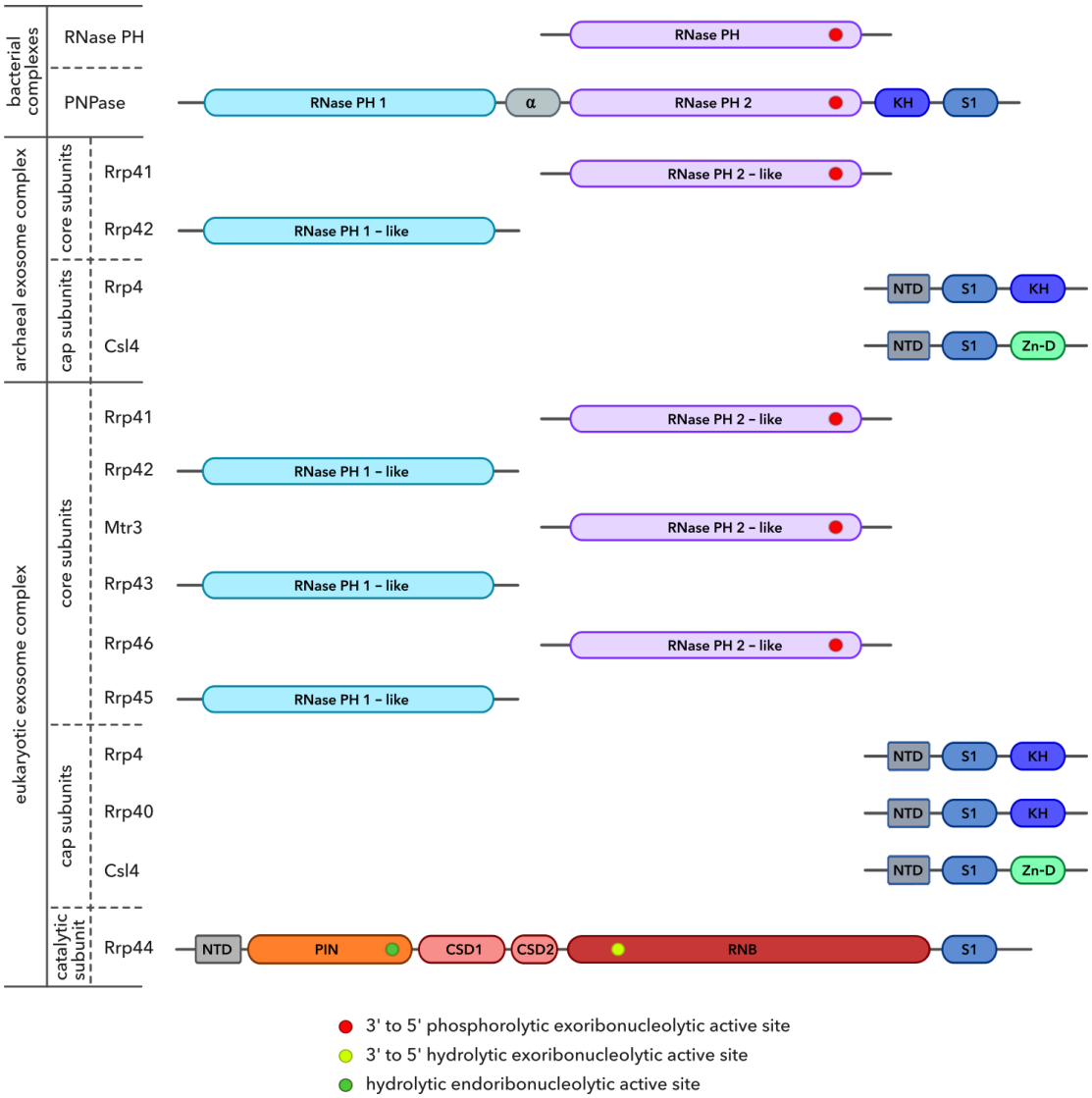


Fig. 3. Domain organisation of the proteins that constitute barrel-shaped RNA degrading complexes. Functional active sites are indicated with appropriate colour. Adapted from *Januszyk and Lima 2011* and *Makino, Baumgartner et al. 2013*.

Transient 3′-end polyadenylation or A-rich polynucleotidylation of RNA, is an important mechanism of targeting RNA molecules for degradation, observed in bacterial, archaeal and eukaryotic organisms *(Slomovic and Schuster 2011)*. Notably, for all phosphorolytic RNA degrading enzymes, including the archaeal exosome complex, the catalysed reaction is reversible, such that the enzyme uses inorganic phosphate to carry out degradation from the 3′ end of an RNA substrate, or it uses

NDPs to add polynucleotide tail to the substrate's 3′ end *(Mohanty and Kushner 2000; Lorentzen, Walter et al. 2005)*. The archaeal exosome complex is, thus, instrumental both for targeting RNA molecules for degradation and for performing degradation. According to *in vitro* studies, the direction of this reversible reaction depends on the concentrations of inorganic phosphate, NDPs and Mg^{2+} *(Evguenieva-Hackenberg, Roppelt et al. 2008)*.

The three active sites of the archaeal exosome are located in the lower part of the complex, facing the interior chamber. Within the active sites, three conserved residues have been identified as being crucial for proper positioning of the two terminal nucleotides of the substrate and the Pi: two arginines and one aspartate *(Lorentzen and Conti 2005)*. Positively charged arginine residues enable that the two phosphate groups, Pi and the one of the 3′-terminal nucleotide, are placed close to each other. The degradation reaction begins with the Pi's nucleophilic attack on the terminal phosphodiester bond, while the two previously mentioned arginine residues stabilize the transition state. The hydroxyl group of the conserved aspartate's side chain is then used as a proton donor, the phosphodiester bond between the last two nucleotides is cleaved and a ribonucleoside diphosphate is released *(Lorentzen and Conti 2005)*. This reaction, common to all phosphorolytic RNA degrading enzymes, requires presence of a divalent cation (Mg^{2+} or Mn^{2+}), that has been confirmed in the active site and that might function as an activator of the Pi *(Nurmohamed, Vaidialingam et al. 2009)*.

1.3.4. Eukaryotic exosome complex

The eukaryotic exosome shares a similar structural organisation to the archaeal complex and consists of a hexameric core and a trimeric cap structure. However, these nine subunits of the eukaryotic exosome complex are all distinct proteins *(Januszyk and Lima 2011; Makino, Baumgartner et al. 2013)*. The six core proteins belong to the RNase PH family and they can be divided into two groups (Fig. 3). In the first group are orthologs of the archaeal Rrp41 (i.e. of the RNase PH 2 domain in the PNPase): the eukaryotic Rrp41, Rrp46 and Mtr3; in the second group are orthologs of the archaeal Rrp42 (i.e. of the RNase PH 1 domain in the PNPase): the eukaryotic Rrp42, Rrp43 and Rrp45. Notably, these core subunits are arranged in three individual heterodimers that comprise one subunit of each group: Rrp41-Rrp45, Rrp46-Rrp43 and Mtr3-Rrp42 *(Januszyk and Lima 2011)*. Hence the core acquires

pseudo–three-fold symmetry. The eukaryotic cap structure always contains one copy of the three subunits: Rrp4, Rrp40 and Csl4. As in their archaeal counterparts, they feature conserved RNA binding domains: S1, KH and Zn ribbon domain (Fig. 3). This eukaryotic assembly of nine subunits is referred to as the Exo-9 complex and has a molecular weight of around 300 kDa in yeast.

All RNase PH-like subunits of the eukaryotic exosome complex are catalytically inactive, according to the studies of the yeast *(Dziembowski, Lorentzen et al. 2007)* and the human complex *(Liu, Greimann et al. 2006)*. Therefore, the Exo-9 requires an additional subunit, Rrp44 (Dis3), that carries out hydrolytic RNA degradation (Fig. 2, in red) and together they form the 400-kDa Exo-10 complex *(Dziembowski, Lorentzen et al. 2007)*. Rrp44 displays homology with bacterial exoribonucleases RNase II and RNase R *(Frazao, McVey et al. 2006)* and comprises several domains: CR3, PIN, CSD1, CSD2, RNB and S1 (Fig. 3) *(Chlebowski, Lubas et al. 2013)*. Besides hydrolytic exoribonuclease activity, conferred by the RNB domain, this subunit also has endoribonuclease active site located in the PIN domain (Schaeffer, Tsanova et al. 2009). In the nucleus, another exoribonuclease is part of the exosome complex and that is Rrp6, an ortholog of the bacterial RNase D that carries out distributive RNA degradation. Therefore, the nuclear exosome complex consists of eleven subunits with a gross molecular weight of around 500 kDa and it is referred to as Exo-11. The different composition of the exosome complex in different cellular compartments is particularly noticeable in human cells *(Tomecki, Kristiansen et al. 2010)*. Furthermore, the eukaryotic exosome complex interacts with different cofactors in different cellular compartments, which is pivotal for the elaborate regulation of the RNA substrate recruitment and decay *(Zinder and Lima 2017)*. One of the principal partners of the exosome complex in the nucleus is the TRAMP complex. This complex consists of a poly(A) polymerase Trf3 or Trf4, a Zn knuckle protein Air1 or Air2, and an RNA helicase Mtr4. The TRAMP complex is bound to the Exo-11 through the nuclear cofactor Rrp47. In the cytoplasm, the major partner of the exosome is an assembly with the Ski complex, which contains an RNA helicase Ski2, tetratricopeptide repeat scaffold protein Ski3 and two β-propeller proteins Ski8p. For the interaction between the exosome and the Ski complex, a bridging cofactor Ski7 is necessary.

2. Aims and Significance of the Project

The exosome complex is a ubiquitous molecular machine with conserved structural organization in all three domains of life. It has a crucial role in the 3′-5′ degradation of a multiplicity of RNA molecules and participates in indispensable processes like RNA turnover, quality control and processing. If its function is disrupted, serious, often fatal, consequences arise for the cell.

The archaeal exosome is an efficient enzymatic complex that processively degrades RNA molecules through a phosphorolytic reaction. Its trimeric cap structure is always part of the complex *in vivo* and, although it is not necessary for degradation *in vitro*, its presence improves the efficiency of the process. Furthermore, the cap structure influences the substrate specificity of the enzyme and its subunits have conserved RNA binding domains as their defining structural features. Since it had been shown that there is no allosteric effect at the active sites after the cap binding *(Audin, Dorn et al. 2013)*, it is reasonable to infer that the direct interaction between the cap and the RNA substrate is of central importance.

The principle aim of my PhD thesis is to shed light on the interactions between the full archaeal exosome complex and the RNA substrate. I am particularly interested in the role of the cap structure in substrate recruitment and degradation. Insights into the recruitment and processing mechanisms are the only way to fully understand how this highly efficient molecular machine works. Results from this work can also help us to grasp how other related barrel-shaped assemblies function.

Another important aim of this PhD project is to establish that newly developed approaches in NMR spectroscopy can be exploited for the studies of large biomolecular assemblies. With around 300 kDa, the archaeal exosome (*Sulfolobus solfataricus*) in complex with RNA is around 10 times the size of the complexes that are normally studied using high-resolution NMR methods. This work, thus, also provides a highly important example that shows that NMR spectroscopy is applicable to obtain detailed insights into molecular machines with molecular weights far over 100 kDa.

3. Results

3.1. Selective labelling of the Rrp4–exosome complex yields high quality NMR spectra

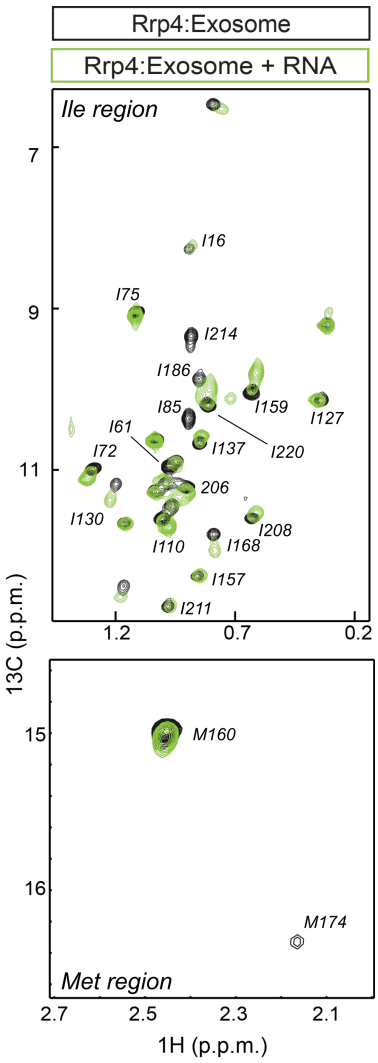
The exosome complex from the thermophile Archaea *Sulfolobus solfataricus* is particularly amenable to NMR studies due to its full three-fold symmetry and its stability at high temperatures. The trimeric cap structure of the archaeal exosome can contain Rrp4 and/or Csl4 subunits in different ratios. We turned to the Rrp4–exosome complex in this study, as Rrp4 expresses better and produces higher quality NMR spectra than the Csl4.

High quality methyl TROSY NMR spectra of the 260-kDa Rrp4–exosome complex were obtained by selectively labelling isoleucine and methionine residues in the Rrp4 cap structure with NMR active methyl groups (Ile-δ1 $[^{13}CH_3]$ and Met-ε $[^{13}CH_3]$ respectively; Fig. 4, left) Several resonances were initially assigned by "divide and conquer" approach (work performed by Stefan Schütz), while all other resonances were assigned through mutagenesis (paper 1, supplementary figure 1). In summary, we were able to assign 69% (18 in total) of all NMR active isoleucine and 100% (2 in total) of all NMR active methionine methyl group resonances. Based on these data, we were able to approximately localise the observed chemical shift perturbations that occurred as the result of RNA binding on the Rrp4 cap (Fig. 4). Importantly, the isoleucine residues that are naturally present in the Rrp4 protein are sufficiently well dispersed over its three domains that they can be used to report on the overall structural integrity of the Rrp4 cap and on the RNA binding.

3.2. Mapping the interaction interface between the Rrp4–exosome and its RNA substrate

It had been previously demonstrated that the presence of the cap structure in the exosome complex is required for the increased efficiency of RNA degradation and for modulation of specificity towards RNA molecules with different nucleotide composition. However, x-ray crystallography was unable to unravel structural details

of the interaction between the cap structure and the substrate RNA due to the lack of electron density detected for the RNA in the cap region. To address this question and map the interaction interface, we resorted to the methyl TROSY NMR spectra.



Assigned chemical shifts for Ss_Rrp4

Residue	*^{13}C (p.p.m.)*	*^{1}H (p.p.m.)*
Ile-16 δ1	8.198	0.865
Ile-61 δ1	10.948	0.959
Ile-72 δ1	11.001	1.281
Ile-75 δ1	9.021	1.099
Ile-85 δ1	10.374	0.880
Ile-110 δ1	11.618	0.992
Ile-127 δ1	10.118	0.327
Ile-130 δ1	11.178	1.185
Ile-137 δ1	10.656	0.832
Ile-157 δ1	12.307	0.840
Ile-159 δ1	10.000	0.614
Ile-168 δ1	11.816	0.776
Ile-186 δ1	9.915	0.836
Ile-206 δ1	11.207	0.886
Ile-208 δ1	11.578	0.616
Ile-211 δ1	12.710	0.967
Ile-214 δ1	9.395	0.871
Ile-220 δ1	10.173	0.807
Met-160 ε	14.989	2.453
Met-174 ε	16.328	2.165

Fig. 4. Methyl TROSY NMR spectra of the Rrp4–exosome complex with a table of assigned resonances. In the presence of RNA substrate, a number of resonances in the spectrum that correspond to the residues in the Rrp4 cap display chemical shift perturbations. The spectrum in black was recorded in the absence of RNA, the spectrum in green after the addition of the RNA.

Based on the RNA titration experiments with the Rrp4–exosome complex and observed CSPs of the assigned resonances, we established that the large area of the Rrp4 surface is involved in the interaction with RNA substrate, in particular those parts of the cap surface that belong to the S1 and KH domains. However, the number and distribution of the isoleucine and methionine residues that are naturally present on the surface of the Rrp4 cap were not sufficient for detailed mapping of its interaction with

the RNA. In order to increase the coverage of the methyl group reporters on the Rrp4–cap surface and map it with high resolution, we deployed methionine scanning (paper 1, supplementary figure 2). This NMR-based approach enables to test whether a particular residue on the protein surface, which is replaced with the reporter methionine, participates in ligand binding or not. In total, 30 residues on the Rrp4–cap surface were successively substituted with reporter methionine residues and the respective mutant Rrp4–exosome complexes were analysed by comparing their methyl TROSY NMR spectra before and after the addition of the RNA substrate (paper 1, figure 2). Methionine scanning also enables identification of hot-spot residues that are pivotal for the interaction to take place, i.e. their substitution with reporter methionine residue abolishes ligand binding. One hot-spot residue was identified through methionine scanning (K_{Rrp4}170M) and it is located in the GKNK loop of the KH domain, on the periphery of the Rrp4 cap structure (Fig. 5). This hot spot, however, obstructs RNA binding only at the periphery of the cap, without influencing interactions in the pore region. A set of RNA degradation experiments led to the identification of three more local hot spots and we confirmed their functional significance in RNA degradation process (paper 1, figure 3). Two of them are located at the periphery of the Rrp4 cap structure (R_{Rrp4}14E and K_{Rrp4}221D), while the third one is located in the neck region of the Rrp41 core subunit (R_{Rrp41}67E).

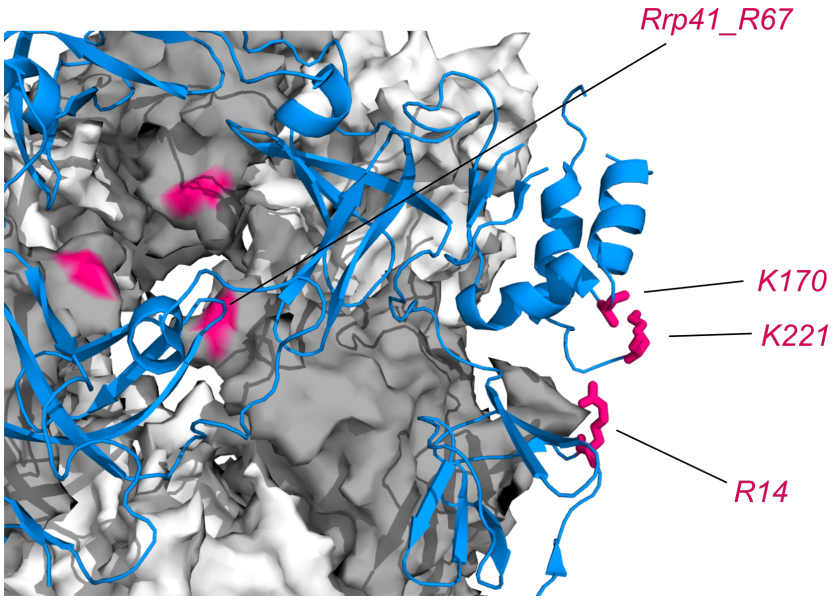


Fig. 5. Location of the identified hot–spot residues. The hexameric core is represented in surface mode, with Rrp41 in grey and Rrp42 in white; the Rrp4 cap is represented in cartoon mode and coloured blue. Residues R14, K170 and K221 outline the KH–NTD groove on the periphery of the Rrp4–cap structure, positioned between the KH domain and the N–terminal domain. Residue R67 is located in the neck region of the hexameric core.

Taken together, NMR data and RNA degradation experiments demonstrated that the substrate RNA is bound to the Rrp4–exosome complex over a 50-Å long RNA binding path (paper 1, figure 3). At the periphery of the Rrp4 cap, this binding path starts with a well-defined KH–NTD groove that includes all three peripheral hot spots (Fig. 5). Notably, our data show that one side of the NTD participates considerably in RNA binding, as part of the KH-NTD groove, although the NTD lacks any known RNA binding motif. From the periphery, the RNA binding path extends towards the S1 domain, through the pore and the neck, into the catalytic chamber of the complex.

3.3. Deconstruction and quantification of the exosome–RNA interaction

Based on known crystal structures and a plethora of NMR data that we collected, we inferred that there are four distinct regions in the Rrp4–exosome complex that interact with the substrate RNA: the peripheral cap region (KH–NTD groove) and the pore region are parts of the Rrp4–cap structure, while the neck region and the active sites belong to the hexameric core (paper 1, figure 3 and 4).

Despite considerable efforts we did not succeed to obtain information about the affinity of the exosome for the substrate RNA through the ITC (isothermal titration calorimetry) or SPR (surface plasmon resonance). Consequently, we resorted to fluorescence anisotropy measurements with fluorescently labelled RNA molecules. As NMR data from RNA titration experiments with hot-spot mutants indicated that RNA binding regions function independently from one another, we aimed to deconstruct this convoluted multivalent interaction in order to reveal the individual contributions of four contact points to the overall affinity of the exosome complex for the substrate RNA. To that end, a series of fluorescence anisotropy measurements were performed employing different versions of the exosome complex and RNA substrate, designed in such way that particular RNA–enzyme contact points were absent or hampered. For easier calculation, all affinity values (K_D) extracted from the fluorescence anisotropy measurements were converted into binding Gibbs free energies (ΔG^0), using the simple thermodynamic equation $\Delta G = - RT \ln(K_D)$. Afterwards, the local intrinsic binding energies for all four contact points (ΔG^i) were calculated by simply adding and/or subtracting ΔG^0 values for different versions of the exosome complex (e.g. $\Delta G^i_A = \Delta G^0_{ABCD} - \Delta G^0_{BCD}$, where A, B, C and D stands for functional contact points) (paper 1,

supplementary table 2). The local intrinsic binding energy can be regarded as an approximate measure of the individual contribution of a respective contact point to the overall exosome's affinity for the RNA substrate.

It is noticeable that the local intrinsic binding energy increases from the contact points on the periphery towards the active sites (paper 1, figure 4). Indeed, the overall intrinsic binding energy of the Rrp4 cap is indeed very small. This can denote the fact that interaction of the highly flexible RNA molecule with a long binding path on the cap surface involves a high entropic cost. In addition, the kink that RNA molecule needs to make in order to simultaneously occupy all four contact points, likely further increases the entropic cost. Consequently, the distal part (5′ end) of the RNA can temporarily detach from the two weak contact points on the cap surface and sample RNA binding path on any of the three Rrp4 protomers. In that process, the proximal part of the RNA (3′ end), which is inside the catalytic chamber, interacts with the neck region and the active site, and remains tightly attached to the exosome core. This finding can explain why RNA electron density was not observed anywhere on the cap surface in the crystal structure of the archaeal exosome, but only in the neck region and around the active sites *(Makino, Baumgartner et al. 2013)*. Importantly, the strongest interaction point with the RNA substrate, calculated for the active sites, ensures that the substrate efficiently ratchets forward base-by-base after every cleavage cycle.

3.4. RNA degradation is processive due to the multivalent RNA–exosome interaction

As previously mentioned, our RNA degradation assays identified an arginine residue ($R_{Rrp41}67$) at the protruding loop of the neck region as an important hot spot (paper 1, figure 3). These three positive charges (one Arg residue for each Rrp4 protomer) at the narrowest point of the RNA entrance funnel are, thus, crucial for the strong RNA–exosome interaction at the neck region. Substitution of the arginine with a residue that does not bear positively charged side chain hampered interaction with the RNA at this point and, in the case of the exosome core alone, it led to the loss of processivity (paper 2, figure 3). The effect on degradation was much stronger when the arginine was replaced with a residue that bears negatively charged side chain, which

slowed down the reaction so much that within a usual RNA degradation time setup it seemed that the reaction was completely abolished (paper 1, figure 3). However, when the Rrp4 cap was added to the exosome core with the neck mutation, processivity was recovered (paper 1, figure 3). This clearly showed that for the processivity of the RNA degradation reaction it is necessary that the exosome complex interact with the RNA at least at one more independent contact point in addition to the active site. In that way the substrate molecule would not be released from the complex between successive cleavage cycles.

3.5. Stoichiometry of the RNA–exosome interaction

The Rrp4 cap structure consists of three Rrp4 protomers, each with identical RNA interaction surface, therefore allowing simultaneous binding of three substrate RNA molecules. This was, indeed, indicated by NMR data from our RNA titration experiments (Fig. 4, left) where symmetry of the RNA–cap binding is preserved on the periphery of the cap structure. On the other hand, the neck opening is wide enough to allow passage of only one RNA molecule at a time, meaning that only one RNA molecule can simultaneously interact with all four RNA contact points of the complex and be actively degraded. This stoichiometry model was further validated by an experiment with the size-exclusion chromatography where it was possible to remove a large fraction of the RNA substrate (weakly bound only to the cap structure) from the saturated exosome complex, while a smaller fraction (tightly bound via all four contact points) remained attached (paper 1, supplementary figure 9).

As the experimental setup for the fluorescence anisotropy experiments requires considerable excess of the exosome complex over the RNA concentration, all RNA molecules were strongly bound to the enzyme via four contact points in one-to-one ratio and there was no available RNA for two additional weaker binding events to occur. To include two weaker RNA bindings to the Rrp4–cap into the affinity/energetics calculation, we resorted to the NMR data from the RNA titration experiments. In these titration experiments a substantial excess of RNA over the exosome complex was reached, rendering them applicable for the study of the weak binding events. By fitting the NMR line shapes for eight different resonances to the binding model, during the 16 titration steps, we concluded that two additional RNA

molecules can be recruited to the cap of the Rrp4–exosome complex with micromolar affinity (paper 1, figure 4c and supplementary table 2). This indicates much stronger binding compared to the contribution of the cap structure to the tight binding of the first recruited RNA molecule, which in turn provides the basis for the efficient mechanism of RNA recruitment and processing.

4. Discussion

The principal aim of this project was to unravel the mechanistic basis that underlies and enables highly efficient degradation of RNA by the complete archaeal exosome complex. We precisely determined RNA interaction path that spans over a long surface of the Rrp4 cap and that extends into the interior of the exosome core. Our results identified four independent RNA binding sites in the Rrp4–exosome complex, two on the Rrp4 surface and two in the interior of the hexameric core. The identical RNA binding path is present on each of the three cap subunits and the complex can thus recruit three RNA molecules simultaneously with micromolar affinity (ΔG^0 = *−6.50 kcal/mol*). Initial recruitment involves only two contact points on the cap surface: the KH–NTD groove on the periphery and the pore region. KH–NTD groove was shown to be very important for the rate of the RNA degradation (paper 1, figure 3a), even though its contribution to the overall Rrp4–RNA interaction is small. This can be explained by a model suggesting that peripheral region of the cap structure is involved in the first contact between the enzyme and the substrate RNA. Any disruption of this interaction would, thus, seriously reflect on the degradation process.

If the entrance to the interior of the complex is free, 3′ end of a recruited RNA molecule moves inside and binds two more contact points: the neck region and the active site. This step is driven by the considerable increase in binding free energy from −6.50 kcal/mol to −9.97 kcal/mol, which corresponds to binding with nanomolar affinity. During each cleavage cycle, the RNA molecule ratchets one base forward, owing to the binding of the newly formed 3′ end of the RNA to the active site, the highest-contributing contact point in the exosome complex (ΔG^i = *−3.1 kcal/mol*) (paper 1, figure 4b). It is important to note that while initial RNA recruitment to the Rrp4 cap corresponds to the binding free energy of −6.50 kcal/mol (micromolar affinity), after the 3′ end of the RNA goes inside the complex and engages with the

neck and the active site, the Rrp4–RNA binding free energy changes to only −1.0 kcal/mol. This is the consequence of energetically unfavourable conformation that RNA backbone needs to adopt by making a kink in order to enter the interior of the complex while still retaining interaction with the cap surface. This mechanism provides means for efficient RNA recruitment to the Rrp4 cap with micromolar affinity, without subsequently compromising translocation of the RNA molecule with too strong Rrp4–RNA interaction. While "active" RNA molecule is being degraded, two additional RNA molecules can be recruited to the remaining two RNA binding surfaces on the Rrp4–cap structure. After degradation of the first RNA molecule is finished and the entrance to the catalytic chamber is free, one of the two "waiting" RNA molecules that are positioned on the Rrp4–cap surface can instantly initiate next degradation cycle (Fig. 6).

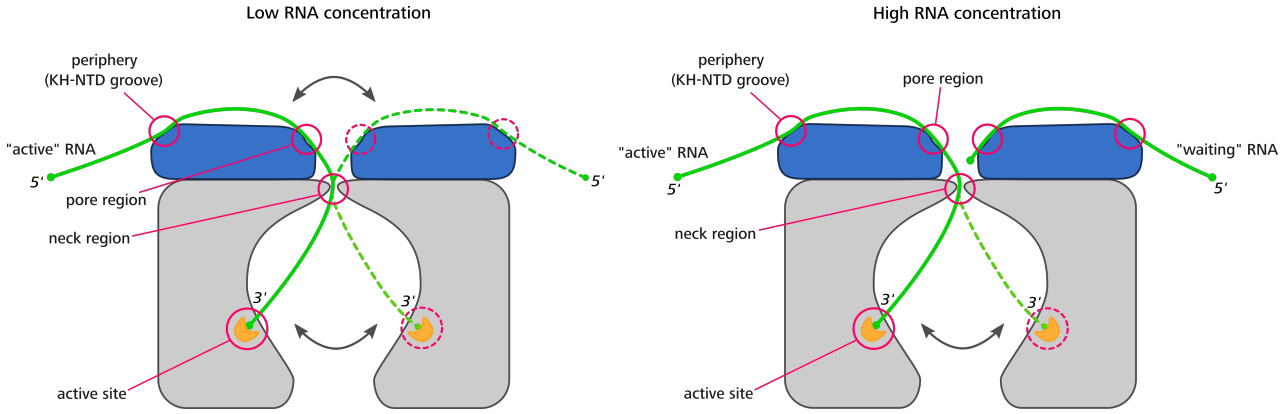


Fig. 6. Illustration summarising the interactions between the Rrp4–exosome complex and the substrate RNA. Only two active sites and two cap subunits are presented in the figure. Motions of the RNA molecule inside the catalytic chamber and on the Rrp4–cap surface are denoted by arrows. Detailed explanation in the text.

Due to the week interaction between the RNA molecule that is being actively degraded and the Rrp4–cap surface, 5′-terminal part of the RNA can temporarily detach from the Rrp4 and sample any of the three RNA binding surfaces on the Rrp4 protomers, provided that they are vacant (Fig. 6). Similarly, it was shown that the 3′-terminal part of the RNA is highly mobile inside the barrel of the exosome complex and that it jumps between three active sites (Fig. 6; paper 2, figure 4 and 5).

In this PhD project a unique mechanism was discovered that provides highly efficient functioning of a complex molecular machine. With more molecular machines to be described and investigated, a similar mechanism for facilitated substrate

recruitment and processing might be revealed, as it offers a relatively simple, yet comprehensive, framework for improving enzyme efficiency.

Our study also further extends applicability and convincingly emphasises advantages of NMR spectroscopy in the studies of large biomolecular complexes.

References

Alberts, B. (1998). "The cell as a collection of protein machines: preparing the next generation of molecular biologists." Cell 92(3): 291-294.

Audin, M. J., G. Dorn, et al. (2013). "The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding." Angew Chem Int Ed Engl 52(32): 8312-8316.

Bandyra, K. J., M. Bouvier, et al. (2013). "The social fabric of the RNA degradosome." Biochim Biophys Acta 1829(6-7): 514-522.

Carpousis, A. J., G. Van Houwe, et al. (1994). "Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation." Cell 76(5): 889-900.

Carthew, R. W. and E. J. Sontheimer (2009). "Origins and Mechanisms of miRNAs and siRNAs." Cell 136(4): 642-655.

Cech, T. R. and J. A. Steitz (2014). "The noncoding RNA revolution-trashing old rules to forge new ones." Cell 157(1): 77-94.

Chlebowski, A., M. Lubas, et al. (2013). "RNA decay machines: the exosome." Biochim Biophys Acta 1829(6-7): 552-560.

Clouet-d'Orval, B., D. K. Phung, et al. (2015). "Universal RNA-degrading enzymes in Archaea: Prevalence, activities and functions of beta-CASP ribonucleases." Biochimie 118: 278-285.

Diribarne, G. and O. Bensaude (2009). "7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor." RNA Biol 6(2): 122-128.

Dziembowski, A., E. Lorentzen, et al. (2007). "A single subunit, Dis3, is essentially responsible for yeast exosome core activity." Nat Struct Mol Biol 14(1): 15-22.

Evguenieva-Hackenberg, E. and G. Klug (2009). "RNA degradation in Archaea and Gram-negative bacteria different from Escherichia coli." Prog Mol Biol Transl Sci 85: 275-317.

Evguenieva-Hackenberg, E., V. Roppelt, et al. (2008). "Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome." Biochemistry 47(50): 13158-13168.

Evguenieva-Hackenberg, E., P. Walter, et al. (2003). "An exosome-like complex in Sulfolobus solfataricus." EMBO Rep 4(9): 889-893.

Frazao, C., C. E. McVey, et al. (2006). "Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex." Nature 443(7107): 110-114.

Harlow, L. S., A. Kadziola, et al. (2004). "Crystal structure of the phosphorolytic exoribonuclease RNase PH from Bacillus subtilis and implications for its quaternary structure and tRNA binding." Protein Sci 13(3): 668-677.

Hartwell, L. H., J. J. Hopfield, et al. (1999). "From molecular to modular cell biology." Nature 402(6761 Suppl): C47-52.

Hasenohrl, D., T. Lombo, et al. (2008). "Translation initiation factor a/eIF2(-gamma) counteracts 5' to 3' mRNA decay in the archaeon Sulfolobus solfataricus." Proc Natl Acad Sci U S A 105(6): 2146-2150.

Hou, L., G. Klug, et al. (2013). "The archaeal DnaG protein needs Csl4 for binding to the exosome and enhances its interaction with adenine-rich RNAs." RNA Biol 10(3): 415-424.

Houseley, J. and D. Tollervey (2009). "The many pathways of RNA degradation." Cell 136(4): 763-776.

Hui, M. P., P. L. Foley, et al. (2014). "Messenger RNA degradation in bacterial cells." Annu Rev Genet 48: 537-559.

Ishii, R., O. Nureki, et al. (2003). "Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus." J Biol Chem 278(34): 32397-32404.

Januszyk, K. and C. D. Lima (2011). "Structural components and architectures of RNA exosomes." Adv Exp Med Biol 702: 9-28.

Labno, A., R. Tomecki, et al. (2016). "Cytoplasmic RNA decay pathways - Enzymes and mechanisms." Biochim Biophys Acta 1863(12): 3125-3147.

Lin, C. L., Y. T. Wang, et al. (2012). "Crystal structure of human polynucleotide phosphorylase: insights into its domain function in RNA binding and degradation." Nucleic Acids Res 40(9): 4146-4157.
Liu, Q., J. C. Greimann, et al. (2006). "Reconstitution, activities, and structure of the eukaryotic RNA exosome." Cell 127(6): 1223-1237.
Lorentzen, E. and E. Conti (2005). "Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core." Mol Cell 20(3): 473-481.
Lorentzen, E., A. Dziembowski, et al. (2007). "RNA channelling by the archaeal exosome." EMBO Rep 8(5): 470-476.
Lorentzen, E., P. Walter, et al. (2005). "The archaeal exosome core is a hexameric ring structure with three catalytic subunits." Nat Struct Mol Biol 12(7): 575-581.
Makino, D. L., M. Baumgartner, et al. (2013). "Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex." Nature 495(7439): 70-75.
Marsh, J. A. and S. A. Teichmann (2015). "Structure, dynamics, assembly, and evolution of protein complexes." Annu Rev Biochem 84: 551-575.
Matera, A. G., R. M. Terns, et al. (2007). "Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs." Nat Rev Mol Cell Biol 8(3): 209-220.
Mohanty, B. K. and S. R. Kushner (2000). "Polynucleotide phosphorylase functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in Escherichia coli." Proc Natl Acad Sci U S A 97(22): 11966-11971.
Mohanty, B. K. and S. R. Kushner (2011). "Bacterial/archaeal/organellar polyadenylation." Wiley Interdiscip Rev RNA 2(2): 256-276.
Mohanty, B. K. and S. R. Kushner (2016). "Regulation of mRNA Decay in Bacteria." Annu Rev Microbiol 70: 25-44.
Neu, A., U. Neu, et al. (2015). "An excess of catalytically required motions inhibits the scavenger decapping enzyme." Nat Chem Biol 11(9): 697-704.
Nguyen, V. T., T. Kiss, et al. (2001). "7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes." Nature 414(6861): 322-325.
Nicholson, A. W. (2014). "Ribonuclease III mechanisms of double-stranded RNA cleavage." Wiley Interdiscip Rev RNA 5(1): 31-48.
Nurmohamed, S., B. Vaidialingam, et al. (2009). "Crystal structure of Escherichia coli polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly." J Mol Biol 389(1): 17-33.
Parker, R. (2012). "RNA degradation in Saccharomyces cerevisae." Genetics 191(3): 671-702.
Pereira-Leal, J. B., E. D. Levy, et al. (2006). "The origins and evolution of functional modules: lessons from protein complexes." Philos Trans R Soc Lond B Biol Sci 361(1467): 507-517.
Phung, D. K., D. Rinaldi, et al. (2013). "Archaeal beta-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor." Nucleic Acids Res 41(2): 1091-1103.
Pieters, B. J., M. B. van Eldijk, et al. (2016). "Natural supramolecular protein assemblies." Chem Soc Rev 45(1): 24-39.
Portnoy, V. and G. Schuster (2006). "RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R." Nucleic Acids Res 34(20): 5923-5931.
Roppelt, V., G. Klug, et al. (2010). "The evolutionarily conserved subunits Rrp4 and Csl4 confer different substrate specificities to the archaeal exosome." FEBS Lett 584(13): 2931-2936.
Schaeffer, D., B. Tsanova, et al. (2009). "The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities." Nat Struct Mol Biol 16(1): 56-62.
Slomovic, S. and G. Schuster (2011). "Exonucleases and endonucleases involved in polyadenylation-assisted RNA decay." Wiley Interdiscip Rev RNA 2(1): 106-123.
Stickney, L. M., J. S. Hankins, et al. (2005). "Function of the conserved S1 and KH domains in polynucleotide phosphorylase." J Bacteriol 187(21): 7214-7221.
Symmons, M. F., G. H. Jones, et al. (2000). "A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation." Structure 8(11): 1215-1226.
Tomecki, R., M. S. Kristiansen, et al. (2010). "The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L." EMBO J 29(14): 2342-2357.
Tucker, B. J. and R. R. Breaker (2005). "Riboswitches as versatile gene control elements." Curr Opin Struct Biol 15(3): 342-348.

Walter, P., F. Klein, et al. (2006). "Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus." Mol Microbiol 62(4): 1076-1089.
Zinder, J. C. and C. D. Lima (2017). "Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors." Genes Dev 31(2): 88-100.

Acknowledgements

First and foremost, I would like to express my sincere gratitude and deep appreciation to Remco for offering me the opportunity to do my PhD in the group, for his continual support, encouragement, kindness and patience. You are a true "dream boss" and above all a great, inspiring person. It was a pleasure to work with you and learn from you.

Matthias, you have been a true friend whom I can always rely on; thank you for making my transition to Tübingen an enjoyable and cheerful experience and for numerous situations when it was good to have a comforting friend around.

Maxime, thank you for introducing me to the exosome story and to the lab routine; it was always nice to have a touch of French *joie de vivre* in the group and a fellow foreigner to defy the German numerical dominance ☺.

Ancilla, you were the first person to welcome me and show me around upon my arrival in Tübingen; thank you for helping me find my way through German bureaucracy and Swabian/German phone calls countless times.

Lisa and Charlotte, you were always a great company to hang out with, both at the institute and outside of it.

Nina, Philip and Iris, I appreciate very much all your help in the lab.

Silke, thank you for all useful and insightful questions asked during our group meetings.

I would also like to mention other members of the Sprangers-Wiesner "supergroup" and thank you all for a nice collaboration, a pleasant atmosphere and kind help whenever I asked for a favour: Stefan, Mira, Fabi, Natalia, Magnus, Anna, Jan, Erik, Sussi, Lena, Karolina, Carsten, Philip (Jr.), Samira and all others who joined the group for a shorter or longer period of time.

I would like to express special gratitude to my IMPRS/MPI friends Ania, Paola and Stefan for amusing, re-energising meetings with food and drinks, for relaxing, thoughtful talks (Stefan, Paola ☺), for our thrilling banter (Ania ☺) and for being there for me.

Sarah, thank you for organising all our IMPRS events, for all deadline reminders and for being always available for any sort of help or advice.

I would like to express my gratitude to prof. Stehle and prof. Jansen for overseeing the progress of my PhD work through our TAC meetings during the last four years and also, together with prof. Schwarzer, for being part of my PhD defence committee.

I would like to use this opportunity to mention my former supervisors, Antonin Morillon, Anne-Catherine Dock-Bregeon and Olivier Bensaude, who helped me a lot to pursue my interests in science.

И на крају, желео бих да изразим огромну захвалност својој породици, мами и секи, на безрезервној подршци, бризи, охрабривању и разумевању током свих мојих успона и падова. Без вас не бих успео све ово да остварим.

Publications:

ARTICLE
PUBLISHED ONLINE: 13 MARCH 2017 | DOI: 10.1038/NCHEMBIO.2328

nature chemical biology

The Rrp4–exosome complex recruits and channels substrate RNA by a unique mechanism

Milos A Cvetkovic, Jan Philip Wurm, Maxime J Audin, Stefan Schütz & Remco Sprangers*

The exosome is a large molecular machine involved in RNA degradation and processing. Here we address how the trimeric Rrp4 cap enhances the activity of the archaeal enzyme complex. Using methyl-TROSY NMR methods we identified a 50-Å long RNA binding path on each Rrp4 protomer. We show that the Rrp4 cap can thus simultaneously recruit three substrates, one of which is degraded in the core while the others are positioned for subsequent degradation rounds. The local interaction energy between the substrate and the Rrp4–exosome increases from the periphery of the complex toward the active sites. Notably, the intrinsic interaction strength between the cap and the substrate is weakened as soon as substrates enter the catalytic barrel, which provides a means to reduce friction during substrate movements toward the active sites. Our data thus reveal a sophisticated exosome–substrate interaction mechanism that enables efficient RNA degradation.

The exosome is a large molecular machine that has an essential role in RNA processing and degradation[1–4]. The archaeal and eukaryotic exosome share a common architecture and consist of a hexameric core and a trimeric cap structure. In eukaryotes, all subunits that constitute the core and the cap are unique proteins[1]. Despite this complexity, the 9-subunit eukaryotic exosome has lost the ability to degrade RNA on its own and requires an auxiliary subunit to perform this function[5]. In archaea, the exosome structure is simpler and has a threefold symmetry[6–8] (**Fig. 1a**). Its hexameric core is composed of three Rrp41–Rrp42 dimers, and the cap structure comprises a trimer of Csl4 or Rrp4. In addition, the archaeal exosome is catalytically active, with the active sites located in the Rrp41 subunits within the interior of the hexameric core. Multiple structures of eukaryotic[9–11] and archaeal[12–16] exosome complexes have been determined in recent years, and these provide fundamental insights into the functioning of the enzyme.

The archaeal exosome structure displays an RNA entrance funnel. During catalytic degradation, the RNA is progressively threaded from the outside into the hexameric barrel, where successive diphosphate nucleotides are removed from the 3′ end of the RNA substrate in a phosphorolytic manner[13,17,18]. The cap proteins feature conserved S1, KH and Zn ribbon RNA-binding domains[6]. The S1 domains of the three cap proteins constitute the pore region, which represents the wider opening of the RNA entrance funnel. The narrowest point of this funnel marks the neck region that is formed at the top of the Rrp41–Rrp42 hexameric barrel[13–15] (**Fig. 1a**).

The Rrp4 protein, encoded side by side with the core proteins Rrp41 and Rrp42 in a highly conserved archaeal superoperon[19], is composed of an N-terminal, an S1 and a KH domain (**Fig. 1a**). The trimeric Rrp4 cap structure increases the catalytic efficiency of the enzyme complex[7,16] and provides substrate specificity for polyadenine (poly(A)) or adenine-rich stretches of RNA[8,16]. Notably, the rate at which the Rrp4–exosome complex degrades its substrate depends on the remaining length of the RNA substrate. For long RNA substrates, the degradation rate is constant from the 3′ end to nucleotide 24, then it increases between nucleotides 24 and 14, after which it rapidly drops[17,20]. These variations were observed for poly(A) RNA and are thus caused by the degradation mechanism of the enzyme and not by the RNA sequence or its structural features. The mechanism that underlies this variation in degradation rate and the way RNA interacts with the Rrp4 cap structure remain elusive.

NMR spectroscopy is a powerful tool to study biomolecular interactions[21], as it can provide quantitative information about which residues participate in the interaction. In combination with methyl group labeling[22,23] and TROSY techniques[24], the method can extract interaction information for complexes >100 kDa[25–27] and in some cases as large as 1 MDa[28]. One drawback of methyl-TROSY NMR spectroscopy is that parts of interaction interfaces can be devoid of methyl groups and would thus be 'invisible' in the experiment, limiting the binding-site mapping precision. Recently, this limitation was overcome with methionine scanning, a method in which reporter methionine residues are introduced into the protein at a location of interest[29,30]. In addition to identifying the binding interface at per-residue resolution, methionine scanning enables determination of hot-spot positions that are essential for the interaction[30].

Here we study the *Sulfolobus solfataricus* archaeal Rrp4–exosome in complex with substrate RNA (~300 kDa) using methyl-TROSY NMR spectroscopy. We identified three 50-Å long binding channels in Rrp4 that span all domains of the cap complex and we localized hot-spot positions in Rrp4 and in the exosome core. Using affinity measurements, we show that the local interaction strength between the substrate and the enzyme increases in a stepwise manner from the periphery of the cap structure toward the pore region, the neck region and the region at the active sites inside the core complex. Our data show that while one substrate molecule enters the enzyme through the neck, two additional substrate molecules can be positioned on the cap structure awaiting degradation. Notably, we also found that the binding energy between Rrp4 cap and the substrate is substantially reduced after the 3′ end of the RNA enters the catalytic barrel, thereby reducing friction during the degradation of the substrate.

RESULTS

NMR spectra of the 270-kDa complex are of high quality

The 270-kDa Rrp4–exosome complex is among the largest complexes studied using high-resolution NMR spectroscopy[31]. To simplify the NMR spectra, we labeled only the Rrp4 subunits with NMR active methyl groups (Online Methods). Using methyl-TROSY NMR techniques[24], we were able to record high-quality spectra to study the interaction between the Rrp4 cap and RNA substrate (**Fig. 1b**). The isoleucine and methionine resonances were assigned by a combination of the "divide-and-conquer"[28] and mutagenesis

Max Planck Institute for Developmental Biology, Tübingen, Germany. *e-mail: remco.sprangers@tuebingen.mpg.de

approaches[32,33] (**Supplementary Results**, **Supplementary Fig. 1**). With these methods we assigned 69% of all NMR active methyl group resonances in our spectra.

To address which residues in Rrp4 have a role in the substrate–enzyme interactions, we added three equivalents of an RNA substrate (Online Methods) to the Rrp4–exosome complex, which resulted in a large number of chemical shift perturbations (CSPs) (**Fig. 1b**). Based on the assigned methyl groups, we conclude that the Rrp4 S1 (pore region) and KH (periphery) domains are most prominently involved in RNA interactions. The number and location of residues that undergo CSPs show that the large area of the Rrp4 surface is used for the interaction with RNA substrate (**Fig. 1**).

Increasing the coverage of methyl groups on the surface

The naturally occurring methyl groups are sparsely distributed over the surface of the Rrp4 complex (**Supplementary Fig. 2**). Methyl-TROSY spectroscopy is thus 'blind' to most of the surface of the cap structure, and as a result these regions cannot provide information about intermolecular interactions. We made those areas on the surface visible using methionine scanning[30], in which we replaced residues at the surface, one at a time, with reporter methionine residues. These methionine residues then directly reported on binding events at the specific location on the surface (**Fig. 2**). One substantial advantage of methionine scanning is that the introduced methyl group appears as a novel resonance in the spectrum and can thus be assigned instantaneously. To improve the spectral quality and to prevent signal overlap in the methionine region, we mutated three natural methionine residues in the flexible and unstructured N-terminal region of Rrp4 to serine residues, which did not interfere with the structure and activity of the enzyme (**Supplementary Fig. 3a**).

In total, we selected 30 residues on the Rrp4 surface and substituted them with reporter methionine residues (**Supplementary Table 1**). Of these, 28 residues gave rise to a readily identifiable single resonance in the corresponding methyl-TROSY NMR spectrum, without interfering with the integrity of the enzyme complex (**Supplementary Fig. 3b**,**c**). To each of these samples we added a threefold molar excess of RNA to probe for binding at the site of the reporter methionine. In brief, we observed three different outcomes. First, the introduced reporter methionine was unaffected by the addition of the RNA (**Supplementary Fig. 4**), whereas the naturally occurring isoleucine residues showed the same CSPs as we observed for the wild-type (WT) complex. These positions are therefore considered to be outside the RNA binding site. Second, the reporter methionine showed a CSP upon addition of the substrate, proving that this residue is inside the RNA binding site. We found these residues in the KH domain of Rrp4 and close to the RNA entrance pore. Depending on the chemical shift differences between the free and RNA-bound state of the Rrp4–exosome, the CSPs were visible as shifting (**Fig. 2a** and **Supplementary Fig. 5**) or splitting (**Fig. 2b** and **Supplementary Fig. 6a**,**b**) methyl group resonance frequencies. It is important to note that signals that split titrate to the RNA bound state upon addition of an excess of RNA. This indicates that the exosome cap structure is symmetric in the fully RNA bound state (see below). Third, we observed a situation in which the reporter methionine resonance was unaffected by the substrate RNA, whereas the naturally occurring isoleucine residues displayed substantially reduced CSPs (**Fig. 2c**). In this case, we identified a hot-spot residue that, when mutated, interfered with substrate binding (**Fig. 2c** and **Supplementary Fig. 7a**). The identified hot spot is in the GKNK loop of the KH domain, which is at the periphery of the cap structure. This confirms previous reports that mutations in this loop of a KH domain abolish RNA binding, without influencing the structure[34]. Notably, this hot spot in Rrp4 abolished RNA interactions only at the periphery of the complex, as interactions between the residues in the pore region and RNA were unaffected

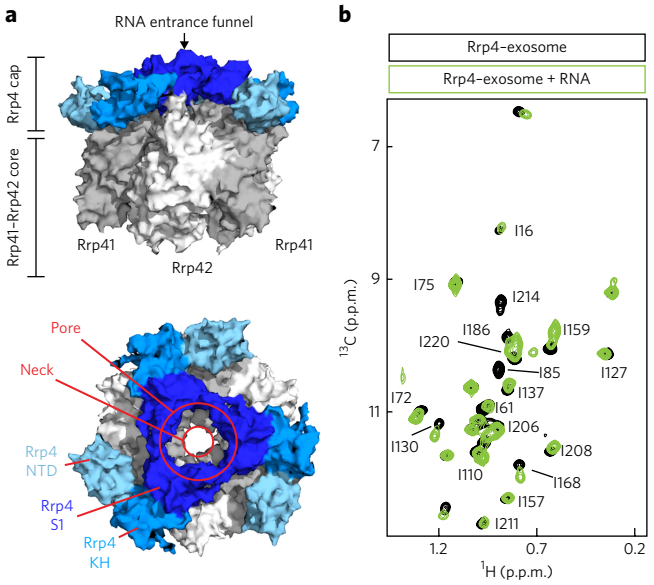


Figure 1 | Structure and NMR spectra of the Rrp4–exosome complex. (**a**) Side view (top) and top view (bottom) of the *S. solfataricus* exosome complex[15] (2JEA). Rrp4 subunits are shown in blue; Rrp41 and Rrp42 are shown in gray and white, respectively. (**b**) Isoleucine region of the methyl-TROSY NMR spectrum of the Rrp4 subunits in the full exosome complex in the absence (black) and presence (green) of RNA substrate. A number of residues showed chemical shift pertubations that report on the RNA–enzyme interaction. Assignments for some resonances are indicated (**Supplementary Fig. 1**).

(**Supplementary Fig. 7b**). The hot spot we identified thus interferes with the RNA interaction only locally and does not affect cap–RNA interactions in other areas. This shows that the S1 and KH domains in Rrp4 interact with RNA independently, which is in agreement with findings for the *Arabidopsis thaliana* Rrp4 protein[35].

Hot spots have a functional role in RNA degradation

To address the functional role of the local hot spot that we identified in the Rrp4 cap structure and to identify additional critical residues in this region, we performed RNA degradation experiments[18] (Online Methods and **Fig. 3a**). In agreement with previous reports[7,16], the presence of the Rrp4 cap proteins increased the catalytic efficiency of the exosome (**Fig. 3a**). However, when the exosome core was in complex with Rrp4 protein containing mutations in hot-spot locations (K170M, KNK170ENQ, K221D and R14E), this increase in catalytic efficiency was no longer observed (**Fig. 3a**). This indicates that the hot spot that we identified has an important role in the RNA degradation process. Using RNA degradation experiments combined with NMR binding experiments, we identified additional hot-spot residues that clustered around the interface between the KH domain and the N-terminal domain (NTD) (**Fig. 3**). Notably, all of these mutants were properly folded and interacted normally with the exosome core (**Supplementary Fig. 7b**). It is unlikely that the hot-spot mutations in Rrp4 influence the active sites in the exosome core through an allosteric mechanism, as addition of the Rrp4 cap to the exosome core does not perturb the resonances close to the active sites[25]. Rather, we postulate that the periphery of the Rrp4–exosome complex has a role in the first contacts between the enzyme and the RNA substrate. Mutations in this region are thus likely to interfere with the RNA recognition and recruitment process.

The narrowest point of the RNA funnel, the neck, is formed by the Rrp41 and Rrp42 and is located just below the Rrp4 pore (**Fig. 1a**). Mutations that invert the positive charge in this region have been

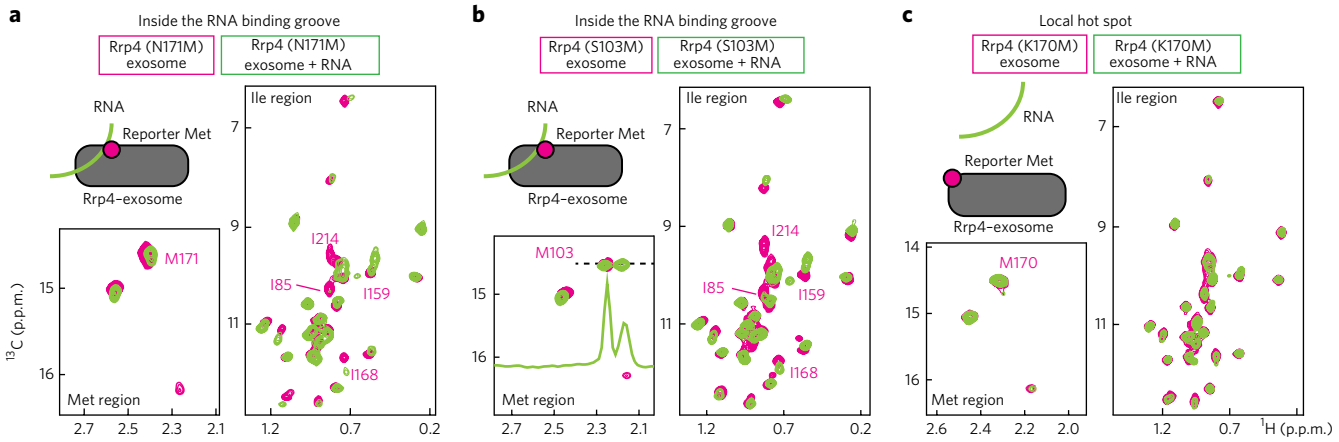


Figure 2 | Methionine scanning identifies residues in Rrp4 that interact with substrate RNA. (**a**) Spectra of the Rrp4–exosome complex that contain a single introduced reporter methionine (M171) in the absence (pink) and presence (green) of RNA substrate. The assignment of the reporter methionine is based on the appearance of a novel resonance in the methionine region of the NMR spectra (**Supplementary Fig. 4**). N171M is inside the RNA binding groove. (**b**) As in **a**, where the signal of the reporter methionine (S103M) experiences resonance splitting upon addition of one RNA molecule per Rrp4 monomer (**Supplementary Fig. 5**). (**c**) As in **a**, where the introduced reporter methionine abolishes RNA binding, on the periphery of the Rrp4 cap. I110, in the pore region, displays the same CSP as observed in the WT Rrp4 protein (**Supplementary Fig. 7**).

reported to block RNA degradation for the *Archaeoglobus fulgidus* exosome[16]. Here we show that the equivalent mutation (R67E in Rrp41) in the neck region abolishes the activity of the *S. solfataricus* core exosome (**Fig. 3a**). Notably, the activity of this mutated exosome core was rescued by addition of the Rrp4 cap to the complex (**Fig. 3a**). This indicates that the inactivity of the neck mutant is due to impairments in the interaction with RNA (see below) and not to blocking of its entrance into the catalytic chamber of the core. In the neck mutant background (Rrp41 R67E), the Rrp4 cap provides RNA binding sites that can then funnel the RNA to the active sites. These findings underscore the importance of the Rrp4 cap in the substrate RNA recognition process.

RNA interacts over a long stretch of the cap surface

Our NMR data (**Figs. 1** and **2**), together with the RNA degradation assays (**Fig. 3a**), reveal that the substrate RNA interacts with the Rrp4 surface over a long binding path (**Fig. 3b**). This path starts on the periphery of the complex around the identified hot spots and extends 50 Å through the KH–NTD groove and into the S1 pore to the neck of the core exosome. The Rrp4 cap thus funnels the substrate through the neck region toward the active sites.

To structurally validate the binding path we identified here, we compared the RNA binding sites in the NTD, S1 and KH domains with known protein–RNA structures (**Supplementary Fig. 8**). To that end, we performed DALI searches[36] with the Rrp4 domains and selected structures that contain RNA. We observed that the binding sites of the Rrp4 KH and S1 domains are compatible with known complex structures. Notably, the affected residues we identified in the Rrp4 NTD do not correspond to residues reported to be involved in RNA recognition in this domain (**Supplementary Fig. 8**) and thus represent a novel RNA binding site.

The energetics of the exosome–RNA interaction

To address the overall affinity of the exosome for the RNA substrate, we used fluorescence anisotropy measurements. For the interaction between the Rrp4–exosome complex and RNA, we extracted an affinity of 67.5 ± 22.7 nM (**Fig. 4a** and **Supplementary Table 2**, complex B), which corresponds to a binding Gibbs free energy (ΔG^0) of −9.97 ± 0.20 kcal/mol ($\Delta G = -RT\ln(K_D)$; T = 30 °C).

Known crystal structures[15], our previous NMR experiments[18] and our methionine scanning data (**Figs. 2** and **3**) indicate that there are four regions in the Rrp4–exosome complex that can interact with the RNA: the active sites and the neck region in the hexameric exosome core[15,18], in addition to the pore and periphery regions in the Rrp4 cap structure (**Fig. 3b**). To dissect the energetic contribution of each of these regions to the overall affinity between the Rrp4–exosome and substrate RNA, we used Rrp4–exosome versions in which individual RNA–enzyme contact points were abolished. The periphery interaction in Rrp4 was disrupted using a KNK170ENQ mutation (**Fig. 3a**); the cap interaction was removed through use of the isolated exosome core; and the neck interaction was disrupted using R67G[18] or R67E (**Fig. 3a**) mutation in the core protein Rrp41. In addition, we prevented interactions between the active sites in the exosome core and the substrate by using a shorter RNA that cannot reach from the neck region to the active sites. For all complexes we quantified the affinity for the RNA substrate using fluorescence anisotropy titrations (**Fig. 4a** and **Supplementary Table 2**). Differences in RNA affinity between these complexes provided insights into local intrinsic binding energies for each contact point ($\Delta G^i_A = \Delta G^0_{ABCD} - \Delta G^0_{BCD}$; where A, B, C and D are the four contact points)[37]. As an example, we can extract the Rrp4 contribution (ΔG^i_{Rrp4}) to the overall affinity between the Rrp4–exosome complex

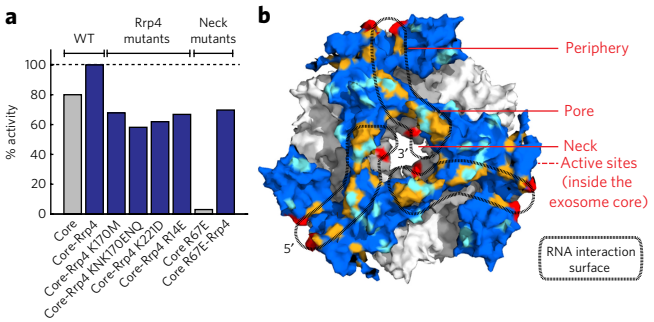


Figure 3 | Activity and RNA interaction surface of the Rrp4–exosome. (**a**) The activity of different versions of the exosome core (gray) and the Rrp4–exosome complex (blue) extracted using degradation series with 16 time points. (**b**) Visualization of the NMR binding studies and activity data. The black dashed line indicates the identified surface that is used by Rrp4 to interact with the RNA substrate. Hot-spot regions are shown in red; sites in contact with RNA are shown in orange; sites outside RNA interaction surface are shown in cyan.

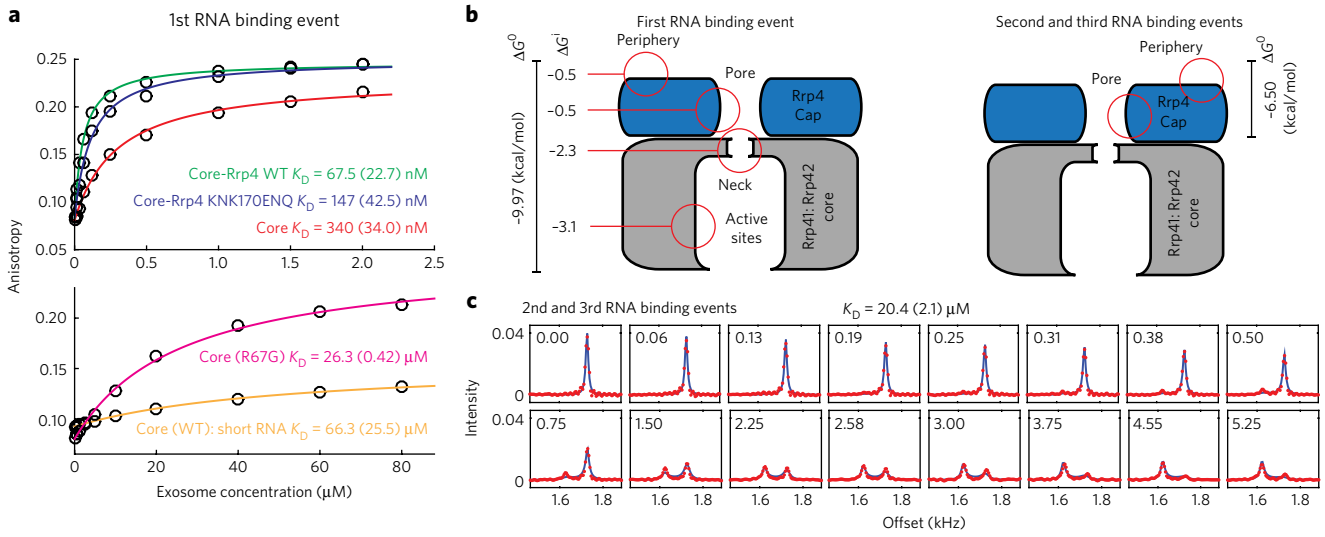


Figure 4 | Quantification of RNA–exosome interactions. (**a**) Interaction of the exosome with substrate RNA based on fluorescence anisotropy measurements. Circles represent experimental data; solid lines indicate best fit to the data (**Supplementary Table 2**). Errors (indicated in parentheses) are derived from 2–5 independent measurements (**Supplementary Table 2**). (**b**) Summary of binding data for the interaction between the Rrp4–exosome complex and the first (actively degraded) RNA (left) and between the complex and the second and third RNA substrates (right). See also **Supplementary Table 2**. (**c**) The interaction between the two additional RNA molecules and Rrp4 protein determined on the basis of fitting the NMR titration experiments to a model that describes the interaction between the RNA substrate and the Rrp4–exosome (**Supplementary Figs. 10–13**). Red dots represent experimental data; blue lines result from a global fit of eight residues that experience CSPs during the NMR titration experiment. For clarity, only the data for reporter M85 are shown; data for all fitted residues are presented in **Supplementary Figure 12**. Numbers in panels indicate the RNA/enzyme ratio and correspond to the concentration of the monomeric exosome subunits.

and the RNA by comparing the RNA affinity of the Rrp4–exosome complex ($\Delta G^0_{Rrp4-exosome}$) (**Supplementary Table 2**, complex B) with that of the isolated exosome core (ΔG^0_{Core}) (**Supplementary Table 2**, complex A). This contribution of the Rrp4 cap to the RNA binding energy can be extracted independently using different versions of the exosome core (for example, with the mutation R67G in Rrp41), with and without the cap proteins (**Supplementary Table 2**, complexes E and F). Using this approach, we were able to extract the intrinsic binding energies for all four RNA contact points in a number of independent ways (**Fig. 4b** and **Supplementary Table 3**).

In summary, our data show that the local intrinsic binding energies between the substrate and the exosome complex increase from the periphery of the complex toward the active sites (**Fig. 4b**). The fact that the highest energy contribution resides at the active sites ensures that the substrate efficiently ratchets one base farther after catalysis. Notably, the intrinsic binding energy of the cap is very small, indicating that Rrp4 provides only a limited energetic contribution to the enzyme–substrate interaction. This probably reflects the high entropic cost involved in the interaction between the highly flexible RNA substrate and the cap surface[38], which prevents a large additive effect of the cap–RNA interaction to the overall binding energy. As a result, the 5′ end of the RNA can temporarily dissociate from the Rrp4 cap structure, whereas the 3′ end remains tightly associated with the core of the exosome complex. It is important to note that the RNA–Rrp4 interactions are nevertheless functionally relevant, as mutations in the periphery of the complex influence the catalytic turnover rates (**Fig. 3a**).

The exosome recruits three RNA substrates simultaneously

Owing to the trimeric nature of the Rrp4 cap structure, three RNA species can interact simultaneously with the exosome complex. The exosome neck is, however, only large enough to accommodate a single RNA substrate that can then interact tightly with the enzyme through all four contact points. The two successive RNA molecules that can be recruited by the cap surface interact with only two Rrp4 contact points: the pore and periphery regions (**Figs. 3b** and **4b**). To experimentally validate the stoichiometry of the RNA–exosome complex, we added a large excess of substrate to the Rrp4–exosome and recorded NMR spectra of the saturated enzyme. We then removed weakly bound RNA substrates from the Rrp4 surface by size-exclusion chromatography. NMR spectra of this sample showed that a large fraction of the RNA dissociated from the complex (**Supplementary Fig. 9**). This supports the idea that one RNA molecule is bound much more strongly to the Rrp4–exosome complex than the two other substrate molecules. Our data thus reveal that the Rrp4 cap can recruit three substrates simultaneously to the enzyme complex. This is in agreement with previous findings that show that Rrp4 can enhance the substrate recruitment to the exosome[17].

In our fluorescence anisotropy measurements, the two weaker binding events remained invisible, as the exosome concentration was in excess of the RNA concentration, and all RNA molecules are recruited to the strongest exosome interaction site. To obtain insights into the interaction of the two additional RNA substrates with the Rrp4 cap, we used NMR titration experiments, where excess RNA will occupy weaker binding sites. We exploited the Rrp4 I85M methionine reporter mutant, as it is located deep inside the Rrp4 pore (**Supplementary Table 1**). During 16 titration steps, we added RNA substrate to the Rrp4–exosome complex and monitored the induced CSPs (**Figs. 1** and **4c** and **Supplementary Figs. 10–12**). To determine the RNA–Rrp4 cap affinity, we then analyzed the NMR line shapes during the titration experiments for eight different resonances. The fitting of NMR line shapes to specific binding models has proven to be an accurate method to extract kinetic parameters[39–41]. Here, we used a model for the interaction between three RNA substrates and the Rrp4–exosome that takes into account that the NMR line shapes are a superposition of the first RNA binding event (with an overall affinity in the high-nM range) (**Supplementary Table 2**) and the two subsequent weaker binding events. In addition, we included the fact that the 5′ end of

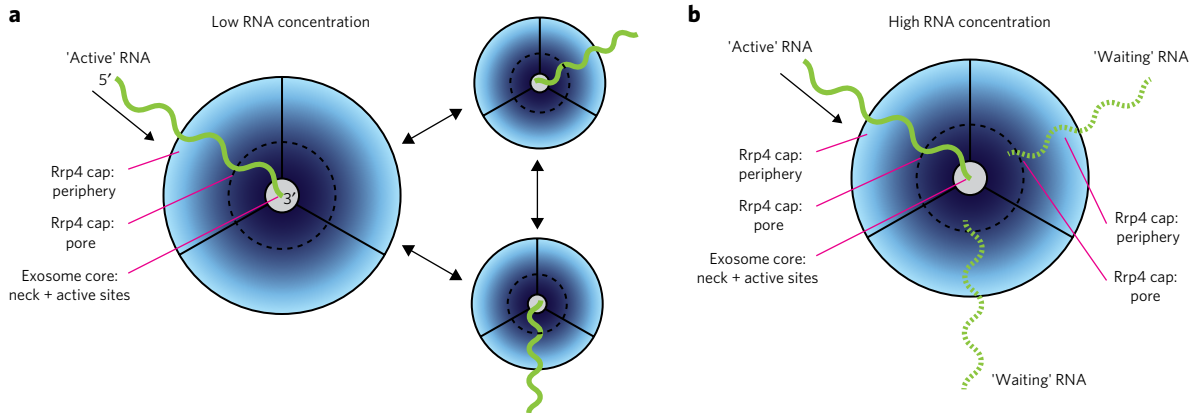


Figure 5 | Illustration of the Rrp4-modulated RNA degradation mechanism. (**a**) At low RNA concentration, a single RNA is bound to four contact points (active sites, neck, pore and periphery) in the Rrp4–exosome complex. The 5′ end is weakly bound to the Rrp4 cap structure, and this part of the substrate can thus temporarily dissociate from the cap and sample all three RNA interaction grooves. The affinity between the RNA and the exosome is indicated by a color gradient (lower interaction strength corresponds to lighter blue). (**b**) At higher RNA concentration, the Rrp4 cap can recruit two additional substrates with micromolar affinity. These RNA molecules interact with the complex through contacts with the pore and periphery regions only and are not actively degraded, as only a single RNA substrate is able to pass the neck region. The interaction energy between the cap protein and the substrate is substantially reduced when the 3′ end of the RNA moves into the catalytic core, thereby reducing molecular friction, which could compromise substrate translocation to the active sites.

the RNA that contacts the exosome core can temporarily dissociate from Rrp4 owing to the small intrinsic binding energy with the cap (**Supplementary Table 3** and **Supplementary Figs. 10** and **11**). Taking all this into consideration, we optimized the kinetic parameters to minimize the square of the difference between the experimental NMR spectra and the NMR spectra that were simulated according to the binding model (**Fig. 4c** and **Supplementary Figs. 12** and **13**).

From the global fit, we extracted that the RNA interacts with the Rrp4 cap with a K_D of 20.4 ± 2.1 μM (**Fig. 4c** and **Supplementary Table 2**, complex K), which corresponds to a free energy of binding of –6.50 ± 0.06 kcal/mol (**Fig. 4b**). From that we conclude that two additional RNAs are recruited to the cap in the Rrp4–exosome complex with micromolar affinity.

We then repeated the NMR titration experiments using the Rrp4–exosome complex, in which mutations at the periphery of the cap structure prevent RNA binding in this region. In that case, the affinity between the exosome and the second and third RNA was reduced to 60.5 ± 5.1 μM, corresponding to a free energy of –5.85 ± 0.04 kcal/mol (**Supplementary Fig. 14** and **Supplementary Table 2**, complex L). This confirms that the intrinsic binding energy of the periphery region is small (**Supplementary Table 3**) despite its functional significance (**Fig. 3a**).

DISCUSSION

Static high-resolution structures are indispensible for understanding the mechanism of molecular machines. These structures, however, hide dynamic features that are important for biological function. Here, we address how the Rrp4–exosome complex recruits substrate RNA molecules and how the enzyme is able to channel these toward the active sites.

Using methyl-TROSY NMR spectroscopy and degradation assays, we identified a 50-Å long RNA interaction channel that covers a large portion of the Rrp4 surface between the periphery and the RNA entrance pore (**Figs. 3** and **5**). This path is in agreement with a very low-resolution SAXS reconstruction of the *A. fulgidus* Rrp4–exosome in complex with RNA[17], where the RNA appears to make contacts with the top of the Rrp4 cap. At the same time, the interactions that we observed on the archaeal cap structure differ from the RNA interactions that have been reported for the eukaryotic exosome cap proteins (**Supplementary Fig. 15**). Whether these differences are due to variations in the substrate recruitment mechanism, the RNA used in the experiment or the complex composition (for example, the presence of Rrp6 in the eukaryotic exosome)[10,11,42] remains to be determined.

Although the large RNA interaction surface on the Rrp4 protein in the archaeal exosome is advantageous for substrate recruitment (**Fig. 5**), it can potentially compromise catalysis, as excessive friction would hamper motions of the RNA toward the active sites[17]. Optimal catalytic efficiency is thus a trade-off between efficient substrate recruitment and rapid substrate translocation. Our binding experiments show that Rrp4 can recruit RNA substrates with micromolar affinity. This RNA is initially in contact with both the periphery and the pore of the Rrp4 cap structure. When the entrance pore is free, the 3′ end of the substrate can move into the core such that it contacts all four interaction sites (periphery, pore, neck and active sites) (**Figs. 4b** and **5**). This step will be driven by a significant change in binding free energy from –6.50 kcal/mol (micromolar affinity) to –9.97 kcal/mol (nanomolar affinity) (**Supplementary Table 2** and **Fig. 4b**).

After the substrate is fully bound to the enzyme, the 3′ terminal nucleotide can be phosphorolyzed at the active sites. This will result in a release of a nucleotide diphosphate product and a loss of the interactions between the substrate and the active sites. The RNA then ratchets one base farther such that the new 3′ end can engage in interactions with the active sites. This movement results in a favorable change of the binding free energy between the enzyme and the substrate of up to –3.15 kcal/mol (**Fig. 4** and **Supplementary Table 2**). This change in free energy is much smaller than was previously calculated for the eukaryotic complex[43], where RNA degradation takes place in the additional exosome component Rrp44. This difference correlates well with the fact that the eukaryotic exosome, unlike the archaeal exosome, is able to process RNA substrates with secondary structure elements that need to unfold before entering the exosome barrel.

The combination of the four independent interaction points can explain the molecular basis for the dependence of the degradation velocity on substrate length[17,20]. Substrates longer than 24 nucleotides will interact with the enzyme through all sites (periphery, pore, neck and active sites). For substrates shorter than 24 nucleotides, the RNA is too short to reach from the active site to the periphery and the friction between the RNA, and the

Rrp4 cap is reduced, which results in an increase in degradation velocity[17]. For substrates shorter than 13 nucleotides, the RNA is too short to interact simultaneously with the active sites and the neck region. This releases the RNA from the pivotal point in the neck of the exosome complex[18] and results in a decrease of the degradation rate, as the substrate is no longer tightly restrained to the barrel of the complex[17].

It is important to note that the ΔG^i values appear more negative when the two binding sites influence each other in a constructive manner and appear less negative when the two binding events influence each other negatively (**Supplementary Fig. 16**). Notably, our data indicate a situation where the sum of the four individual intrinsic binding energies (ΔG^i) is smaller than the overall binding energy ($\Delta G^0_{Rrp4-exosome}$). Assuming that there are no additional interaction sites between the substrate and the enzyme, this highlights two important features of the RNA–enzyme interaction. First, there is no additivity in the multivalent RNA–exosome binding processes[37,44]. The lack of additive binding effects has also been observed, for example, for the interaction between the tau protein and microtubules[45] and for the interaction between the trigger factor chaperone and the unfolded alkaline phosphatase substrate[46]. In those examples, and in the case of the flexible RNA we study here, the interaction with one binding site is unable to position other motifs in the proper binding position and at each interaction site a large entropic cost has to be paid. Second, the small sum of the intrinsic binding energies suggests that the RNA backbone adopts an energetically unfavorable conformation upon interaction with the enzyme, which is plausible, as the RNA must make a tight turn upon entering the barrel of the Rrp4–exosome complex. Mechanistically this has important advantages for the degradation process. In particular, the initial recruitment of the substrate by the Rrp4 cap involves a binding energy of ΔG = –6.50 kcal/mol (**Supplementary Table 2**). This energy would invoke substantial friction between the substrate and the enzyme such that degradation rates would be substantially reduced. However, upon translocation of the 3′ end of the RNA substrate into the exosome core (which is driven by an increase in the binding free energy of –3.47 kcal/mol; **Supplementary Table 2**) the Rrp4–RNA binding energy is substantially reduced to an intrinsic binding energy of around –1 kcal/mol (**Supplementary Table 3**). The binding groove that recruits substrates does therefore not add substantially to the enzyme–RNA interaction during degradation. The Rrp4–exosome complex can thus combine efficient substrate recruitment without compromising the movement of the substrate toward the active sites (**Fig. 5**).

In summary, we reveal here a unique mechanism by which the exosome ensures efficient recruitment and motion of the RNA substrate. Our data thus enhance understanding of the exosome complex and reveal functionally important molecular details that are hidden in static crystal structures. It will be interesting to see whether the mechanisms we identified here are general principles that are also exploited by other complex molecular machines.

Received 19 June 2016; accepted 20 December 2016;
published online 13 March 2017

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

References

1. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′→5′ exoribonucleases. *Cell* **91**, 457–466 (1997).
2. Bousquet-Antonelli, C., Presutti, C. & Tollervey, D. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**, 765–775 (2000).
3. Mitchell, P., Petfalski, E. & Tollervey, D. The 3′ end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev.* **10**, 502–513 (1996).
4. van Hoof, A., Frischmeyer, P.A., Dietz, H.C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262–2264 (2002).
5. Dziembowski, A., Lorentzen, E., Conti, E. & Séraphin, B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat. Struct. Mol. Biol.* **14**, 15–22 (2007).
6. Evguenieva-Hackenberg, E., Walter, P., Hochleitner, E., Lottspeich, F. & Klug, G. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep.* **4**, 889–893 (2003).
7. Walter, P. *et al.* Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* **62**, 1076–1089 (2006).
8. Roppelt, V., Klug, G. & Evguenieva-Hackenberg, E. The evolutionarily conserved subunits Rrp4 and Csl4 confer different substrate specificities to the archaeal exosome. *FEBS Lett.* **584**, 2931–2936 (2010).
9. Wasmuth, E.V., Januszyk, K. & Lima, C.D. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **511**, 435–439 (2014).
10. Liu, Q., Greimann, J.C. & Lima, C.D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223–1237 (2006).
11. Makino, D.L., Baumgärtner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495**, 70–75 (2013).
12. Lorentzen, E. *et al.* The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.* **12**, 575–581 (2005).
13. Navarro, M.V., Oliveira, C.C., Zanchin, N.I. & Guimarães, B.G. Insights into the mechanism of progressive RNA degradation by the archaeal exosome. *J. Biol. Chem.* **283**, 14120–14131 (2008).
14. Lorentzen, E. & Conti, E. Structural basis of 3′ end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. *Mol. Cell* **20**, 473–481 (2005).
15. Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B. & Conti, E. RNA channelling by the archaeal exosome. *EMBO Rep.* **8**, 470–476 (2007).
16. Büttner, K., Wenig, K. & Hopfner, K.P. Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol. Cell* **20**, 461–471 (2005).
17. Hartung, S., Niederberger, T., Hartung, M., Tresch, A. & Hopfner, K.P. Quantitative analysis of processive RNA degradation by the archaeal RNA exosome. *Nucleic Acids Res.* **38**, 5166–5176 (2010).
18. Audin, M.J., Wurm, J.P., Cvetkovic, M.A. & Sprangers, R. The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation. *Nucleic Acids Res.* **44**, 2962–2973 (2016).
19. Koonin, E.V., Wolf, Y.I. & Aravind, L. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11**, 240–252 (2001).
20. Niederberger, T., Hartung, S., Hopfner, K.P. & Tresch, A. Processive RNA decay by the exosome: merits of a quantitative Bayesian sampling approach. *RNA Biol.* **8**, 55–60 (2011).
21. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.* **35**, 60–67 (2015).
22. Kerfah, R., Plevin, M.J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* **32**, 113–122 (2015).
23. Gardner, K.H. & Kay, L.E. Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119**, 7599–7600 (1997).
24. Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E. & Kay, L.E. Cross-correlated relaxation enhanced 1H–13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.* **125**, 10420–10428 (2003).
25. 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).
26. Gelis, I. *et al.* Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131**, 756–769 (2007).
27. Rosenzweig, R., Moradi, S., Zarrine-Afsar, A., Glover, J.R. & Kay, L.E. Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. *Science* **339**, 1080–1083 (2013).
28. Sprangers, R. & Kay, L.E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).
29. Mari, S. *et al.* Structural and functional framework for the autoinhibition of Nedd4-family ubiquitin ligases. *Structure* **22**, 1639–1649 (2014).
30. Stoffregen, M.C., Schwer, M.M., Renschler, F.A. & Wiesner, S. Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. *Structure* **20**, 573–581 (2012).

31. Rosenzweig, R. & Kay, L.E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* **83**, 291–315 (2014).
32. 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).
33. Amero, C. *et al.* A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR* **50**, 229–236 (2011).
34. Hollingworth, D. *et al.* KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res.* **40**, 6873–6886 (2012).
35. Chekanova, J.A., Dutko, J.A., Mian, I.S. & Belostotsky, D.A. *Arabidopsis thaliana* exosome subunit AtRrp4p is a hydrolytic 3→5′ exonuclease containing S1 and KH RNA-binding domains. *Nucleic Acids Res.* **30**, 695–700 (2002).
36. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
37. Jencks, W.P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. USA* **78**, 4046–4050 (1981).
38. Searle, M.S. & Williams, D.H. On the stability of nucleic acid structures in solution: enthalpy-entropy compensations, internal rotations and reversibility. *Nucleic Acids Res.* **21**, 2051–2056 (1993).
39. Waudby, C.A., Ramos, A., Cabrita, L.D. & Christodoulou, J. Two-dimensional NMR lineshape analysis. *Sci. Rep.* **6**, 24826 (2016).
40. Kovrigin, E.L. NMR line shapes and multi-state binding equilibria. *J. Biomol. NMR* **53**, 257–270 (2012).
41. Bain, A.D., Rex, D.M. & Smith, R.N. Fitting dynamic NMR lineshapes. *Magn. Reson. Chem.* **39**, 122–126 (2001).
42. Makino, D.L. *et al.* RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* **524**, 54–58 (2015).
43. Vuković, L., Chipot, C., Makino, D.L., Conti, E. & Schulten, K. Molecular mechanism of processive 3′ to 5′ RNA translocation in the activesSubunit of the RNA exosome complex. *J. Am. Chem. Soc.* **138**, 4069–4078 (2016).
44. Zhou, H.X. & Gilson, M.K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* **109**, 4092–4107 (2009).
45. Butner, K.A. & Kirschner, M.W. Tau protein binds to microtubules through a flexible array of distributed weak sites. *J. Cell Biol.* **115**, 717–730 (1991).
46. Saio, T., Guan, X., Rossi, P., Economou, A. & Kalodimos, C.G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **344**, 1250494 (2014).

Acknowledgments

We acknowledge all members of R.S.'s laboratory for discussions. We thank S. Wiesner for suggestions regarding the methionine scanning experiments, I. Holdermann and J. Petters for excellent technical assistance and V. Truffault for maintenance of the NMR infrastructure. M.A.C. and S.S. acknowledge funding from the International Max Planck Research School "From Molecules to Organisms". This work was supported by the Max Planck Society and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013), ERC grant agreement 616052 (R.S.).

Author contributions

M.A.C. performed and analyzed NMR, binding and degradation experiments. J.P.W. assisted with the binding experiments, M.J.A. assisted with the degradation experiments and S.S. performed NMR assignment experiments. R.S. conceived the project, analyzed data and wrote the manuscript. All authors commented on the data, the analysis and the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are 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 production. The *S. solfataricus* Rrp41–Rrp42 exosome core was obtained by co-expression of the two proteins in LB medium using BL21 (DE3) RIL *E. coli* cells (Stratagene). The core complex was purified using Ni-affinity chromatography and gel filtration as previously described[25]. The *S. solfataricus* Rrp4–exosome cap protein was obtained by overexpression in D_2O-based minimal medium in the presence of $^{2}H^{12}C$ glucose. Methionine (^{1}H-^{13}C; 100 mg/l) and α-ketobuteric acid (4-$^{1}H_3$-^{13}C, 3-$^{12}C^{2}H_2$; 60 mg/l) were added to the growth medium 1 h before the induction of protein expression to ensure that the methyl groups in methionine and isoleucine (δ1) residues were NMR active. Cells were lysed in buffer A (50 mM $NaPO_4$ (pH 7.5), 150 mM NaCl, 10 mM Imidazol, 1 mM DTT), and Rrp4 was bound to Ni-NTA resin. The resin was washed extensively using buffer A. Subsequently, the exosome core complex was added to the Ni-resin to allow for the formation of the Rrp4–exosome complex during 2 h at 4 °C with constant tilting. In this way we ensure an equimolar composition of the complex (**Supplementary Fig. 17**). The reconstituted complex was eluted using buffer A plus 330 mM imidazole. The complex was dialyzed into buffer A without imidazole, and the affinity tag was simultaneously removed using TEV protease. Subsequently the complex was incubated at 50 °C for 2 h and purified to homogeneity using gel filtration in GF buffer (30 mM KPO_4 (pH 7), 100 mM NaCl) (**Supplementary Fig. 17**). It is worth mentioning that the Rrp4 protein could be purified in isolation at low concentration but that it was not possible to record high-quality NMR spectra on the isolated protein.

RNA production. The RNA substrate for NMR experiments was prepared using *in vitro* transcription. The RNA substrate contained a 5′ GC-based hairpin structure followed by 32 adenines (5′-GCCCCCCCCGAAAGGGGGGG GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-2′-3′ cyclic phosphate). We used a poly(A) sequence, as this has been shown to be the preferred substrate for the Rrp4–exosome. According to crystal structures of the Rrp4–exosome complex, the adenine tail is substantially longer than the distance from the Rrp4 periphery to the active sites of the exosome. A homogenous 3′ end of the transcript was ensured by HDV ribozyme cleavage. This resulted in a 2′-3′ cyclic phosphate that also prevented the degradation of the RNA substrate by the exosome complex. The 5′ hairpin structure prevents potential binding of the single stranded substrate RNA in the reverse direction. For degradation experiments, the same RNA was prepared using run-off transcription that results in an RNA that contains a 3′ hydroxyl group, which is an ideal substrate for the exosome complex. This RNA contained three extra bases at the 3′ end that resulted from the linearization of the DNA template (5′-GCCCCC CCCGAAAGGGGGGGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAGCU-3′). The RNAs that were used for binding experiments contained a single 4-thiouridine (4-S-U) (long RNA, 5′-GCCCCCCCCGAAAGGGGG GGGAAAAAAAAAAAAAAAAAAAAA-4-S-U-AAAAAAAAAAAGCU-3′, short RNA, 5′-4-S-U-GCCCCCCCCGAAAGGGGGGGAAAA) and were obtained from Dharmacon.

NMR spectroscopy. NMR spectra were recorded at 50 °C on an AVIII-800 spectrometer with room temperature probe-heads. HMQC methyl-TROSY spectra were recorded with a carbon acquisition time of 45 ms. Spectra were processed using the NMRPipe/NMRDraw software suite[47], using zero-filling to 2k (1k) points in the direct (indirect) dimension to increase digital resolution. For NMR titration experiments, the substrate RNA was added to a 15.75-fold molar excess (RNA concentration over the concentration of the full exosome complex) in 15 steps. For methionine scanning experiments, RNA was added to an approximately 3:1 ratio. Small differences in the RNA/protein ratio in the methionine scanning spectra result in differences in the saturation of the Rrp4–exosome with RNA. This does not influence the interpretation of the methionine scanning data, that determine only whether a residue is outside or inside the RNA interaction groove or whether the residue interferes with RNA binding.

NMR methyl groups of isoleucine residues were assigned using the 'divide-and-conquer' approach, in which parts of the large complex are assigned in isolation, then the assignments are transferred to the intact complex. The 28-kDa full-length Rrp4 protein is not stable without the exosome core complex, but we found that a truncated form of Rrp4 that contains only the S1 and KH domains can be purified. Methyl group assignments of this monomeric 21-kDa Rrp4 fragment were obtained using traditional TROSY-based methods[48] and could be partially transferred to the 270-kDa Rrp4–exosome complex (**Supplementary Fig. 1a**). The methyl group assignments obtained in this step were complemented with a number of assignment mutants[32,33] (**Supplementary Table 1**). In that case, single isoleucine or methionine methyl groups were replaced with an alternative amino acid, which ideally results in the disappearance of a single resonance from the methyl-TROSY NMR spectrum (**Supplementary Fig. 1b**).

To extract the binding constant for the interaction between the RNA and the exosome cap, one-dimensional (1D) traces were extracted from the 2D spectra using the nmrPipe/nmrDraw software suite[47]. These 1D spectra were fitted using numerical equations for NMR resonance lines taking into account the model that is described in detail in **Supplementary Figures 10**–**13**. Errors in the determined parameters were obtained through a Jackknife approach, where single residues were omitted from the fitting procedure. Details of the fitting procedure and of the used model are described in **Supplementary Figures 10** and **11**.

Fluorescence anisotropy. For fluorescence anisotropy measurements, the substrate RNA containing a 4-thiouridine was coupled to 6-(iodoacetamido)-fluorescein[49]. RNA (10 nM) was mixed with increasing amounts of several variants of the exosome (0 to 2,000 nM or 0 to 80 μM full exosome complex, depending on the affinity) in 96-well plates. After 2h incubation changes in fluorescence anisotropy were detected using a plate reader (Tecan, Infinite F200; filter linear polarization XP38; excitation at 485 nm and emission at 535 nm). Binding curves were fitted to the standard equation for a one-site binding model[50] using in-house scripts. Errors in the measurements were extracted from fully independent measurements as indicated in the legend of **Supplementary Table 2.**

Degradation assay, HPLC. RNA degradation experiments were performed by mixing RNA substrate (25 μM) with different versions of the exosome complex (60 nM exosome) in 180 μl reaction buffer (20 mM HEPES, pH 6.5, 60 mM KCl, 0.1 mM EDTA, 2 mM DTT, 8 mM $MgCl_2$, 10 mM Na_2HPO_4) at 50 °C. 10 μl aliquots of the reaction mix were taken at 16 different defined time points, and the reaction was quenched by mixing the aliquots 1:1 with 8 M Urea, 20 mM EDTA, 2 mM Tris, pH 8. The amounts of substrate and product (a 5′ GC-based hairpin structure followed by 10 adenines) were quantified on a DNAPac PA100 column (Dionex) using a linear gradient from buffer A (5 M Urea, 20 mM Tris, pH 8, 100 mM NaCl) to buffer B (5 M Urea, 20 mM Tris, pH 8, 2 M NaCl). Peak intensities were translated into concentrations from which the turnover numbers were extracted by linear fitting of the 16 time points[18].

Data availability. All data generated or analyzed during this study are included in the published article (and its supplementary information files) or are available from the corresponding author upon reasonable request.

47. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
48. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 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 (1997).
49. Ramos, A. & Varani, G. A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).
50. Johnson, P.E., Tomme, P., Joshi, M.D. & McIntosh, L.P. Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC 2. NMR and ultraviolet absorption spectroscopy. *Biochemistry* **35**, 13895–13906 (1996).

Supplementary information

The Rrp4-exosome complex recruits and channels substrate RNA by a unique mechanism.

Milos A. Cvetkovic[1], Jan Philip Wurm[1], Maxime J. Audin[1], Stefan Schütz[1] & Remco Sprangers[1*]

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

[*] e-mail: remco.sprangers@tuebingen.mpg.de

Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Results

Supplementary Table 1 | List of used proteins.

Methionine scanning and NMR		
Rrp4_3MS_I130M [a]	Assignment	*1306*[b]
Rrp4_3MS_I186M	Assignment	*1307*
Rrp4_3MS_I194M	Assignment	*1308*
Rrp4_3MS_M160I	Assignment	*1310*
Rrp4_3MS_I220M	Assignment	*1311*
Rrp4_3MS_I220L	Assignment	*1312*
Rrp4_3MS_I16M	Assignment	*1315*
Rrp4_3MS_I85M	NMR titration (Fig. 4c, d)	*998*
Rrp4_3MS_I85M_KNK170EMQ	NMR titration (Supplementary Fig. 9)	*1309*
Rrp4_3MS_K38M	outside	*997*
Rrp4_3MS_K54M	outside	*1015*
Rrp4_3MS_I72M	outside	*1114*
Rrp4_3MS_N73M	outside	*1294*
Rrp4_V83M	outside	*966*
Rrp4_S109M	outside	*963*
Rrp4_3MS_S109M	outside	*1104*
Rrp4_3MS_N111M	outside	*1116*
Rrp4_3MS_R118M	outside	*1003*
Rrp4_3MS_K144M	outside	*1117*
Rrp4_3MS_K146M	outside	*1110*
Rrp4_3MS_N153M	outside	*1106*
Rrp4_3MS_I61M	inside	*1303*
Rrp4_3MS_D74M	inside	*1298*
Rrp4_3MS_L106M	Inside	*1524*
Rrp4_3MS_G107M	inside	*1115*
Rrp4_3MS_R108M	inside	*1000*
Rrp4_3MS_I137M	inside	*1005*
Rrp4_3MS_N171M	inside	*1119*
Rrp4_3MS_N190M	inside	*1107*
Rrp4_3MS_I206M	inside	*1007*
Rrp4_3MS_I85M	splitting	*998*
Rrp4_3MS_K97M	splitting	*1105*
Rrp4_3MS_S103M	splitting	*999*
Rrp4_3MS_I110M	splitting	*1001*
Rrp4_3MS_V112M	splitting	*1002*
Rrp4_3MS_V140M	splitting	*1109*
Rrp4_3MS_K170M	local hot-spot	*1118*
Rrp4_3MS_KNK170ENQ	local hot-spot, tested protein integrity	*1297*
Rrp4_3MS_K221D	local hot-spot, tested protein integrity	*1313*
Rrp4_3MS_R14E	local hot-spot, tested protein integrity	*1314*
Rrp4_3MS	Supplementary Fig. 3a	*986*
Rrp4_3MS_K71M	Supplementary Fig. 3b; excluded from Met scanning; structure disrupted;	*1108*

Nature Chemical Biology: doi:10.1038/nchembio.2328

Rrp4_3MS_S203M	Supplementary Fig. 3c; excluded from Met scanning; partial disulfide bond	*1006*
Rrp41 Rrp42	Exosome core	*545*

Degradation experiments		
Rrp41 Rrp42	Exosome core	*545*
Rrp4_3MS_K170M	local hot-spot; degradation rate decreased	*1118*
Rrp4_3MS_KNK170ENQ	local hot-spot; degradation rate decreased	*1297*
Rrp4_3MS_K221D	local hot-spot; degradation rate decreased	*1313*
Rrp4_3MS_R14E	local hot-spot; degradation rate decreased	*1314*
Rrp41 (R67E) Rrp42	Exosome core hot-spot; degradation abolished	*1316*

Binding experiments		
Rrp41 (D182A) Rrp42	Inactive exosome core	*1113*
Rrp41 (R67G, D182A) Rrp42	Inactive exosome core hotspot mutant	*1414*
Rrp41 (R67E, D182A) Rrp42	Inactive exosome core hotspot mutant	*1525*
Rrp4_3MS	WT Rrp4	*986*
Rrp4_3MS_KNK170ENQ	Rrp4 periphery mutant	*1297*

a) 3MS refers to the Rrp4 proteins where the 3 N-terminal methionine residues are replaced with serine residues (see Supplementary Fig. 3).
b) Internal reference.

Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Table 2: Energetics of RNA binding

	Complex [a]	K_D	Δ G (kcal/mol) [b]	Active Sites [c]	Neck [c]	Pore [c]	Periphery [c]
First RNA binding event							
A	Core WT	340 ± 34.0 nM [d]	-8.97 ± 0.06	✓	✓	-	-
B	Core WT Rrp4 WT	67.5 ± 22.7 nM	-9.97 ± 0.20	✓	✓	✓	✓
C	Core WT Rrp4 KNK170ENQ	147 ± 42.5 nM	-9.49 ± 0.17	✓	✓	✓	-
D	Core WT + Short RNA	66.34 ± 25.5 μM	-5.82 ± 0.20	-	✓	-	-
E	Core R67G (Rrp41)	26.3 ± 0.42 μM	-6.35 ± 0.01	✓	-	-	-
F	Core R67G (Rrp41) Rrp4 WT	4.67 ± 1.24 μM	-7.40 ± 0.15	✓	-	✓	✓
G	Core R67G (Rrp41) Rrp4 KNK170ENQ	7.14 ± 3.35 μM	-7.18 ± 0.26	✓	-	✓	-
H	Core R67E (Rrp41)	2.4 ± 0.08 μM	-7.79 ± 0.02	✓	-	-	-
I	Core R67E (Rrp41) Rrp4 WT	0.56 ± 0.03 μM	-8.67 ± 0.03	✓	-	✓	✓
J	Core R67E (Rrp41) Rrp4 KNK170ENQ	2.34 ± 0.53 μM	-6.66 ± 0.28	✓	-	✓	-
Second and third RNA binding events							
K	Core WT Rrp4 WT	20.4 ± 2.1 μM	-6.50 ± 0.06	-	-	✓	✓
L	Core WT Rrp4 KNK170ENQ	60.5 ± 4.6 μM	-5.85 ± 0.04	-	-	✓	-

a) Core refers to the Rrp41:Rrp42 exosome core complex.

b) The ΔG values were calculated for 303K.

c) The contact points that are present in the RNA:enzyme interaction are indicated

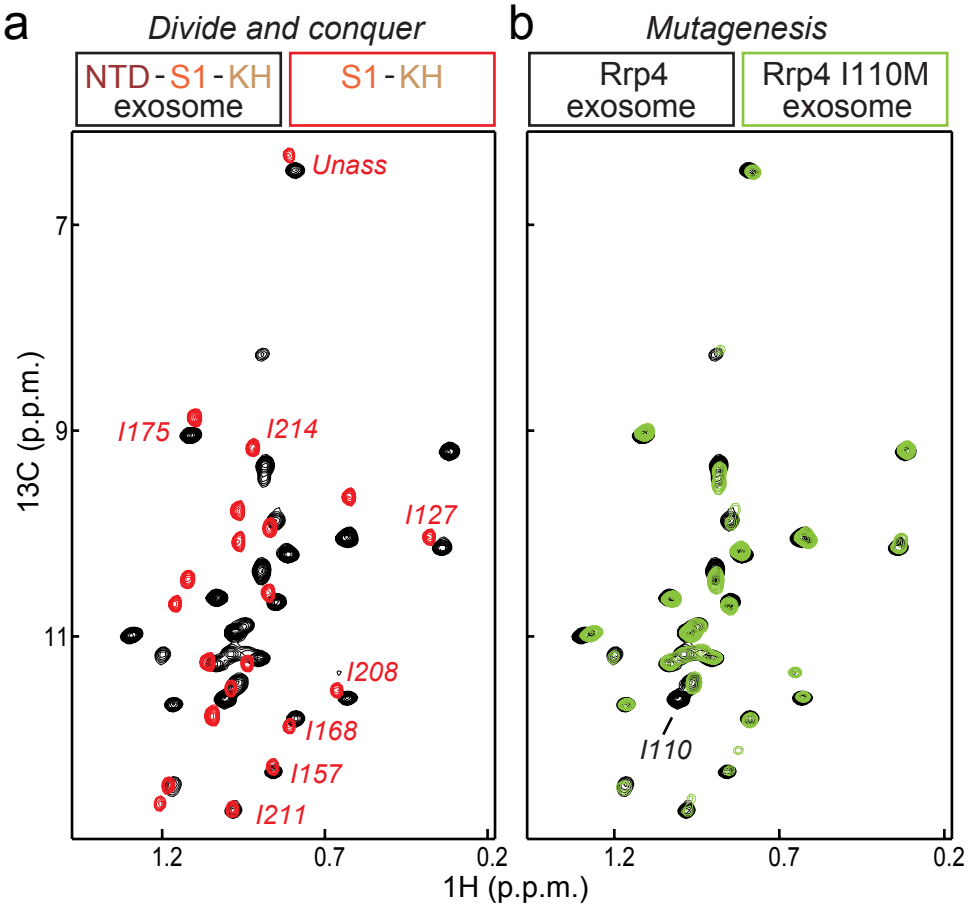
d) The error is extracted based on 2 (complex E, H), 3 (F, G, I, J) or 5 (A, B, C, D) independent measurements.

Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Table 3: RNA binding summary

RNA contact point	Intrinsic binding energy ΔG^i (kcal/mol)	Based on complexes
First RNA binding event		
Rrp4, Periphery	-0.52 ± 0.35	(B, C), (F, G) and (I, J)
Rrp4, Pore	-0.46 ± 0.39	(A, C), (E, G) and (H, J)
Rrp4, Cap (Periphery + Pore)	-0.98 ± 0.17	(A, B), (E, F) and (H, I)
Core, Neck	-2.30 ± 0.43	(A, E), (B, F), (C, G) and (A, H)
Core, Active sites	-3.15 ± 0.21	(A, D)
Second and third RNA binding events		
Rrp4, Periphery	-0.65 ± 0.08	(B, C)

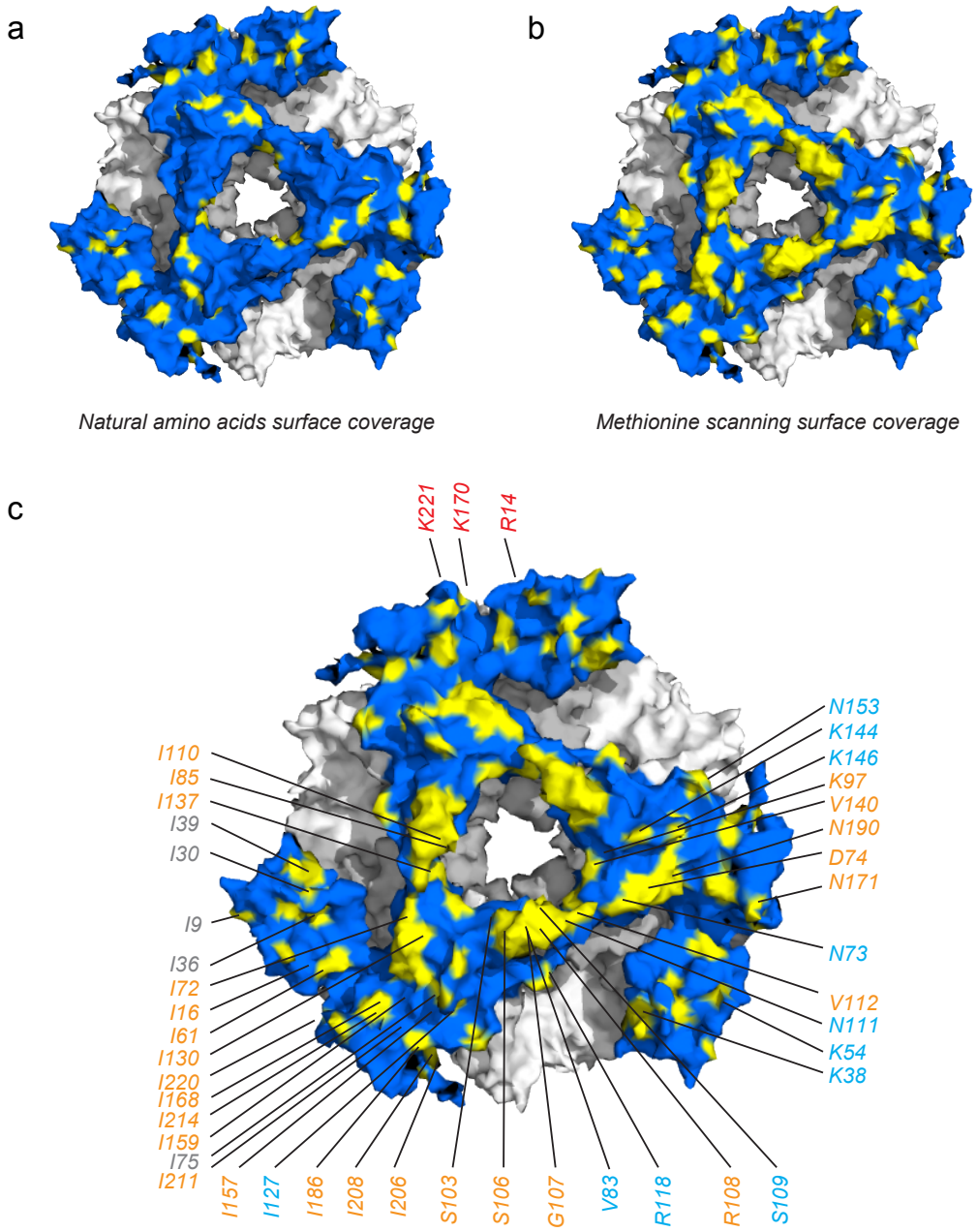
Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 1 | Assignment strategy
(a) Methyl TROSY spectra of Ile-δ1 labeled full-length Rrp4 in complex with the NMR inactive Rrp41-Rrp42 exosome core (270 kDa; black) and the monomeric Rrp4 S1-KH domain (21kDa; red). Multiple resonances of the small building block overlap with the resonances in the full complex and could be transferred. The methyl group assignments of the small building block were obtained by traditional TROSY-based NMR methods.

(b) Assignments of methyl group resonances were completed using a mutagenesis approach where individual isoleucine residues were mutated to alternative amino acids such that the corresponding resonance disappears from the methyl TROSY spectrum. Here, I110 was mutated to a methionine, which resulted in the straightforward assignment of the methyl group of residue 110.

Nature Chemical Biology: doi:10.1038/nchembio.2328



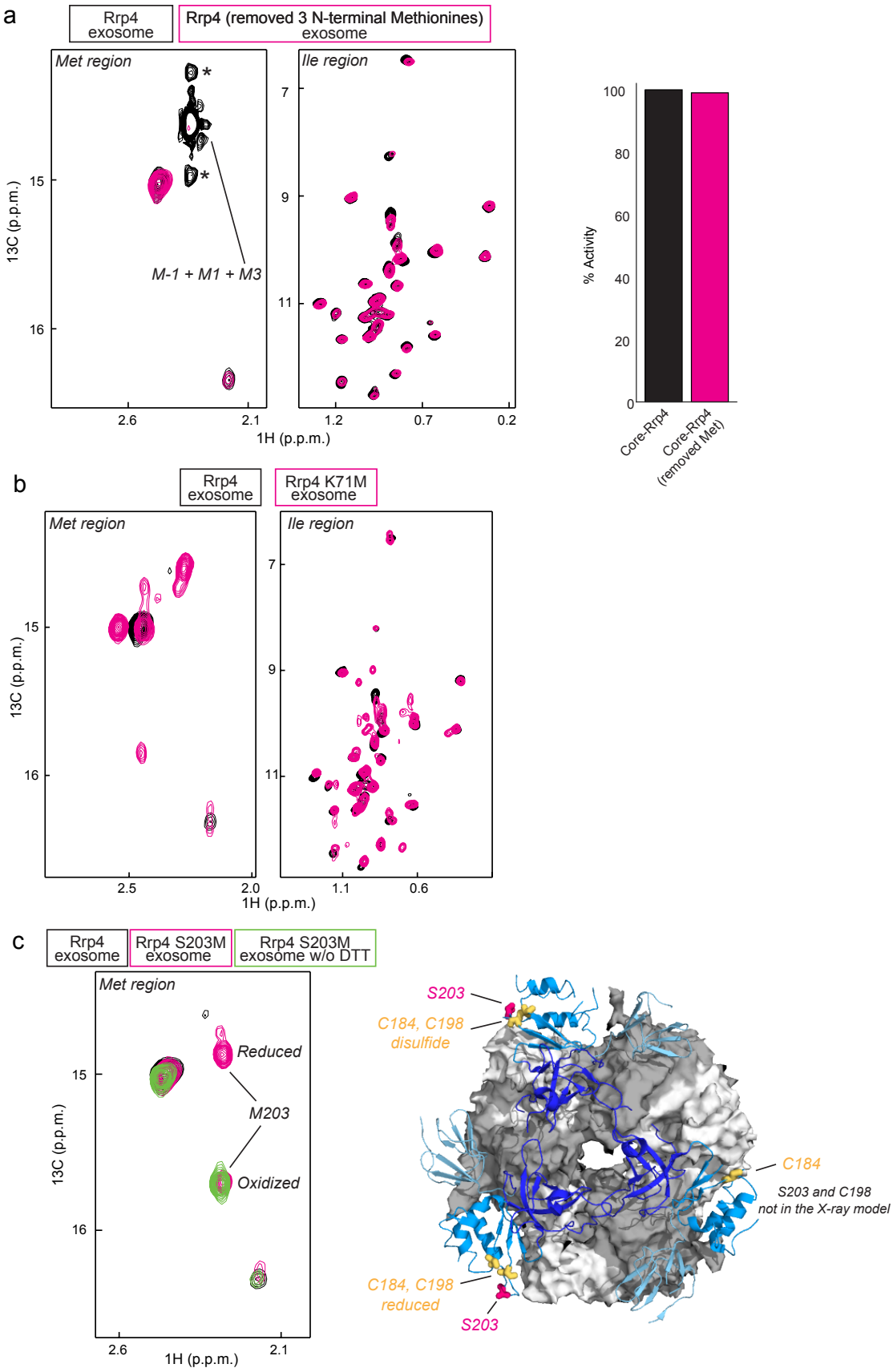
Supplementary Figure 2 | Coverage of methyl groups on the surface of the Rrp4 cap.

(a) Naturally occurring isoleucine residues (shown in yellow) only cover a small portion of the surface of the Rrp4 cap (shown in blue). The NMR invisible Rrp41-Rrp42 exosome core is shown in gray.

(b) As in (a), but in this case all methionine reporter residues that we introduced (one at a time) are also displayed in yellow. Methionine scanning thus significantly increases the surface coverage on the Rrp4 cap allowing for an accurate mapping of the RNA interaction surface.

(c) Location of all natural isoleucine and methionine residues as well as the residues that are used for methionine scanning. Note that all residues appear three times on the surface, but have only been labeled once for clarity. Residues labeled in grey are unassigned, residues labeled in cyan are outside the RNA binding groove and residues labeled in orange are inside the RNA binding groove.

Nature Chemical Biology: doi:10.1038/nchembio.2328



Nature Chemical Biology: doi:10.1038/nchembio.2328

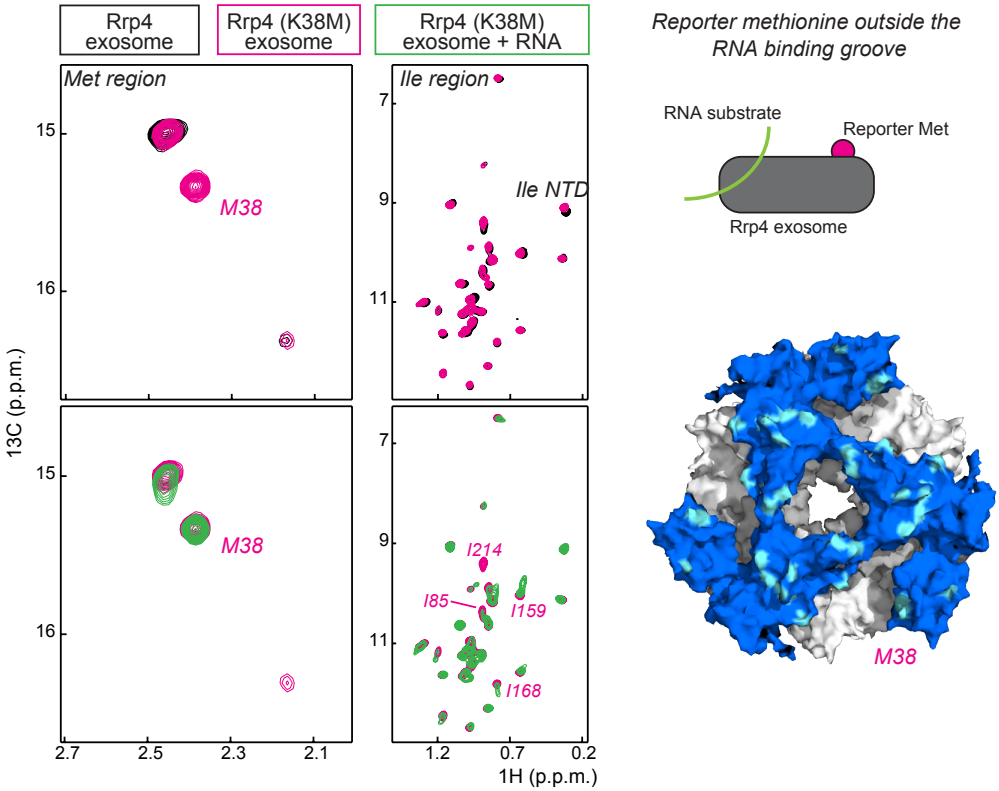
Supplementary Figure 3 | Methionine spectra
(a) The unstructured N-terminus of Rrp4 contains three methionine residues (M-1, M1 and M3). These flexible methyl groups result in a strong signal in the methyl TROSY spectrum (black). Mutation of these methionines into serines results in a significant simplification of the methionine part of the methyl TROSY spectrum (pink; left), without changing the structure of the complex as judged from the minimal changes in the isoleucine part of the spectrum (right spectrum). In addition, these mutations had no effect on the activity of the complex (right). In our study, we refer to this Rrp4 protein (Rrp4_3MS) that lacks these strong signals in the methionine part of the spectrum as the WT protein. Resonances indicated with an asterisk result from truncation artifacts that arise during the processing of the NMR data. Position M-1 was introduced in the Rrp4 protein due to the cloning of the DNA construct into the overexpression vector.

(b) Example of a reporter methionine that resulted in a partial destabilization of the Rrp4 protein in the Rrp4-exosome complex (K71M). It should be noted that such a destabilizing effect was only observed in one case and that this is not a general bottleneck of the applied methodology. Nevertheless, the close inspection of the spectra is required to assess the effects of the introduced mutations. Here, the K171M mutation shows multiple new resonances in the methionine part of the spectrum (left). At the same time, a significant number of additional resonances arise in the isoleucine part of the spectrum. Both are clear indications that the K171M mutation is not folded in the same stable manner as the WT Rrp4 protein in complex with the exosome. The K171M Rrp4 methionine reporter mutant was not used for further binding studies.

(c) For the S203M mutant we observed two additional resonances in the methionine part of the spectrum that are due to the partial formation of the C184-C198 disulfide bond within the Rrp4 monomer. Since we observed that the exosome's interaction with the RNA substrate remains unaltered under oxidizing conditions and that the presence of reducing agents causes doubling of certain resonances from isoleucine residues around the mentioned cysteines, we performed our NMR measurements in the absence of any reducing agent. The formation of this disulfide bond is also visible in one of the published crystal structures of the Rrp4-exosome complex (Right, PDB: 3L7Z) [1]. In that study the disulfide bond leads to a break in the symmetry of the trimetric Rrp4 cap in the crystal structure.

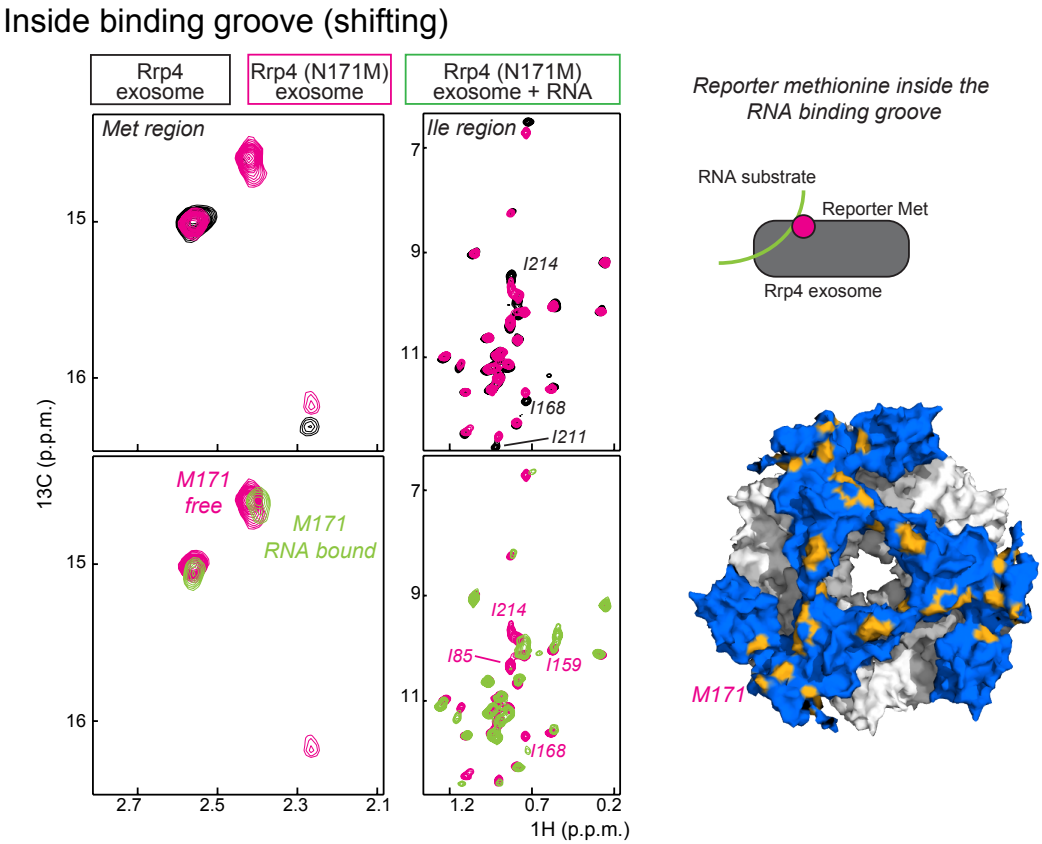
Nature Chemical Biology: doi:10.1038/nchembio.2328

Outside binding groove



Supplementary Figure 4 | Reporter methionine outside the RNA binding groove.
Example of a reporter methionine that is outside the RNA binding groove. The introduced methionine reporter (K38M) is located in the NTD and results in the appearance of a single new resonance in the methionine part of the NMR spectrum (top left, black spectrum: WT; pink spectrum: K38M). The reporter methionine resonance can be assigned instantaneously (top left spectrum) and only causes minor CSPs in the naturally occurring isoleucine and methionine resonances (top left and top right spectra). Addition of RNA to the K38M Rrp4-exosome complex results in CSPs of the naturally occurring isoleucine residues (bottom right spectrum) that are comparable to the CSPs observed for the WT protein (compare: Fig. 1b), showing that the RNA interacts normally with the complex. The K38M reporter methionine resonance fails to show CSPs (bottom left spectrum) indicating that residue K38 is not part of the RNA interaction interface. A cartoon displaying this scenario is shown on the top right. All reporter methionine and natural isoleucine residues that we determined to be outside the RNA interaction site are shown in cyan on the structure of the complex (bottom right).

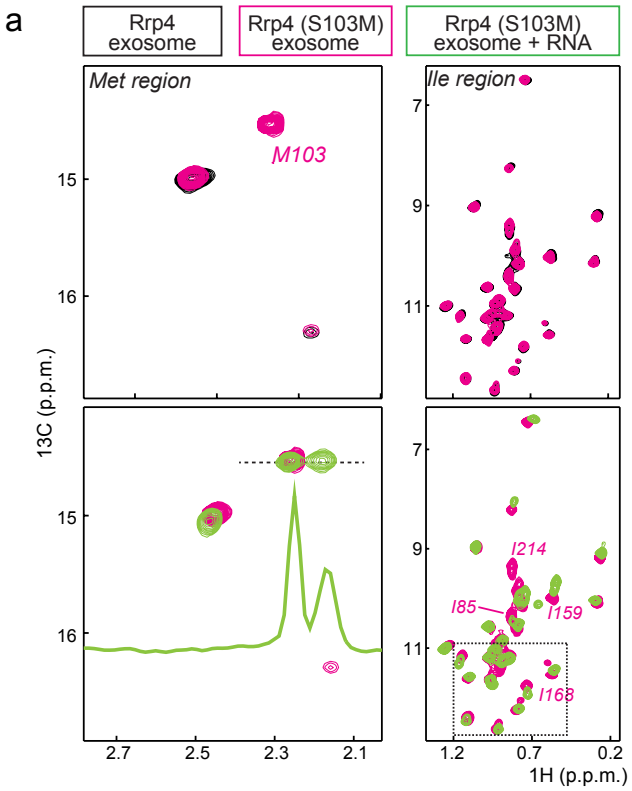
Nature Chemical Biology: doi:10.1038/nchembio.2328

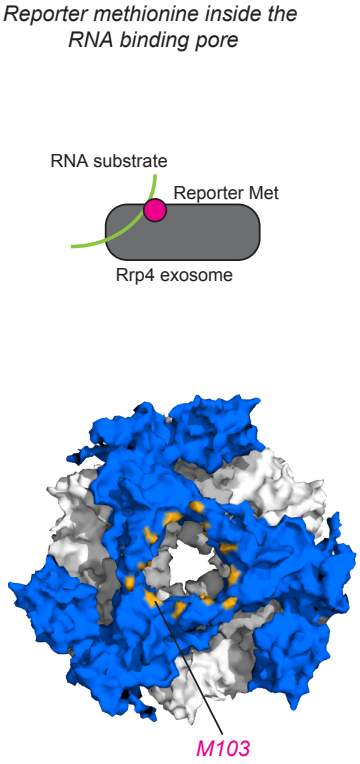


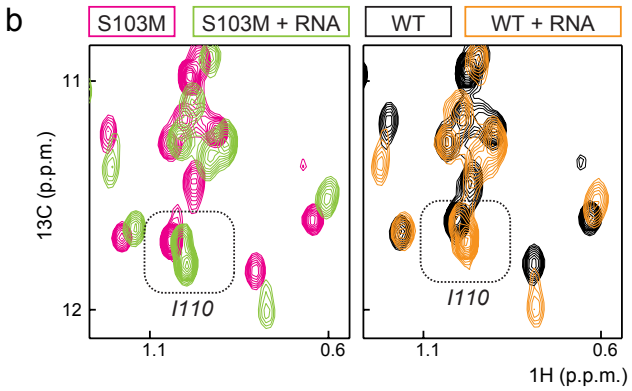
Supplementary Figure 5 | Reporter methionine inside the RNA binding groove.
As in Supplementary Fig. 4, but for the case where the reporter methionine (N171M) is inside the RNA interaction groove. The reporter methionine appears as a single novel resonance in the methionine part of the spectrum (top left spectrum) and causes minor CSPs of resonances close to the site of mutation (top right spectrum). Addition of RNA results in CSPs of the introduced reporter methionine (bottom left spectrum) and of the naturally occurring isoleucine residues (bottom right spectrum). A summary of this scenario is shown as a cartoon on the top right and occurred for all residues that are indicated in orange on the surface of the complex (bottom right)

Nature Chemical Biology: doi:10.1038/nchembio.2328

Inside binding groove (splitting)





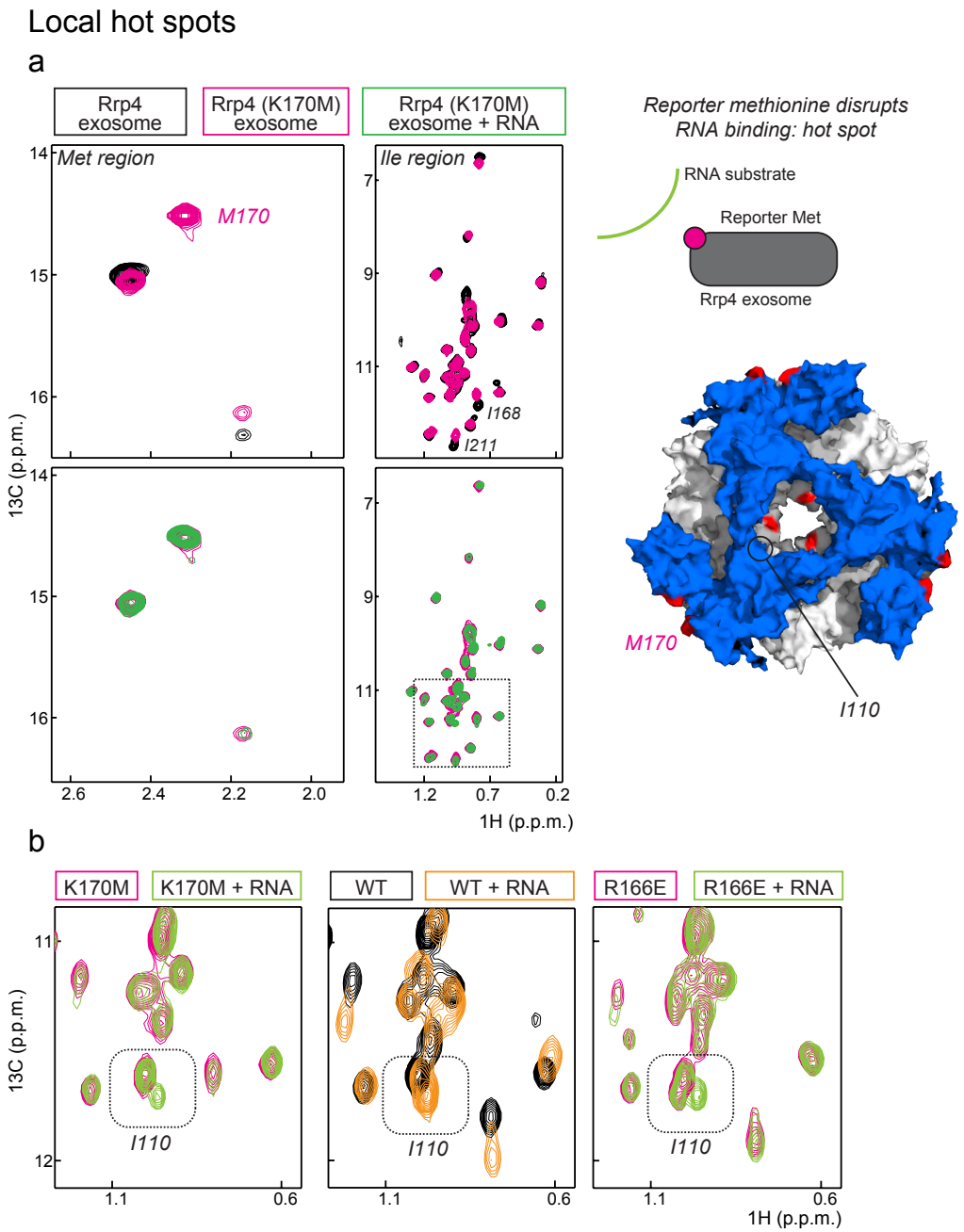


Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Figure 6 | Reported methionine residues in the pore of Rrp4.
(a) As in S4 and S5, where the reporter methionine is located in the RNA entrance pore of Rrp4. The reporter methionine (S103M) can be assigned in a straightforward manner (top left spectrum) and causes only minor CSPs in residues that are close to the site of mutation (top right spectrum). Addition of RNA results in a splitting of the resonance of the reporter methionine (bottom left spectrum) and CSPs in the naturally occurring isoleucine residues that are comparable to the ones observed in the titration experiments with WT Rrp4-exosome (compare: Fig. 1b). A cartoon for this situation is shown on the top right and all sites that show this behavior are indicated with orange on the surface of the protein (bottom right). The splitting of resonances indicates that for those residues the RNA is in slow exchange on the NMR chemical shift timescale between the free and the bound form. This suggests that the RNA has a different affinity for the complex as was observed in S5.

(b) The boxed region of the bottom right spectrum in (A) is enlarged (left spectrum) and shows that I110, that is located in the Rrp4 pore, also undergoes a signal splitting upon addition of RNA. The observed signal splitting is not due to the reporter methionine (S103M) as we also observe the same splitting of I110 in the WT protein upon addition of RNA (right spectrum). The signal splitting we observe for a number of residues is thus a general phenomenon and not due to the introduced methionine residues.

Nature Chemical Biology: doi:10.1038/nchembio.2328



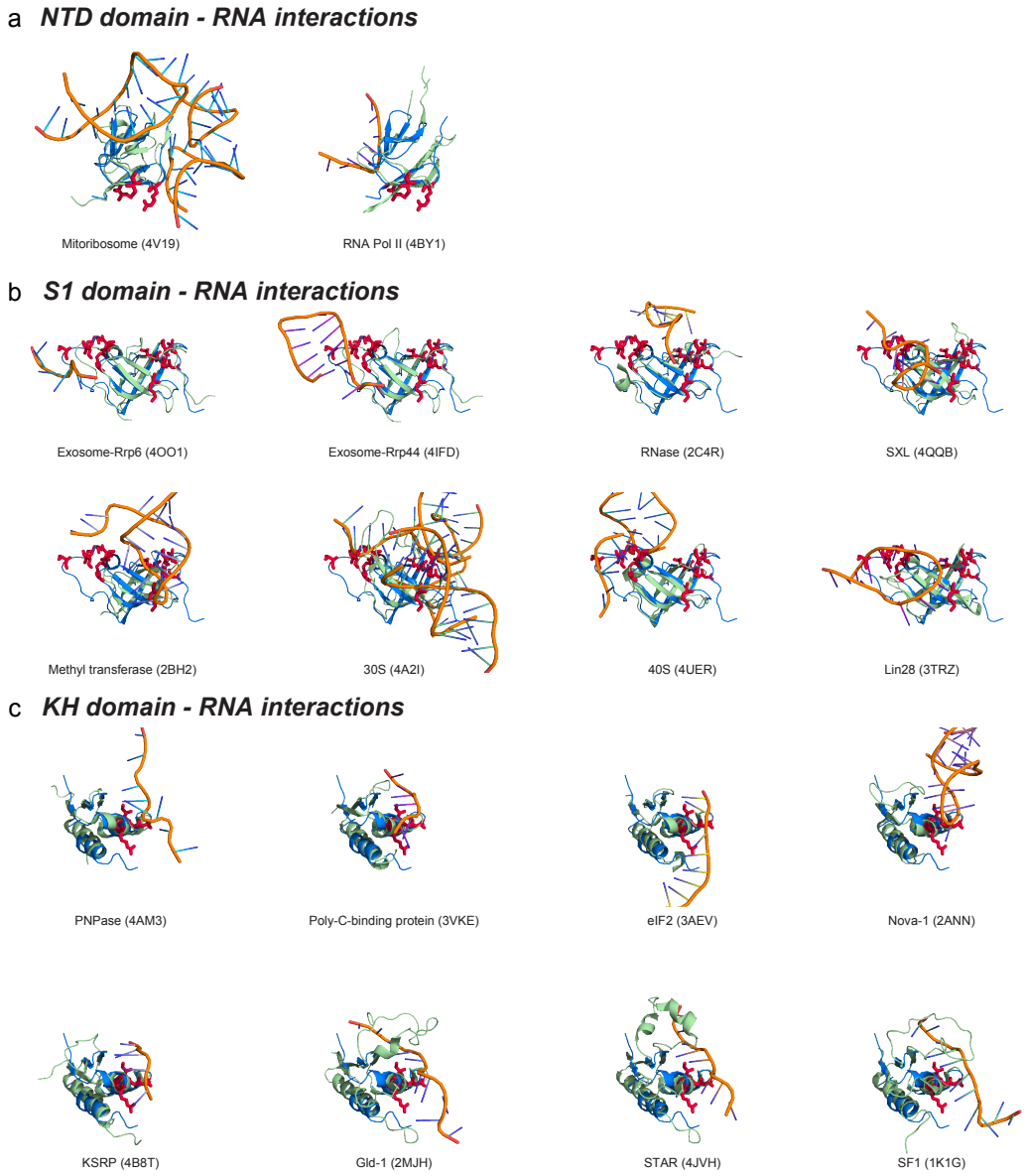
Supplementary Figure 7 | Identification of local hot spot residues.
(a) The resonance of the reporter methionine K170M is identified in a straightforward manner (top left spectrum) and only causes minor CSPs in the isoleucine residues (top right spectrum). Addition of RNA to the K170M mutant does not result in CSPs of the reporter methionine (bottom left spectrum), but also fails to induce the CSPs that we observed for the WT protein in the isoleucine residues (bottom right spectrum, compare: Fig. 1b). This indicates that the K170M mutation interferes with the RNA interaction as indicated by the cartoon (top tight). A summary of the hot-spots is shown in red on the surface of the complex (bottom right).

(b) Right: Enlarged view of the boxed region of the bottom right spectrum in (a). Despite the loss of many CSPs in the RNA titration of the K170M mutant, residue I110, that is located in the pore, displays CSPs that are the same as for the RNA

Nature Chemical Biology: doi:10.1038/nchembio.2328

titration with the WT protein (middle). This shows that the K170M mutation resulted in a local loss of the RNA:Rrp4 interaction at the periphery of the complex, but that the RNA was still able to interact with the Rrp4 pore region. K170 is thus a local hot-spot in the complex. Not only K170 is part of the hot-spot region and also other mutations in the KH loop region (e.g. R166E) resulted in a local loss of the RNA affinity without perturbing the interactions at the RNA pore (right spectrum).

Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 8 | Known RNA interactions of NTDs, S1 and KH domains.
A DALI search [2] using the NTD, S1 or KH domain of the Rrp4 protein (2JEA, REF [3]) was performed and structures containing nucleic acids were selected. These protein-RNA complexes were then superimposed on the NTD (a), S1 (b) or KH (c) domains of the Rrp4 protein to assess if the RNA binding site that we map here compares to the RNA interaction sites that have previously been found in these protein domains. The Rrp4 domains are shown in blue, where the residues that we mapped to the interaction interface are shown in red. The domains found using the DALI search are shown in light green, where the RNA in those complexes is shown in orange. For the NTD (a), residues that we identified have not been identified to be involved in RNA interactions in structural homologues. For the S1 and KH domains (b, c), the surface that we identified corresponds to surfaces of these domains that have previously been shown to be involved in RNA interactions.

Nature Chemical Biology: doi:10.1038/nchembio.2328

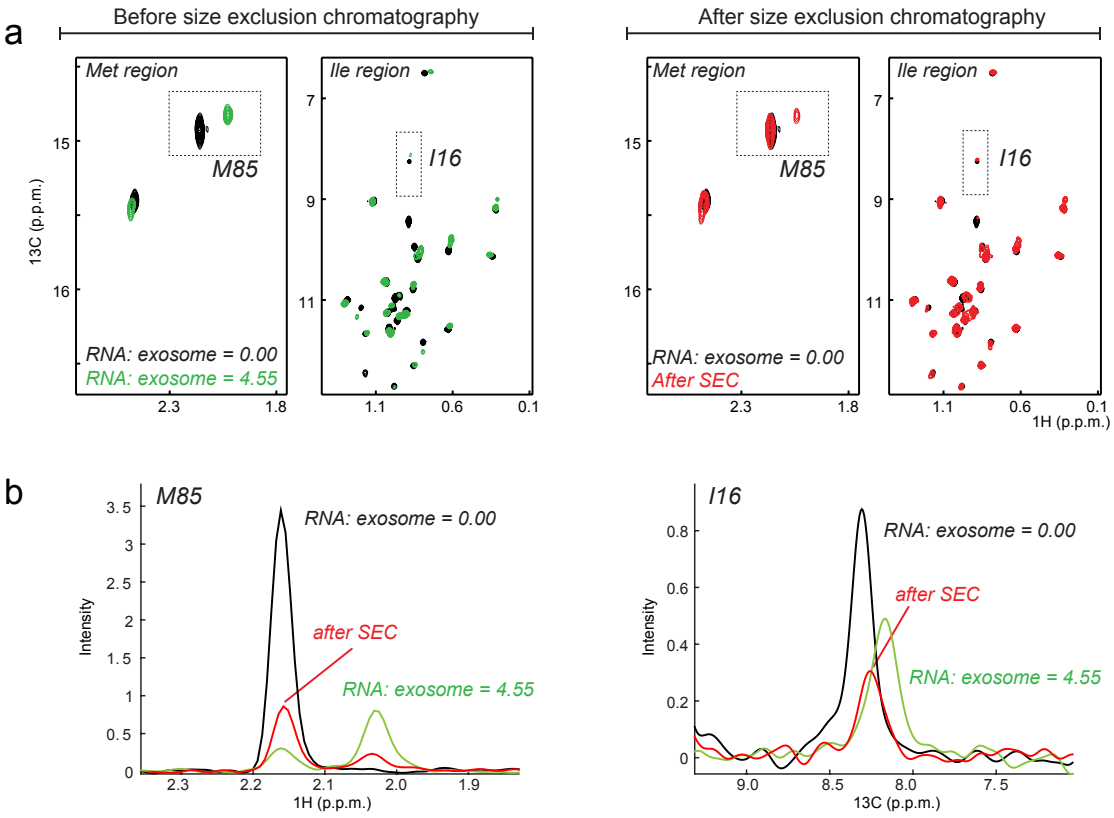
The following structures are shown (PDB code and reference):

For the NTD: 4V19 (REF [4]), 4BY1 (REF [5])

For the S1 domain: 4OO1 (REF [6]), 4IFD (REF [7]), 2C4R (REF [8]), 4QQB (REF [9]), 2BH2 (REF [10]), 4A2I (REF [11]), 4UER (REF [12]), 3TRZ (REF [13]).

For the KH domain: 4AM3 (REF [14]), 3VKE (REF [15]), 3AEV (REF [16]), 2ANN (REF [17]), 4B8T (REF [18]), 2MJH (REF [19]), 4JVH (REF [20]), 1K1G (REF [21])

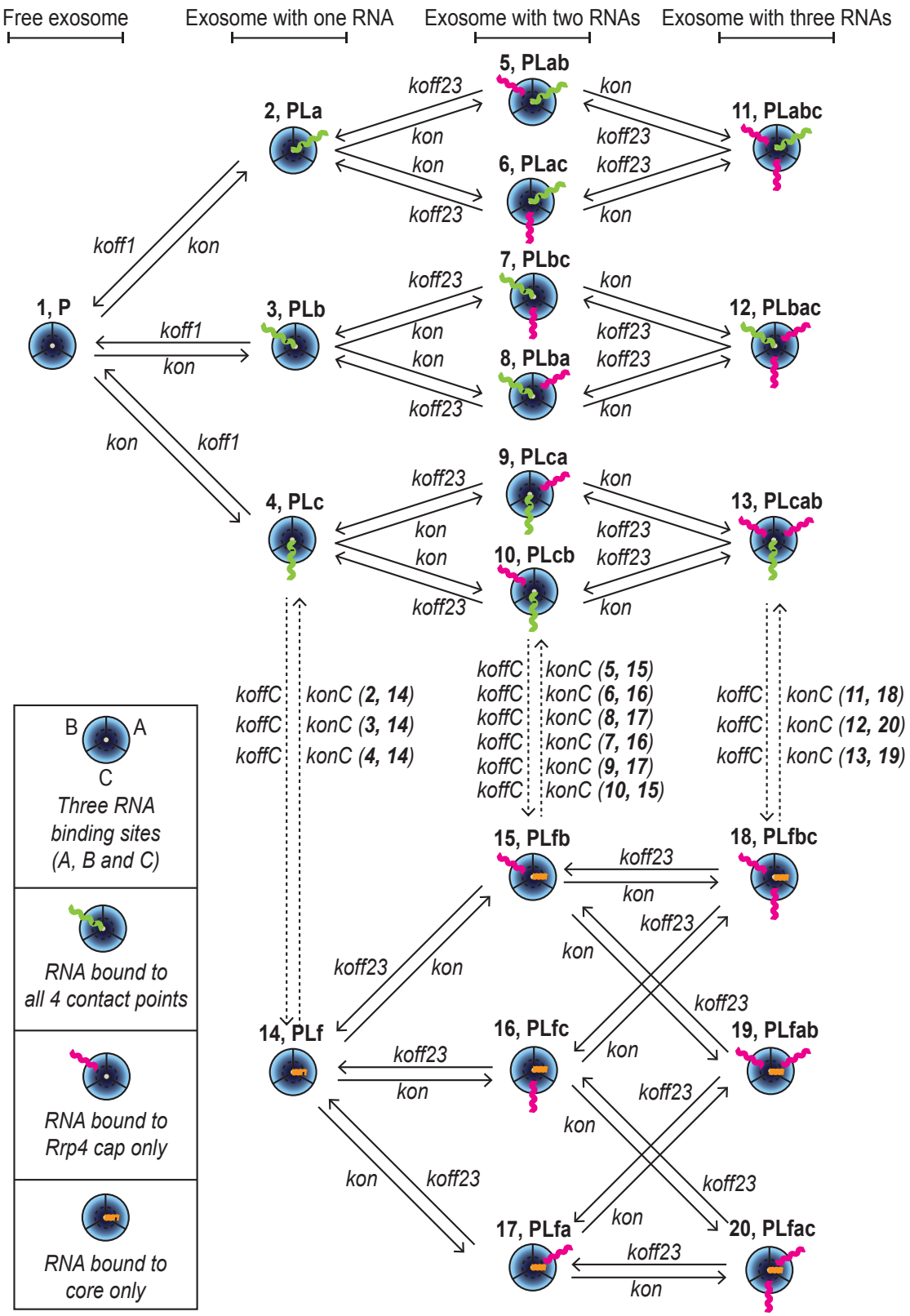
Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 9 | Removal of weakly bound RNA
(a) NMR spectra of the Rrp4-exosome complex in the absence (black) and presence of an excess of RNA substrate (green). The excess substrate was removed using size exclusion chromatography, which results in the red spectrum. Regions that were used to extract one-dimensional projections (b) are indicated with dashed boxes.

(b) Projections of the boxed regions in (a). The projection of M85 was done in the carbon dimension, whereas the projection of I16 was done in the proton dimension. The fully RNA loaded complex (green) lost a significant amount of RNA during size exclusion chromatography (red) and is in-between the free (black) and fully loaded (green) complex, both for residue M85 (that is in slow exchange on the NMR chemical shift timescale) and for I16 (that is in fast exchange on the NMR chemical shift timescale).

Nature Chemical Biology: doi:10.1038/nchembio.2328



Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Figure 10 | Model for the interaction between the exosome and RNA substrate.
The free exosome (complex 1, where P refers to the exosome protein complex) can recruit a first RNA with nM affinity (Table 1 of the main text). This RNA can be recruited to site A, B or C on the Rrp4-cap to form a complex with 1 strongly bound RNA (complex 2, 3 or 4, where the RNA that interacts strongly with the 4 Rrp4-exosome contact points is indicated in green; L refers to the RNA ligand and PL to the protein complex with one ligand. PLa, PLb and PLc refer to the situations where the ligand is bound to site A, B or C, respectively). The on-rate (kon) and the off-rate (koff1) that are associated with this binding event are indicated. In the fitting procedure, we fix the affinity of the first interaction to the affinity that we determined based on fluorescence anisotropy experiments (Talbe 1 of the main text).

Subsequently a second RNA can be recruited to the Rrp4-exosome to form a complex with two bound RNAs (complexes 5 to 10, where the RNA that only interacts with the Rrp4 cap is indicated in pink. PLxy refers to a protein complex that has ligand bound to sites X and Y). The second RNA is bound weaker than the first RNA. The on-rate for this complex formation (kon) and the off-rate (koff23) are indicated. To reduce the number of fitting parameters, we assume that the on-rate for the binding of the first RNA and the second RNA are identical.

The Rrp4-exosome complex that contains two RNA species can recruit a third RNA (complex 11, 12 or 13, where the third RNA only interacts with Rrp4 cap is also indicated in pink). The associated on- and off-rates for the third binding event are assumed to be identical to those for the second RNA binding event as the same interaction points are involved.

The 5' end of the RNA that is bound to the exosome core (first RNA interaction; indicated in green) can dissociate from the Rrp4 cap due to the very small intrinsic binding energy for the interaction between the Rrp4 cap and the 5' end of the RNA (see Table 1). In case this occurs for e.g. complex 2 (or 3 or 4), this leads to the formation of complex 14 (where PLf, refers to a protein complex with one ligand that has a 5' end that is released or freed from the cap). The RNA that is bound only to the core (and not to the Rrp4 cap) is indicated in orange in the cartoon. Likewise, the dissociation of the 5' end of the substrate for the first bound RNA can convert complexes 5 to 10 into complexes 15 to 17 and complexes 11 to 13 into complexes 18 to 20. We describe this exchange process with the on- and off-rates konC and koffC (where "C" refers to the Cap").

The interaction between three RNA species and the Rrp4-exosome complex can thus be accurately described using 20 different microscopic states and a set of 5 exchange parameters (kon, koff1, koff23, konC and koffC). The exchange parameters kon and koff1 determine the affinity for the first RNA binding event, the exchange parameters kon and koff23 determine the affinity for the second and third RNA binding event and the exchange parameters konC and koffC determine the fraction where the RNA that binds to the core has a 5' end that is released from the cap.

Nature Chemical Biology: doi:10.1038/nchembio.2328

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-3*kon*L	koff1	koff1	koff1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	kon*L	-koff1-2*kon*L-koffC	0	0	koff23	koff23	0	0	0	0	0	0	0	koffC	0	0	0	0	0	0
3	kon*L	0	-koff1-2*kon*L-koffC	0	0	0	koff23	koff23	0	0	0	0	0	koffC	0	0	0	0	0	0
4	kon*L	0	0	-koff1-2*kon*L-koffC	0	0	0	0	koff23	koff23	0	0	0	koffC	0	0	0	0	0	0
5	0	kon*L	0	0	-koff23-kon*L-koffC	0	0	0	0	0	koff23	0	0	0	konC	0	0	0	0	0
6	0	kon*L	0	0	0	-koff23-kon*L-koffC	0	0	0	0	koff23	0	0	0	0	konC	0	0	0	0
7	0	0	kon*L	0	0	0	-koff23-kon*L-koffC	0	0	0	0	koff23	0	0	0	konC	0	0	0	0
8	0	0	kon*L	0	0	0	0	-koff23-kon*L-koffC	0	0	0	koff23	0	0	0	0	konC	0	0	0
9	0	0	0	kon*L	0	0	0	0	-koff23-kon*L-koffC	0	0	0	koff23	0	0	0	konC	0	0	0
10	0	0	0	kon*L	0	0	0	0	0	-koff23-kon*L-koffC	0	0	koff23	0	konC	0	0	0	0	0
11	0	0	0	0	kon*L	kon*L	0	0	0	0	-2*koff23-koffC	0	0	0	0	0	0	konC	0	0
12	0	0	0	0	0	0	kon*L	kon*L	0	0	0	-2*koff23-koffC	0	0	0	0	0	0	0	konC
13	0	0	0	0	0	0	0	0	kon*L	kon*L	0	0	-2*koff23-koffC	0	0	0	0	0	konC	0
14	0	koffC	koffC	koffC	0	0	0	0	0	0	0	0	0	-3*konC-3*kon*L	koff23	koff23	koff23	0	0	0
15	0	0	0	0	koffC	0	0	0	0	koffC	0	0	0	kon*L	-2*konC-koff23-2*kon*L	0	0	koff23	koff23	0
16	0	0	0	0	0	koffC	koffC	0	0	0	0	0	0	kon*L	0	-2*konC-koff23-2*kon*L	0	koff23	0	koff23
17	0	0	0	0	0	0	0	koffC	koffC	0	0	0	0	kon*L	0	0	-2*konC-koff23-2*kon*L	0	koff23	koff23
18	0	0	0	0	0	0	0	0	0	0	koffC	0	0	0	kon*L	kon*L	0	-2*koff23-konC	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	koffC	0	kon*L	0	kon*L	0	-2*koff23-konC	0
20	0	0	0	0	0	0	0	0	0	0	0	koffC	0	0	0	kon*L	kon*L	0	0	-2*koff23-konC

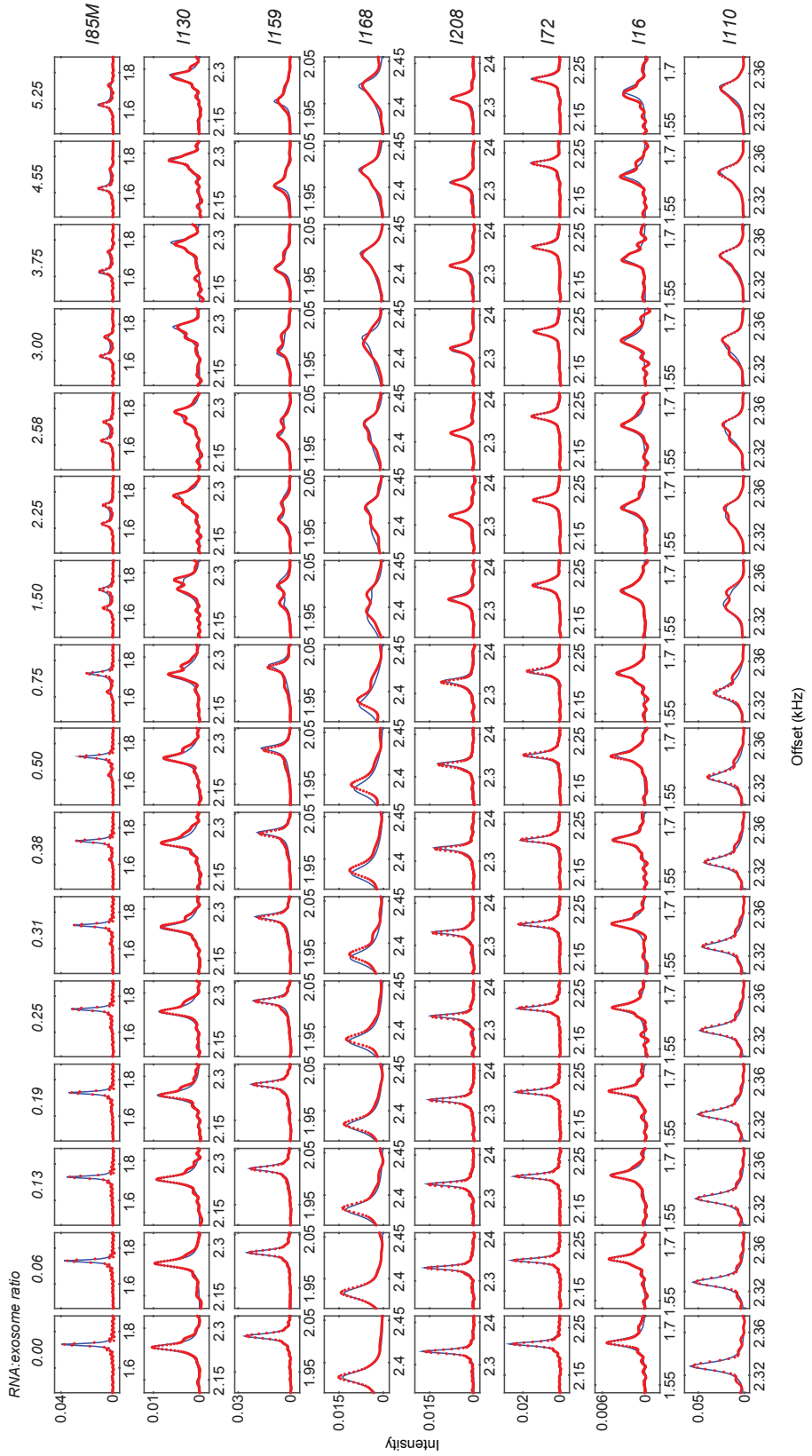
Supplementary Figure 11 | Exchange matrix for the binding between three RNA substrates and the Rrp4-exosome

To simulate the NMR spectra during the NMR titration experiment the exchange matrix (above) is multiplied with the diagonal relaxation and chemical shift matrixes and with the population vector. The populations of the different states (1 to 20) were determined based on the extracted exchange parameters in an iterative manner. The relaxation rates and resonance frequencies of the free resonance were directly extracted from the NMR spectra. The relaxation rates and resonance frequencies of the bound state were estimated from the spectra and subsequently optimized in the global fitting procedure, as these parameters cannot be accurately determined based on the spectra (due to less than 100% saturation of all binding sites at the last step of the titration experiment).

To extract the exchange parameters (kon, koff23, konC and koffC) these were optimized to minimize the square of the difference between the experimental and simulated NMR spectra. Koff1 is determined based on the measured affinity for the first binding event (67.5 nM; Table 1) and kon. In addition, for each residue a single scaling factor is optimized to correct for differences in signal intensity between residues and for each spectrum a single scaling factor is optimized to correct for slight changes in the intensity in different titration points (e.g. due to loss in protein during the titration experiment due to extended measurement times and due to small differences that result from potential imperfections shimming and pulse calibrations).

In total 8 different residues (M85, I130, I159, I168, I208, I72, I16 and I110) were fitted to the global exchange parameters simultaneously. Errors in the exchange parameters were based on a Jackknife approach, where the fitting routine was repeated eight times with a dataset where a single residue was excluded.

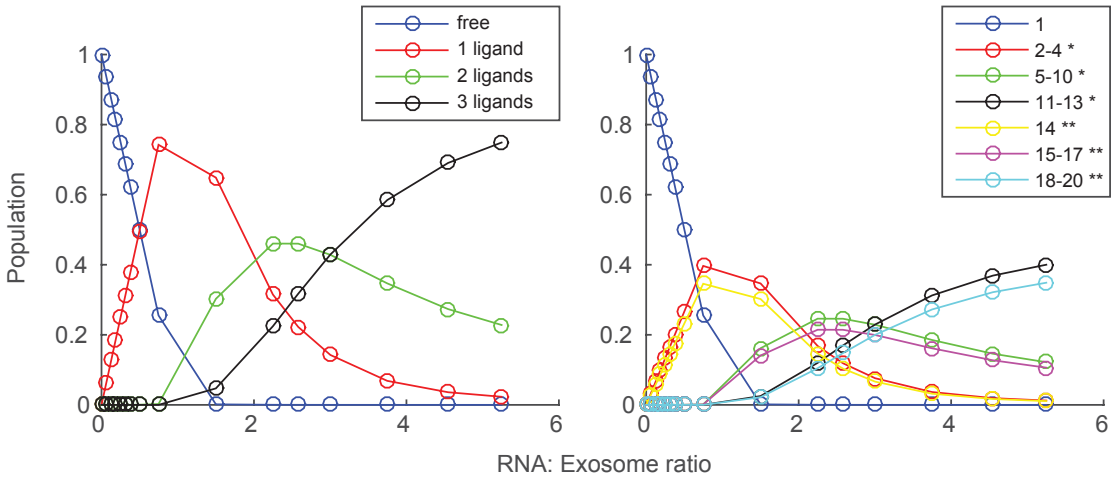
Nature Chemical Biology: doi:10.1038/nchembio.2328



Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Figure 12 | Experimental and back-calculated NMR spectra during the addition of RNA to the Rrp4-exosome.
1D traces extracted from the 2D NMR spectra (red dots; experimental data) superimposed on the spectra that are back-calculated based on the determined exchange parameters (blue lines). All residues and all spectra have been simultaneously fitted to determine the associated exchange parameters.

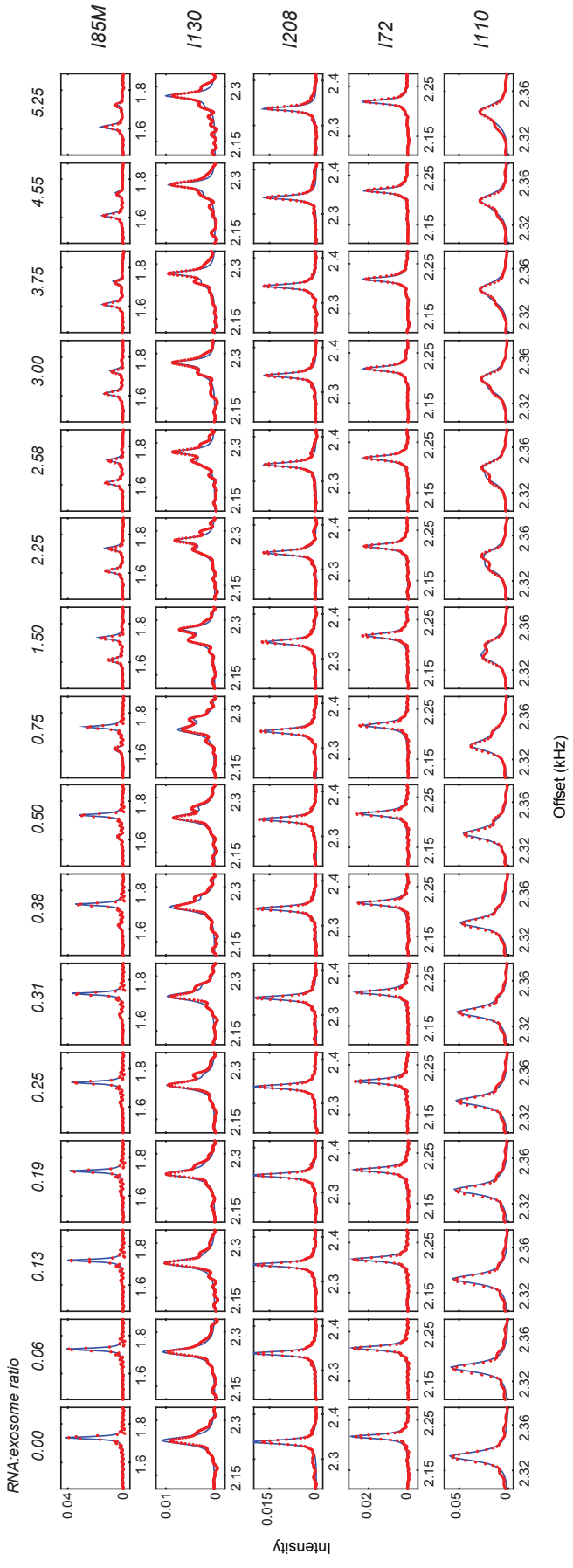
Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 13 | Populations of the different states
Left: Populations of the Rrp4 exosome complex with no, one, two or three RNA molecules bound during the NMR titration experiment. The curves are based on the affinities of the first and subsequent (second and third) RNA binding events. Right, as the plot on the left, but where is distinguished between states where the 5' end that is bound (*) or released (**) from the cap structure. The states refer to the states in Supplementary Fig. 10a.

The KonC and KoffC values that we obtain from the fitting are 165 ± 26 Hz and 145 ± 26 Hz respectively. Based on these exchange parameters, the apparent kD for the interaction between the 5' end of the RNA and the cap structure is very low and the energy involved ($\Delta G = -RTln(\frac{K_{on,C}}{K_{off,C}})$) is 0.081 ± 0.15 kcal/mol. The error in ΔG is large (twice the measured value), which is due to the large uncertainties in the determined KonC and KoffC rates. Importantly, the small ΔG value is in good qualitative agreement with the small intrinsic binding energy (ΔG^i = -0.98 ± 0.17kcal/mol) that we determined for the Rrp4 cap.

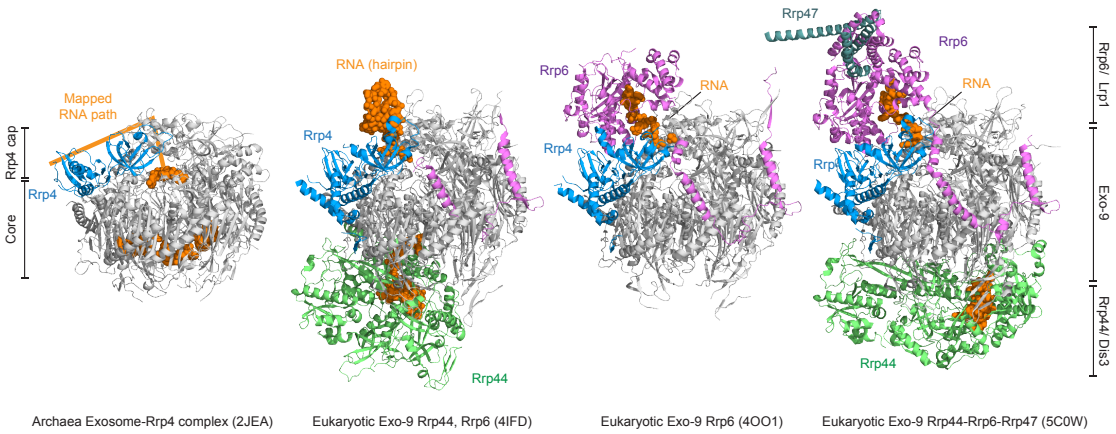
Nature Chemical Biology: doi:10.1038/nchembio.2328



Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary Figure 14 | Experimental and back-calculated NMR spectra during the addition of RNA to the periphery mutant Rrp4-exosome.
1D traces extracted from the 2D NMR spectra (red dots; experimental data) superimposed on the spectra that are back-calculated based on the determined exchange parameters (blue lines). All residues and all spectra have been simultaneously fitted to determine the associated exchange parameters. Note that less residues are used here, compared to the fitting of the WT enzyme as residues that are close to the periphery of the enzyme don't display CSPs and are thus omitted from the fitting routine.

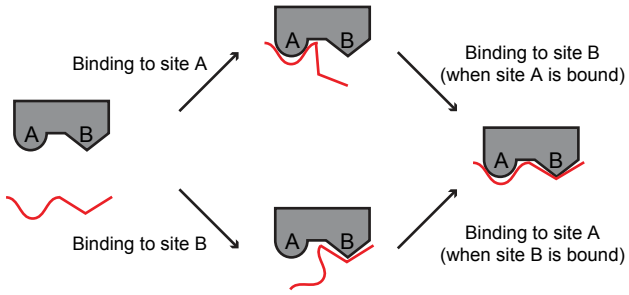
Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 15 | Comparison between the archaeal and eukaryotic exosome-RNA complexes

Structure of the archaeal exosome complex [3] (left, 2JEA), the structure of the eukaryotic exo-9 (core + cap proteins) in complex with RNA, Rrp44 and a part of Rrp6 [7] (4IFD), the eukaryotic exo-9 in complex with RNA and Rrp6 [6] and the eukaryotic exo-9 in complex with RNA and Rrp44, Rrp6 and a part of Rrp47 [22] (5C0W). The path of the RNA we identify here in the archaeal complex covers a large part of the Rrp4 cap protein (indicated by the orange line). The corresponding surface of Rrp4 is not used in the known structures of eukaryotic exosome-RNA complexes. In those structures, the RNA is either not contacting the Rrp4 surface (potentially due to the fact that the 5' end of the RNA that was used forms a hairpin structure) or the RNA contacts adaptor proteins that dock onto the eukaryotic Rrp4-Rrp40-Csl4 exosome cap. This indicates that RNA recognition by the eukaryotic exosome complex evolved further and is modulated by additional exosome components.

Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 16 | Relation between ΔG^i and ΔG^0
Consider the situation above, where a protein (grey) contains two binding sites for a ligand (red).

For the system described above ΔG^0_{AB}, ΔG^0_A and ΔG^0_B can be determined experimentally. These are the free energies of binding of the ligand to sites A and B simultaneously, to site A only and to site B only, respectively. From that we calculate the intrinsic binding energies for site A: $\Delta G^i_A = \Delta G^0_{AB} - \Delta G^0_B$ and for site B: $\Delta G^i_B = \Delta G^0_{AB} - \Delta G^0_A$. The ΔG^i values are larger (more negative) when the two binding sites influence each other in a constructive manner and they are lower (less negative) when the two binding events influence each other unfavorably [23,24], as illustrated below.

First, in the case **the ligand is rigid**, one binding event can position the other interface optimally for the second interaction. In that scenario, only the first binding step includes an entropic penalty (the ligand will loose degrees of freedom upon complex formation in the first binding step). We here refer to this energetic penalty as ΔG^S. When calculating ΔG^i_A and ΔG^i_B (as defined above) the term ΔG^S is cancelled out (ΔG^0_{AB}, ΔG^0_A and ΔG^0_B contain the same amount of ΔG^S). As a result $\Delta G^i_A + \Delta G^i_B < \Delta G^0_{AB}$ (the sum of the intrinsic binding energies is more negative than the free energy of binding).

Second, in case when **the ligand is highly flexible**, both the first and second binding event needs to overcome an entropic barrier (ΔG^S). The loss of entropy for the interaction of the ligand when bound to both site (A and B; twice a ΔG^S term) is thus larger than it is for the interaction with site A only or with site B only (only one ΔG^S term). The extracted ΔG^i_A and ΔG^i_B values are, in that scenario, thus less negative by ΔG^S. When ΔG^S is the same in all steps, $\Delta G^i_A + \Delta G^i_B = \Delta G^0_{AB}$

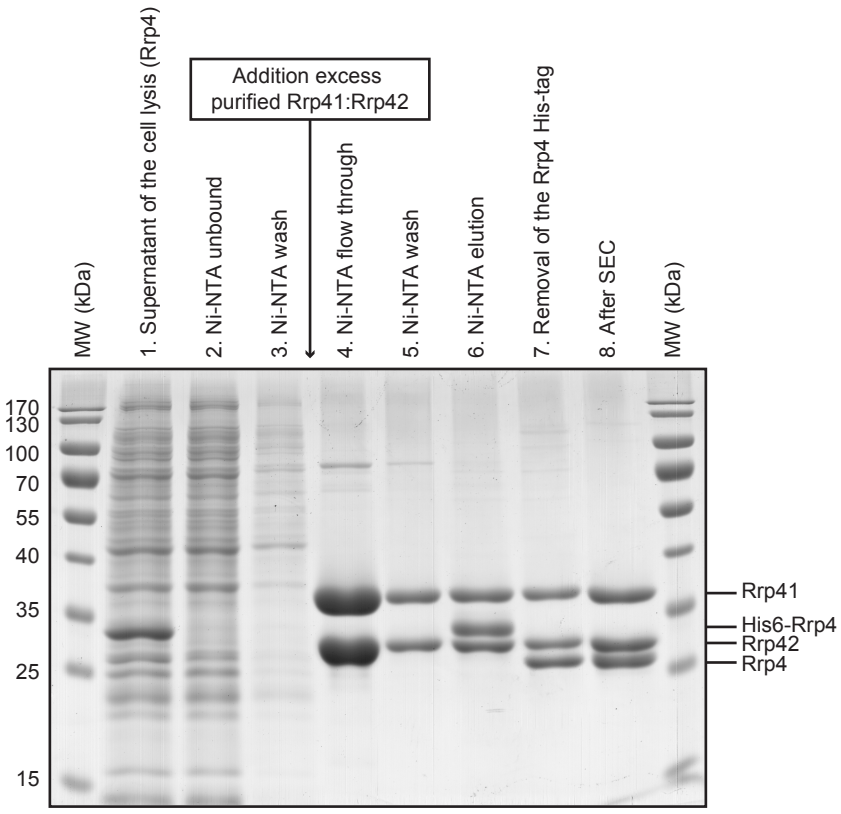
Third, in it is possible that **a flexible ligand is forced to adopt a high-energy structure to interact with both sites**. In that case, the interaction of a flexible ligand with the second site is energetically less favorable when the first site is already bound (due to strain in the fully bound ligand). In that case, the first binding event will need to overcome an entropic barrier ΔG^S, however, the second binding event will need to overcome both a ΔG^S barrier and an additional energy barrier (that is required to induce the higher energetic conformation in the ligand). The extracted ΔG^i_A and ΔG^i_B values are then less

Nature Chemical Biology: doi:10.1038/nchembio.2328

negative by one ΔG^S term and by the energy cost related with the unfavorable structure of the ligand. ΔG^0_{AB}, on the other hand, contains two ΔG^S terms and the energy term that relates to the unfavorable conformation of the fully bound ligand. In this case $\Delta G^0_{AB} < \Delta G^i_A + \Delta G^i_B$ (the sum of the intrinsic binding energies is less negative than the free energy of binding).

The magnitude of the sum of the intrinsic binding energies with respect to the total free energy of binding can thus provide insights into the mode of interaction between the ligand and the protein. In our current case, where a highly flexible RNA ligand interacts with the exosome, we have a situation where $\Delta G^0_{ABCD} < \Delta G^i_A + \Delta G^i_B + \Delta G^i_C + \Delta G^i_D$. From that we conclude that there is tension in the RNA molecule when it interacts with all 4 interactions sites in the exosome.

Nature Chemical Biology: doi:10.1038/nchembio.2328



Supplementary Figure 17 | Preparation of the NMR samples
IM-labeled Rrp4 is overexpressed in *E. coli*. Cells are lysed and the supernatant (lane 1) is applied to Ni-NTA resin. The unbound proteins (lane 2) and the un-specifically bound proteins (lane 3) are removed. The Rrp4 protein that is bound to the resin is supplemented with an excess of the exosome core. The excess core is removed (lane 4 and 5). The complex between Rrp4 and the exosome core that is formed on the column is eluted (lane 6) and subjected to TEV cleavage to remove the His-tag from Rrp4 (lane 7). The final sample is obtained after size exclusion chromatography of the complex (lane 8). Note that the stoichiometry of the final complex is 1:1:1.

Nature Chemical Biology: doi:10.1038/nchembio.2328

Supplementary references

1. Lu, C., Ding, F. & Ke, A. Crystal structure of the S. solfataricus archaeal exosome reveals conformational flexibility in the RNA-binding ring. *PLoS One* **5**, e8739 (2010).
2. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-9 (2010).
3. Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B. & Conti, E. RNA channelling by the archaeal exosome. *EMBO Rep* **8**, 470-6 (2007).
4. Greber, B.J. et al. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283-6 (2014).
5. Kinkelin, K. et al. Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proc Natl Acad Sci U S A* **110**, 15277-82 (2013).
6. Wasmuth, E.V., Januszyk, K. & Lima, C.D. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* **511**, 435-9 (2014).
7. Makino, D.L., Baumgartner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495**, 70-5 (2013).
8. Callaghan, A.J. et al. Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature* **437**, 1187-91 (2005).
9. Hennig, J. et al. Structural basis for the assembly of the Sxl-Unr translation regulatory complex. *Nature* **515**, 287-90 (2014).
10. Lee, T.T., Agarwalla, S. & Stroud, R.M. A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120**, 599-611 (2005).
11. Jomaa, A. et al. Cryo-electron microscopy structure of the 30S subunit in complex with the YjeQ biogenesis factor. *RNA* **17**, 2026-38 (2011).
12. Aylett, C.H., Boehringer, D., Erzberger, J.P., Schaefer, T. & Ban, N. Structure of a yeast 40S-eIF1-eIF1A-eIF3-eIF3j initiation complex. *Nat Struct Mol Biol* **22**, 269-71 (2015).
13. Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. & Sliz, P. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**, 1080-91 (2011).
14. Hardwick, S.W., Gubbey, T., Hug, I., Jenal, U. & Luisi, B.F. Crystal structure of Caulobacter crescentus polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. *Open Biol* **2**, 120028 (2012).
15. Yoga, Y.M. et al. Contribution of the first K-homology domain of poly(C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides. *Nucleic Acids Res* **40**, 5101-14 (2012).
16. Jia, M.Z., Horita, S., Nagata, K. & Tanokura, M. An archaeal Dim2-like protein, aDim2p, forms a ternary complex with a/eIF2 alpha and the 3' end fragment of 16S rRNA. *J Mol Biol* **398**, 774-85 (2010).
17. Teplova, M. et al. Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. *Structure* **19**, 930-44 (2011).
18. Nicastro, G. et al. Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nat Struct Mol Biol* **19**, 1282-6 (2012).

19. Daubner, G.M. et al. Structural and functional implications of the QUA2 domain on RNA recognition by GLD-1. *Nucleic Acids Res* **42**, 8092-105 (2014).
20. Teplova, M. et al. Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites. *Genes Dev* **27**, 928-40 (2013).
21. Liu, Z. et al. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**, 1098-102 (2001).
22. Makino, D.L. et al. RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* **524**, 54-8 (2015).
23. Jencks, W.P. On the attribution and additivity of binding energies. *Proc Natl Acad Sci U S A* **78**, 4046-50 (1981).
24. Butner, K.A. & Kirschner, M.W. Tau protein binds to microtubules through a flexible array of distributed weak sites. *J Cell Biol* **115**, 717-30 (1991).

2962–2973 *Nucleic Acids Research, 2016, Vol. 44, No. 6* *Published online 2 February 2016*
doi: 10.1093/nar/gkw062

The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation

Maxime J. C. Audin, Jan Philip Wurm, Milos A. Cvetkovic and Remco Sprangers*

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

Received December 14, 2015; Revised January 20, 2016; Accepted January 25, 2016

ABSTRACT

The exosome plays an important role in RNA degradation and processing. In archaea, three Rrp41:Rrp42 heterodimers assemble into a barrel like structure that contains a narrow RNA entrance pore and a lumen that contains three active sites. Here, we demonstrate that this quaternary structure of the exosome is important for efficient RNA degradation. We find that the entrance pore of the barrel is required for nM substrate affinity. This strong interaction is crucial for processive substrate degradation and prevents premature release of the RNA from the enzyme. Using methyl TROSY NMR techniques, we establish that the 3′ end of the substrate remains highly flexible inside the lumen. As a result, the RNA jumps between the three active sites that all equally participate in substrate degradation. The RNA jumping rate is, however, much faster than the cleavage rate, indicating that not all active site:substrate encounters result in catalysis. Enzymatic turnover therefore benefits from the confinement of the active sites and substrate in the lumen, which ensures that the RNA is at all times bound to one of the active sites. The evolution of the exosome into a hexameric complex and the optimization of its catalytic efficiency were thus likely co-occurring events.

INTRODUCTION

The exosome is a large molecular machine that plays a role in the processing and degradation of the 3′ end of a large variety of RNA molecules (1). Complexes that belong to the exosome and exosome-like family share the same three-dimensional architecture and are found in all three domains of life. The simplest form of the complex is the bacterial RNase PH that has a 3′ to 5′ exoribonuclease activity (2) (Supplementary Figure S1A). The biological unit of this complex is a homo-hexamer that comprises three RNase PH dimers that assemble into a ring with six active sites (3). During the degradation reaction, the enzyme uses inorganic phosphate to release nucleotide di-phosphates from the 3′ end of the RNA. The second exosome-like complex is the polynucleotide phosphorylase (PNPase) (Supplementary Figure S1B) that is found in bacteria, chloroplasts and mitochondria. The building block of this enzyme contains two consecutive RNase PH domains, a KH and an S1 domain that are linked in one protein chain (4). Six RNase PH domains from three PNPase monomers assemble into a hexameric ring structure that contains three active sites. The exosome complex itself is found in archaea (5) and eukaryotes (6). In archaea, the core of the complex contains the two RNase PH domain proteins Rrp41 and Rrp42 (7). Three Rrp41:Rrp42 dimers assemble into a hexameric ring structure with three active sites (8,9)(Figure 1, Supplementary Figure S1C). The active sites are located in Rrp41, whereas the Rrp42 protein has lost its catalytic activity. The archaeal exosome core recruits three copies of the cap proteins Rrp4 or Csl4 that contain the RNA binding domains (9,10). The interaction of the archaeal exosome core with these cap proteins enhances the RNA degradation rates and provides substrate specificity (11,12). Besides the similar structures of the PNPase, RNase PH and archaeal exosome, these complexes share a similar phosphorolytic mechanism. The eukaryotic exosome (6) has evolved further into a fully asymmetric complex where all protein chains that form the core and all protein chains that form the cap are different (Supplementary Figure S1D and E). In plants (13), only a single subunit in the core (Exo-9) appears to be catalytically active. In other eukaryotes, all exosome subunits are inactive and form a scaffolding complex (14–16). Catalytic activity is added to the Exo-9 complex by the Rrp44 protein that harbors both exoribonucleolytic and endoribonucleolytic activity (16). Interestingly, the catalytic mechanism of the eukaryotic complex moved from phosphorolytic to hydrolytic. The removal of the phosphorolytic catalytic activities in the eukaryotic exosome barrel might have occurred to prevent polymerase activity that could result in non-specific 3′ elongation of RNA (17).

RNase PH (3), PNPase (4), the archaeal and eukaryotic exosome complexes (8,9,15,18) all assemble into barrel-like

*To whom correspondence should be addressed. Tel: +49 7071 601 1330; Fax: +49 7071 601 1308; Email: remco.sprangers@tuebingen.mpg.de

structures with a (pseudo) 3-fold symmetry (Supplementary Figure S1). The active sites are located inside the lumen of these barrels and these are thus secluded from the cellular environment. As a result access to the active sites can be regulated and erroneous RNA degradation can be prevented. For the exosome and exosome-like complexes, RNA-binding domains that are not part of the catalytic RNase PH ring can be used to provide substrate selectivity (19). Based on previous biochemical data and on published crystal structures, it can be concluded that the RNA substrate is threaded to the catalytic chamber through a central pore (the neck region), which is only large enough to accommodate one single-stranded RNA (10) (Figure 1). The functional advantage of substrate selectivity that results from the formation of the quaternary structure is, however, counterbalanced by catalytic disadvantages, as oligomerization of an enzyme into a multimeric complex reduces the number of substrates that can be degraded at the same time. As an example, three isolated Rrp41:Rrp42 dimers will be able to degrade three substrate RNAs simultaneously, whereas a trimer of Rrp41:Rrp42 dimers, as found in the archaeal exosome, is only able to degrade a single substrate at a time. The oligomerization of enzyme complexes, as is seen in the exosome family of exonucleases, is thus a trade-off between a decrease in the number of available active sites per substrate and an increase in substrate selectivity.

Here, we address whether the quaternary structure of the archaeal exosome complex from *Sulfolobus solfataricus* provides catalytic advantages. We focus on two aspects in the exosome complex that arise due to the oligomerization of the enzyme: the creation of the neck region and the establishment of a high local concentration of active sites in the lumen of the barrel. In brief, we combine methyl TROSY NMR, RNA degradation and binding experiments and find that the neck region is essential for the processivity of the enzyme. In addition, we conclude that the sequestering of active sites inside a small lumen of the complex favors RNA degradation as it ensures that the substrate is always in contact with one of the active sites. The formation of a hexameric complex thus provides significant functional advantages for the exosome and exosome-like complexes.

MATERIALS AND METHODS

Protein expression and purification

The genes for the Rrp41 and Rrp42 proteins from *Sulfolobus solfataricus* were cloned into modified pET vectors that carried an N-terminal TEV cleavable His_6-tag. In addition, a construct for the coexpression of both proteins was constructed in a modified pET vector, where only the Rrp41 protein carried an N-terminal TEV cleavable His_6-tag. Point mutations were introduced using standard site-directed mutagenesis methods.

Escherichia coli BL21 codon plus cells were transformed with the appropriate plasmids (Supplementary Table S1). Cells were grown at 37°C and proteins were over-expressed at 25°C by addition of 1 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) when an OD 600 of 0.8 was reached. Twelve hours later, the cells were pelleted by centrifugation and lysed in buffer A (50 mM $NaPO_4$ pH 7.5, 150 mM NaCl, 1 mM DTT) complemented with 10 mM imidazole, lyzozyme and 0.1% Triton X100. The cell lysate was cleared from insoluble debris by centrifugation and the supernatant was loaded on Ni-NTA resin. The resin was washed with buffer A that was complemented with 10 mM imidazole. The protein bound to the resin was eluted with buffer A complemented with 300 mM imidazole. TEV protease was added to the eluted protein and dialyzed overnight into buffer A. To remove the His_6-tagged TEV protease and the cleaved His_6-tag, the dialysate was applied to Ni-NTA resin. The final purification step was performed using size exclusion chromatography on a Superdex 200 column in buffer B (30 mM Hepes pH 6.9, 100 mM NaCl, 1 mM DTT).

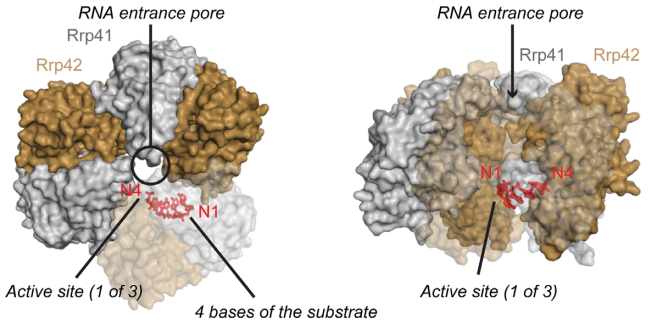


Figure 1. Structure of the *Sulfolobus solfataricus* exosome in complex with a short RNA. (PDB: 2C38) (27). Rrp41 subunits are colored gray, Rrp42 subunits are colored in light brown and the 4 bases of the RNA substrate that are visible in the crystal structure are colored in red. Left: topview of the complex, where the substrate entrance pore is indicated with a circle. Right: sideview of the complex, where the substrate entrance path is indicated with an arrow. One Rrp41:Rrp42 dimer is shown transparent to allow visualization of the inside of the barrel.

The exosome complex was reconstituted from separately purified Rrp42 (after size exclusion chromatography) and Rrp41 (after dialysis and TEV cleavage). Equal amounts of both proteins were mixed and incubated for several hours at room temperature. Uncomplexed Rrp41 or Rrp42 was removed by incubation at 50°C for 2 h. After removal of the precipitated proteins, the sample was further purified using size exclusion chromatography as described above.

Exosome complexes with a different number of active sites (on average) were obtained by mixing catalytically active and catalytically inactive Rrp41 before the addition of Rrp42. The percentage of active Rrp41 was varied between 10 and 100%. After reconstitution, the complexes were purified as described above.

Exosome complexes with exactly one, two or three active sites were obtained with plasmids containing the gene coding for Rrp42 together with three copies of the Rrp41 gene (Supplementary Table S1). These coexpression plasmids were designed as previously described (20). The first copy of Rrp41 contains a His_6-tag, the second one contains a MBP-tag and the third one contains a Strep-tag. For exosomes with a single active site, the His_6- and Strep-tagged versions of Rrp41 were catalytically inactive; for exosomes with two active sites, the His_6-tagged version of Rrp41 was catalytically active; for the exosome with three active sites, all Rrp41 versions were catalytically active (Supplementary Table S1). Cells that coexpressed Rrp42 and the three versions of Rrp41 were grown and induced as described above.

After lysis, the proteins were purified using Ni-affinity chromatography as described above. The protein that eluted from the Ni column contained at least one His_6-tagged version of Rrp41. These complexes were then applied to amylose resin to select for complexes that contained a MBP-tagged version of Rrp41 in addition to a His_6-tagged version of Rrp41. The complexes that were eluted from the amylose resin were subsequently applied to Strep-Tactin resin to select for exosome complexes that contained all three tagged versions of Rrp41. The eluted complex was subsequently treated with TEV protease to remove all affinity tags and dialyzed, prior to performing a size exclusion chromatography as described above. It is worth mentioning that the yield of the exosome complexes purified this way is significantly reduced as only 22% of the exosome complexes that are formed during over-expression contain all three different tags.

Labeling with NMR active nuclei was achieved by over-expression in minimal medium that was based on 100% D_2O. $^{12}C^2H$ glucose was used as the carbon source and methyl labeling was achieved by addition of 100 mg/l U-[^1H,^{13}C] methionine, 60 mg/l 4-methyl $^{13}CH_3$ α-ketobutyric acid (labeled isoleucine precursors) or 100 mg/l methyl $^{13}CH_3$ α-ketoisovaleric acids (labeled valine/leucine precursors) 1 h before induction with 1 mM IPTG.

Coexpression of Rrp41 and Rrp42 was used for the preparation of the complex that contained NMR active methyl groups in both Rrp41 and Rrp42. To that end, both proteins were coexpressed in NMR active growth medium (Supplementary Table S1). Purification was performed as for the single proteins.

RNA *in vitro* transcription and purification

RNA was prepared using *in vitro* transcription with T7 polymerase. The DNA template was obtained from a linearized vector. RNAs used in NMR experiments were transcribed with an HDV-ribozyme that cleaves at the end of the target RNA sequence, resulting in the presence of a 3′ cyclic phosphate. The 3′ cyclic phosphate prevents degradation of the RNA by the exosome and thus allows for long-term NMR measurements. RNAs used in binding and degradation experiments were produced using run-off transcription, where the final RNA contained a 3′ GCT that resulted from the linearization of the template vector with the HindIII restriction enzyme. All RNA constructs contained a hairpin structure (GGCCCCCCCCGAAAGGGGGGGG) followed by 32, 63, 92 or 118 adenines (Supplementary Table S1). The DNA vector containing 63, 92 or 118 adenines were obtained from gene-synthesis (GenScript USA Inc.). *In vitro* transcribed RNA was purified natively with weak ion exchange chromatography using a DEAE-sepharose column as described (21). The pooled fractions were concentrated and buffer exchanged into H_2O with a PD10 column, followed by SpeedVac concentration.

Degradation assays

RNA degradation assays were performed in 180 μl reaction buffer (20 mM Hepes pH 6.5, 60 mM KCl, 0.1 mM EDTA, 2 mM DTT, 8 mM $MgCl_2$, 10 mM Na_2HPO_4) that contained 60 nM exosome (hexameric complex) and 25 μM RNA. The 10 μl samples were taken at different time-points and the reaction was quenched by addition of 10 μl 8M Urea, 20 mM EDTA, 2mM Tris pH 8.

HPLC analysis

Ten microliters of the quenched reaction were automatically injected onto an analytical DNAPac PA100 column (Dionex) that was heated to 80°C. Substrate and product were separated using a linear gradient from buffer A (5 M Urea, 20 mM Tris pH 8, 100 mM NaCl) to buffer B (5 M Urea, 20 mM Tris pH 8, 2 M NaCl) and detected using the absorption at 260 nm. To convert peak intensities to absolute concentrations, the detector response was calibrated by injecting known amounts of RNA (Supplementary Figure S6).

Analysis of degradation data

For each time-point the product concentration was divided by the total concentration [product + substrate] to normalize the signal. The progression of the reaction was then fitted from data at several time-points (Supplementary Figures S4 and S5). Based on the known amounts of enzyme and substrate together with the length of the substrate, the progression curves were translated into number of nucleotides cleaved per second per exosome. To estimate the error in the extracted catalytic rates, we used a jackknife approach, where we fitted the data multiple times after randomly removing a subset of the data.

Fluorescence anisotropy

RNA (GCCCCCCCCGAAAGGGGGGGG-A(21)-4-S-U-A(11)-GCU) for Fluorescence anisotropy measurements was obtained from Dharmacon. The attachment of the 6-(Iodoacetamido)-fluorescein (Sigma-Aldrich) to the thio-uridine (4-S-U) was performed according to the Ramos *et al.* (22). Dilution series of the inactive exosome (2000, 1000, 500, 250, 125, 60, 30, 15, 10 and 0 nM) or of the neck mutant exosome (80, 60, 40, 20, 10, 5, 2, 1, 0.5 and 0 μM) were mixed with 10 nM of RNA labeled with 6-(Iodoacetamido)-fluorescein. For the competition assays, 20 nM of either 32, 63, 98, or 118As RNA was added to the exosome:fluorescent RNA mixture. In all measurements, buffer (30 mM KPO4 pH 6.9, 100 mM NaCl, 0.005% Triton X-100) was used as a reference. Fluorescence anisotropy was recorded every 5 min using a plate reader (Tecan, Infinite F200; filter linear polarization XP38: excitation at 485 nm and emission at 535 nm). Affinity constants were obtained from the data using in-house written scripts using standard equations (23).

NMR

All NMR samples were in buffer B, based on 100% D_2O. NMR spectra were recorded on AVIII-600 and AVIII-800 spectrometers with room temperature probe-heads. Methyl TROSY spectra were recorded at 50°C using a carbon

chemical shift evolution time of 40 ms. SQ (single quantum) dispersion experiments were recorded at 600 and 800 MHz using a relaxation delay of 50 ms and Car-Purcell-Meiboom-Gill (CPGM) frequencies ranging from 40 to 1000 Hz. Relaxation dispersion data were fitted numerically using in-house written scripts using published equations (24). For the final analysis, two residues (I71 and I85), two magnetic fields (600 and 800 MHz) and three temperatures (308, 315 and 323 K) were fitted together to one intrinsic R_2 rate per curve, one exchange rate per temperature and one chemical shift difference per residue. Errors in the extracted parameters were obtained using a Monte Carlo analysis, where the measured data-points were randomly varied around the experimental error. Chemical exchange saturation transfer (CEST) experiments were recorded on an 800 MHz spectrometer at 20°C and using a 400 ms, 5 Hz B_1 field, at 42 different carbon offsets that were spaced by 10 Hz. All NMR spectra were processed using the NMRPipe/NMRDraw software suite (25). Figures displaying NMR spectra and molecular structures were produced using NMRview (onemoonscientific.com) and Pymol (pymol.org), respectively.

RESULTS

The exosome interacts tightly with RNA substrate

The first step in an enzymatic cycle is the formation of an enzyme:substrate complex. Here, we used fluorescence anisotropy measurements to determine the affinity between the exosome and an RNA substrate that contains 32 adenines downstream of a stable GC hairpin structure. To visualize the RNA, we introduced a single 4-thiouridine 15 bases downstream of the 3′ end and coupled this base to 6-(Iodoacetamido)-fluorescein. We then added increasing amounts of a catalytically inactive version of the exosome (D182A in Rrp41). Upon substrate:enzyme complex formation, the rotational lifetime of the RNA is changed, from which we extracted an affinity of 11.7 (0.9) nM (Figure 2A) for the interaction between the exosome and the RNA. This indicates that the exosome interacts tightly with substrates and that substrates can thus be recruited to the complex very efficiently.

We then asked if the interaction between the exosome and RNA depends on the length of the RNA substrate. To that end we performed fluorescence anisotropy competition experiments where we added increasing amounts of non-fluorescently labeled RNA to preformed exosome:fluorescently labeled RNA complex (Figure 2B). As competitors, we used RNA species that contain 32, 63, 92 or 118 adenines downstream of the stable GC hairpin. In these experiments, the fluorescently labeled RNA is competed away from the exosome, which results in a decrease in the fluorescence anisotropy. We then used the program DynaFit (26) to extract the kD for the competitor and found that all RNA species we tested interact with an affinities around 2.7 (0.9) nM (Figure 2B). Note that the competition experiments yield somewhat lower affinities, which is potentially due to minor interference of the fluorescence label with the binding. These data show that the affinity of the enzyme for the substrate is independent of the length of the

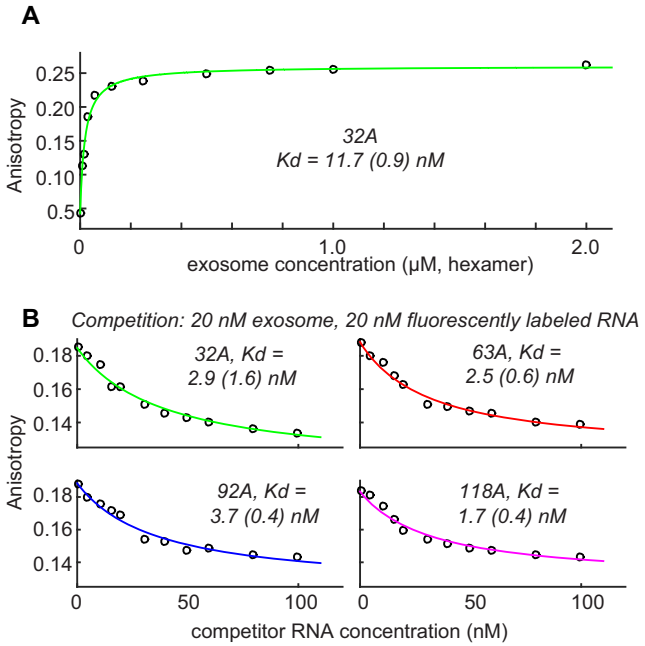


Figure 2. (**A**) Fluorescence anisotropy measurements to determine the affinity between an RNA substrate and the exosome. 32A refers to an RNA that contains a 5′ GC hairpin structure, followed by 32 adenine bases. The extracted error is based on three independent measurements. (**B**) The interaction strength between RNA substrates and the exosome is independent of the length of the RNA. Shown are fluorescence anisotropy measurements, where 20 nM exosome and 20 nM fluorescently labeled 32A RNA (see above), were complemented with increasing amounts of non-fluorescently labeled RNA. The extracted affinities are very similar for RNA species that contain 32, 63, 92 or 118 adenine bases 3′ to a GC hairpin structure. The errors in the extracted parameters result from independent measurements.

the substrate and implies that the 3′ end of the RNA is the prime recognition site for the exosome.

The tight exosome substrate interaction provides processivity

The RNA degradation of the exosome is highly processive, where RNA substrates are not released from the enzyme prior to complete degradation (27). Here, we confirm this processivity and show that an RNA substrate that contains a stable GC hairpin followed by 32 adenines is degraded into an RNA species that contains 10 adenines in addition to the hairpin (Figure 3A). This product results from the fact that the stable hairpin prevents entrance of the substrate into the exosome barrel and from the distance between the entrance pore and the active sites that spans 10 bases (27). During the degradation reaction no intermediate products are observed, indicating that the substrate is not released from the enzyme until degradation has been completed.

To shed light on the interactions that are responsible for this processive degradation we introduced a point mutation in the neck region of the exosome complex (R67G in Rrp41). Interestingly, this mutation that, due to the symmetry of the complex removes three positive charges, caused a reduction in the affinity between the RNA and the exosome from 11.7 nM (Figure 2A) to 27 μM (Figure 3B).

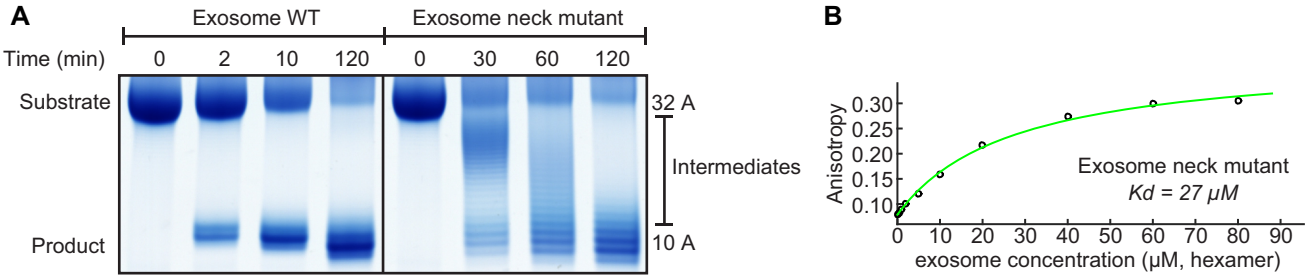


Figure 3. (**A**) Left: Substrate RNA (32A) is processively degraded into product (10A; a GC hairpin with 10 adenine bases) by WT exosome as no intermediate degradation products are detected. Right: a single point mutation in the neck region (Rrp41 R67G) abolishes the processivity as intermediate degradation products appear during the reaction. (**B**) Fluorescence anisotropy binding curve of the neck mutant exosome (Rrp41 R67G) with the fluorescently labeled 32A RNA. The single point mutation results in a 1000-fold reduction of the affinity between the RNA and the enzyme (compare: Figure 2A).

The removal of the positive charge at the entrance pore thus reduces the affinity between the exosome complex and the substrate RNA by three orders of magnitude. At the same time, the processivity of the degradation reaction has been lost as intermediate degradation products appear during the reaction (Figure 3A, right). Over time, these intermediates decrease in length and disappear as they act as substrates in subsequent rounds of degradation by the exosome complex. These results indicate that the neck region of the exosome is responsible for the tight interaction between the enzyme and the substrate and that this tight interaction results in processive RNA degradation.

The RNA is mobile inside the exosome barrel

Based on RNA degradation experiments that use a substrate that contains a stable 5′ hairpin structure, the distance between the neck region of the exosome and the active sites is 10 nucleotides (Figure 3) (27). The exact path of the RNA between the neck region and the active sites has, however, not been revealed. In crystal structures of the *Sulfolobus solfataricus* exosome, the substrate RNA is only visible for the 4 bases (bases N1 to N4; where base N1 is the 3′ terminal base; Figure 1) encompassed in the active site as well as the base N10 in the entrance pore (10,27). Except for these bases, no electron density is observed for the RNA inside the chamber. This suggests that parts of the RNA are not well ordered in the barrel of the exosome.

To obtain additional insights into the path of the substrate RNA within the exosome core, we turned to methyl TROSY NMR spectroscopy (28). This NMR method results in high quality proton-carbon NMR spectra (29) that can be used to identify intermolecular interactions (30). To that end, we recorded methyl TROSY NMR spectra of the exosome in the absence and presence of RNA substrate. Methyl groups that come close to the RNA undergo chemical shift perturbations (CSPs) that can report on the path of the substrate.

First, we prepared an exosome complex that contained NMR invisible Rrp41 and a version of Rrp42 that was NMR active in methyl groups of isoleucine, valine and leucine residues (ILV labeling). We recently obtained resonance assignments for these methyl groups and reported that the RNA interacts with residues around the pore region and with the isoleucine 85 of Rrp42 (31) that is close to the active sites of the enzyme. In case the RNA adopts a static structure with one of the three active sites in the exosome barrel, one would expect that resonances of residues close to the RNA split into two, where 2/3 of the original resonance remains and a novel resonance with intensity 1/3 arises. For residue 85, however, we observe that the resonance intensity is reduced by a factor of ∼8, whereas we failed to identify a novel resonance. This indicates that the RNA substrate is not stably bound to a single active site when it is inside the exosome barrel. To determine if other resonances from Rrp42 also undergo line broadening upon RNA interaction, we divided resonance intensities observed in the free exosome by the resonance intensities of the complex with RNA (Figure 4A). Although most resonance intensities are unaffected by the RNA interaction, we noticed that isoleucine 71 and valine 137 are also significantly weakened in the presence of the RNA. Like isoleucine 85, these residues cluster close to the four nucleotides of the 3′ end of the RNA substrate. The peak broadening of isoleucine 85 is more pronounced because it interacts directly with the RNA, whereas isoleucine 71 and valine 137 are sensing the presence of RNA through conformational changes of the loop underneath the active site (32). Interestingly, the line broadening effect is more prominent at higher temperature (50°C) than at lower temperature (20°C) (Figure 4A). Most likely, this temperature dependence is due to transient interactions between the RNA and the enzyme, where the RNA binding-unbinding causes exchange broadening of the methyl groups in the vicinity. The reduction of the line broadening at lower temperatures can be explained by slower motions of the RNA at 20°C than at 50°C. In summary, our NMR titration data suggest that the RNA is mobile inside the barrel of the exosome.

To identify residues in Rrp41 that are important for the RNA interaction, we prepared an exosome complex that contained NMR active methyl groups for the isoleucine residues of Rrp41 and Rrp42 (Figure 4B). Resonances from Rrp41 can be identified in a straightforward manner based on spectra that were recorded on an exosome that was only labeled in Rrp42 (Supplementary Figure S2). Addition of RNA to the Rrp41:Rrp42 isoleucine labeled sample resulted in a number of CSPs (Figure 4B) where CSPs in Rrp42 confirmed the data that we obtained from the Rrp42 ILV labeled sample. To identify amino acids in Rrp41

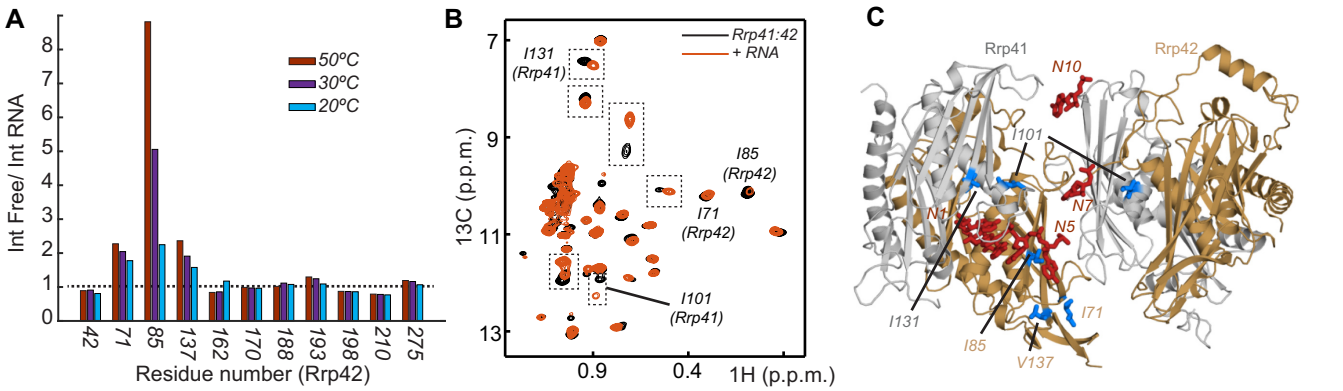


Figure 4. (**A**) The methyl group resonance intensities of Rrp42 resonances in the RNA free exosome divided by the corresponding resonance intensity in the presence of RNA. High bars indicate significant line broadening upon RNA binding, notably observed for residues Ile 71, Ile 85 and Val 137 of Rrp42. The line broadening is shown for three different temperatures, 50, 30 and 20°C. Note that a large number of resonances are not affected by the interaction with RNA (below the dashed bar). (**B**) Methyl TROSY NMR spectrum of the isoleucine region of the exosome that contains NMR active groups in Rrp41 and Rrp42. NMR spectra in the absence (black) and presence (red) of RNA are shown and a number of assignments are indicated. Regions that are highlighted with a dashed box correspond to resonances in Rrp41 that experience large CSPs upon interaction with the RNA substrate. Spectra were recorded at 323 K. (**C**) Structure of the *Sulfolobus solfataricus* exosome (PDB: 2C38) (27) superimposed onto the RNA substrate (in red) visible in the structure of the *Pyrococcus abyssi* exosome (35) (PDB: 2PO1). Assigned residues that show CSPs in the presence of RNA are indicated in blue. Only a single RNA is present per hexameric exosome complex, as the narrow entrance pore does not allow for the recruitment of multiple substrates simultaneously.

that are affected by the RNA, the NMR resonances need to be assigned to the residues in the complex. A full resonance assignment of the Rrp41 methyl groups was, in our hands, not feasible, as Rrp41 in isolation had a high tendency to aggregate. Nevertheless, we assigned a number of Rrp41 methyl group resonances using a mutational approach, where we mutated Rrp41 isoleucine residues into closely related amino acids (33,34). This ideally results in the disappearance of a single isoleucine resonance from the methyl TROSY spectrum. In that way we identified that isoleucines 101 and 131 in Rrp41 interact with the substrate RNA (Figure 4B). Interestingly, these residues shift rather than broaden upon RNA binding. In addition, we noticed that an additional set of unassigned residues in Rrp41 experiences resonance shifts (boxed regions in Figure 4B). Shifting of resonances in NMR titration experiments takes place when the binding–unbinding process is fast on the NMR chemical shift timescale. This suggests that a number of residues in Rrp41, including isoleucine 101 and 131, interact weaker with the RNA than the residues close to the active sites in Rrp42. This observation is in agreement with the lack of electron density for the nucleotides that come close to the Rrp41 protomer.

In comparison with the structure of the *Sulfolobus solfataricus* exosome (27), the structure of the *Pyrococcus abyssi* exosome (35) in complex with RNA shows additional electron density for RNA bases N5 and N7, albeit with low quality. To validate our NMR titration experiments, we superimposed the RNA of the *Pyrococcus abyssi* complex onto the structure of the *Sulfolobus solfataricus* exosome (Figure 4C). This reveals that CSPs that we observe in our NMR titration experiments are in some cases further than 5 Å away from the substrate. This is especially true for isoleucine 101 in Rrp41 that is more than 9 Å away from base N7. We can, however, explain the CSPs of isoleucine 101 in Rrp41 with the mobility of the RNA in the exosome barrel, where structural changes in the RNA result in shorter distances between the substrate and isoleucine 101.

Taken together, our NMR data and the lack of electron density of the RNA substrate (10,27,35) point to a high mobility of the substrate RNA inside the barrel of the exosome.

Quantification of the RNA motions inside the exosome barrel

To directly measure motions of the substrate RNA in the barrel of the exosome, we made use of methyl group relaxation dispersion experiments (36,37). In those experiments, the line broadening that is induced by an exchange process can be quantified and exchange rates can be extracted. We focused our analysis on isoleucine residues of Rrp42 because concentrated samples that are only labeled in Rrp42 can be produced (31). In agreement with the line broadening that we observed in isoleucine 85 and isoleucine 71 (Figure 4A), we detected significant dispersion profiles for these residues in the presence of substrate RNA (Figure 5A). Importantly, these dispersion profiles are solely due to the interaction of the enzyme with the substrate, as they were not observed in the absence of RNA (Supplementary Figure S3). In total, we measured dispersion data at three different temperatures (35, 42 and 50°C) and two magnetic field strengths (600 and 800 MHz; Supplementary Figure S3). To extract the underlying exchange parameters, we fitted all data together and assumed that the chemical shift difference was temperature independent. In addition, we assumed that the excited (RNA bound) state had a population of 1/3, as the substrate RNA can only interact with one of the three active sites at a time. Based on that, we extracted exchange rates of 1021 (163), 1615 (230) and 1744 (249) per second at 35, 42 and 50°C, respectively. The extracted chemical shift differences between the free state and the RNA bound state are 0.25 (0.01) and 0.17 (0.02) p.p.m. for isoleucine 85 and 71, respectively. Isoleucine 71 and isoleucine 85 are located close to the active site of the enzyme and the motions that we detect through those methyl groups thus report on

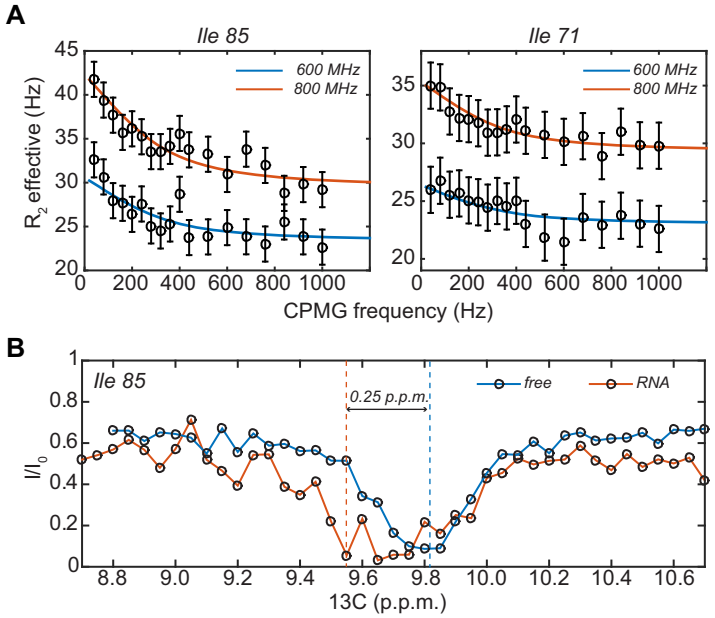


Figure 5. (**A**) Single quantum relaxation dispersion profiles in the presence of RNA. The profiles for isoleucine 71 and isoleucine 85 are shown at two magnetic field strengths and at 50°C. For clarity, the dispersion profiles at 600 MHz are plotted 5 Hz lower than they actually are. The circles indicate the measurement points, where the error bars have been derived from duplicate measurement points. The drawn lines are the best fit to the data, where both residues at three temperatures were fitted simultaneously to one exchange rate and one carbon chemical shift difference per residue. (**B**) CEST profiles for isoleucine 85 in the absence (blue) and presence (red) of the RNA substrate. In the presence of RNA, an invisible state 0.25 downfield of the main resonance of isoleucine 85 appears.

the binding–unbinding of the substrate RNA with the active site region of the exosome. We conclude that the RNA moves from one active site to the next one with a frequency between 1000 and 1700 Hz, depending on the temperature. These data reinforce the notion that the RNA is highly mobile in the lumen of the exosome.

To validate the analysis of the relaxation dispersion experiments, we used CEST experiments (38) to identify the resonance frequency of the invisible RNA-bound state of isoleucine 85. The exchange rates that we determined between 35 and 50°C are too fast for efficient CEST to occur. Therefore, we lowered the experimental temperature to 20°C. Unfortunately, at temperatures lower than 20°C, the signal to noise ratio of the NMR spectra drops significantly, which prevented us from measuring at even lower temperatures. In the CEST experiment, we clearly observe the presence of a second and invisible state for isoleucine 85 (Figure 5B). Interestingly, this invisible state is located 0.25 p.p.m. downfield from the resonance frequency of residue 85. This is in excellent agreement with the chemical shift difference extracted from the relaxation dispersion experiments and confirms that the parameters that we extracted from the relaxation dispersion experiments are accurate and reliable.

The exosome exploits all three active sites during catalysis

Our NMR data show that the 3′ end of the RNA substrate is highly mobile inside the lumen of the exosome. Based on the relaxation dispersion experiments, the substrate interacts with the active sites around 1700 times per second at 50°C. This raises the question whether all three active sites are used or potentially even required during the degradation process. To address this, we used two complementary biochemical experiments where we measured the activity of exosome complexes that contained different numbers of active sites.

First, we reconstituted exosome complexes from separately expressed Rrp42 and Rrp41. To vary the number of active sites in the reconstituted complex, we used different mixtures of catalytically active and catalytically inactive (D182A, that does not interfere with the RNA binding) Rrp41. For example, when a mixture of 40% active and 60% inactive Rrp41 is used in the reconstitution process, statistically the following complexes will form: exosome complexes without any active sites (22%), with one active site (43%), with two active sites (29%) and with three active sites (6%). The relation between the average number of active sites and the activity of the complex depends on the mechanism that is used during the degradation process. In case all active sites equally and independently contribute to the reaction, the activity will linearly increase with the average number of active sites (Figure 6A, red curve). Alternatively, in case only one active site is used in the degradation process (e.g. when the substrate stays on a single active site during the degradation process), the activity will level off with increasing average number of active sites (Figure 6A, blue curve). In case all three active sites are essential for efficient catalysis, the activity of the exosome will only reach high levels when there are on average a high number of active sites (Figure 6A, green curve). Experimentally we can distinguish between these three scenarios by measuring the catalytic activity of exosomes that (on average) contain a different number of active sites. Here, we increased the average number of active sites in a stepwise manner from 10% to 100% in steps of 10%

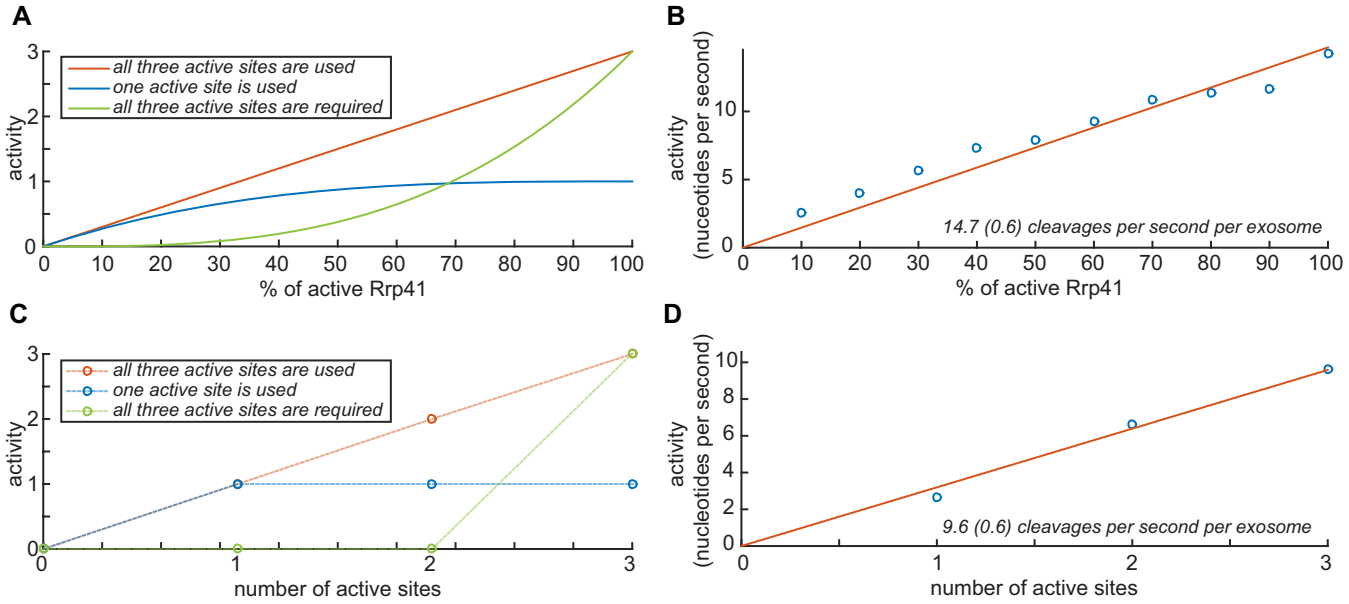


Figure 6. (**A**) Theoretical relationship between the activity and number of active sites inside the exosome complexes, which were reconstituted, with different ratios of active and inactive Rrp41. In case all three active sites in Rrp41 are independently used in the RNA degradation process, the predicted relation between the activity and the percentage of active Rrp41 is shown in red. The blue and green relationships apply to the situation where only one active site is used and where the exosome requires all three sites for full activity, respectively. (**B**) Experimental relationship between the percentage of active Rrp41 in the exosome and the activity. The blue circles indicate the degradation rates (nucleotides cleaved per second per exosome). The red line is the best fit to the data from which the activity of the fully active exosome is determined to be 14.7 cleavages per exosome at 50°C. The data clearly show that all three sites in the exosome are used in the degradation process. (**C**) Theoretical relationship between the activity and the discrete number of active sites inside the exosome. The dashed lines are shown for clarity only. See also (A). (**D**) Experimental relationship between the number of discrete active sites in the exosome complex and the activity. The blue circles refer to the measurements, the red line is the best fit to the data. As in (B), the data show that all three sites in the exosome are used in the degradation process.

(see Methods). Each of these exosome complexes was incubated with substrate RNA and the reaction was quenched at different time-points. In those assays, we choose to use the RNA substrate that contains a 5′ hairpin structure followed by 32 adenine nucleotides. The product of the degradation reaction will then be the hairpin structure with 10 nucleotides, which can be readily detected (see above). Initially, we quantified the levels of substrate and product during the reaction using an Urea PAGE analysis (Supplementary Figure S4), however, we found that quantification using an HPLC approach was more accurate (Supplementary Figure S5) and thus we used this method for all degradation experiments. To obtain degradation rates, we fitted the substrate and product concentrations to a progression curve (Supplementary Figures S4, S5, S6 and methods). In summary, we observe that the number of cleaved nucleotides increases linearly with increasing amounts of active Rrp41 (Figure 6B). These data thus show that all three active sites in the exosome are used during the degradation process. Based on these data, we conclude that one substrate RNA is degraded by three active sites in one exosome. At 50°C, this results in 14.7 (0.6) nucleotide cleavages per second per exosome (Figure 6B).

In a second and independent approach to determine how many active sites in the exosome play a role in the RNA degradation process, we prepared three different samples with exosome complexes that contained exactly one, two or three active sites. These complexes were obtained by overexpression of untagged Rrp42 with three copies of Rrp41, each of them fused to a different affinity tag (His-tag, MBP-tag or Strep-tag). Importantly, we were not able to detect any subunit exchange between different exosome complexes (Supplementary Figure S7), establishing that the exosome complexes with a discrete number of active sites are extremely stable. Subsequently, three consecutive purification steps were used to obtain complexes that contained exactly one of each Rrp41 affinity tags. During expression of the exosome complex, we used one, two or three catalytically active versions of the differently tagged Rrp41 proteins and were thus able to prepare exosome complexes with a discrete number of active sites. As described before, the activity of these complexes can provide information on the mechanism that is used during RNA degradation (Figure 6C, see above). We then experimentally determined the activity of the exosome complexes that harbored a discrete number of active sites and found that all three active sites are equally involved during the degradation process (Figure 6D). This confirms the experiments that we performed using exosomes with mixed number of active sites. The overall activity of the exosome complexes with a discrete number of active sites appears somewhat lower (9.6 cleavages per second per exosome at 50°C) than the activity of the exosome complexes that contain mixture of active and inactive sites. This is most likely due to a loss in activity of the exosome complex during the long purification protocol that is required for the preparation of the exosome complexes with a discrete number of active sites.

The RNA motions are much faster than the enzymatic turnover rates

Our degradation assays show that the RNA substrate uses all three active sites in the exosome barrel. This is in agreement with the NMR data that show that the substrate rapidly exchanges between the active sites. Interestingly, at 50°C the rate of exchange (1700 s^{-1}) is two orders of magnitude faster than the number of cleavages per second per exosome (~10 s^{-1}). It is worth noting that these differences are not a peculiarity of this temperature. Indeed, we compared degradation experiments at 20, 35, 50 and 65°C (Figure 7A) and found that the temperature dependence of the degradation rate follows the Eyring relationship (Figure 7B). The temperature dependence of the RNA hopping frequency in the exosome barrel, as we determined using relaxation dispersion experiments, follows a very similar trend, albeit at much higher frequencies. This indicates that the RNA hopping frequency is significantly larger than the degradation rates for biologically relevant temperatures.

From the temperature dependence of the activity of the exosome core (Figure 7B) the RNA degradation rate at 75°C (the optimal growth temperature of *Sulfolobus solfataricus*) can be predicted to be around 65 nucleotides/second. This degradation rate is comparable to the RNA degradation rate that was previously determined for the PNPase enzyme (120 nucleotides/second) (39). In addition, the rate of RNA degradation by the exosome core is in the same order of magnitude as the elongation rate of the archaeal RNA polymerases that was determined to be around 20 nucleotides per second in *Methanothermobacter thermautotrophicus* (40).

The degradation rate is independent on the RNA length

The catalytic cycle of the exosome includes multiple steps. If substrate binding or product release is very slow, the exosome would require a significant time between finishing the degradation of one substrate and initiating the degradation of the next one. Such a time would reduce the average cleavage frequency that we measured in our biochemical experiments. To probe whether substrate binding or product release are rate limiting in the catalytic cycle of the exosome we performed degradation assays with RNA substrates of increasing length (32, 63, 92 and 118 adenine repeats, that all interact with the exosome with similar affinities, Figure 2B). For short RNA substrates, the exosome would need to reload the substrate significantly more often than for long substrates, which would reduce the overall turnover rate. In our experiments, we designed the substrates such that short (less than ~10 nucleotides) single stranded RNA stretches are not formed (Supplementary Figure S4A), as these have been shown to be degraded at very low rates (41). Interestingly, in our degradation experiments, we find that the activity of the exosome (number of nucleotides cleaved per second) is very similar for the four RNAs used, showing that the activity of the exosome is largely independent of the length of the substrate (Figure 7C). This shows that substrate binding and product release are not significantly limiting catalytic turnover. The nucleotide cleavage frequencies that we measured thus directly report on the activity of the active sites. This validates our conclusion that the RNA jumping frequencies are much faster than the catalytic cleavage rates.

DISCUSSION

Self-compartmentalization is a principle that is exploited not only by the exosome, but also by other enzymes, including the proteasome (42). The sequestering of active sites inside a small space prevents degradation of substrates that are not actively targeted to the enzyme. For the archaeal exosome, substrate selection takes place through RNA interacting proteins that dock around the entrance pore of the barrel (9,10,19). In the eukaryotic exosome, these RNA-binding proteins have evolved further and include helicases that are able to unfold RNA species that contain secondary structure (43).

Despite the functional advantages that are related to substrate selection, self-compartmentalization comes at a cost: multiple protomers in a larger complex can only act on a single substrate whereas the same number of monomeric proteins could act on multiple substrates simultaneously. For the *Archaeoglobus fulgidus* exosome, the activity of the native $(Rrp41:Rrp42)_3$ exosome complex was previously compared to the activity of a version of the complex that only assembles into $(Rrp41:Rrp42)_1$ dimers (41). Interestingly, for long RNA substrates, it was found that the catalytic activity of one $(Rrp41:Rrp42)_3$ hexameric complex is higher than that of three $(Rrp41:Rrp42)_1$ dimers. One $(Rrp41:Rrp42)_3$ exosome that acts on one substrate is thus more efficient than three $(Rrp41:Rrp42)_1$ dimers that act on three substrates simultaneously. For the *Archaeoglobus fulgidus* exosome complex, the advantages of self-compartmentalization thus outweigh the disadvantages. The molecular basis that underlies this gain in activity upon assembly of the Rrp41 and Rrp42 protomers into a barrel like quaternary structure remained undetermined.

Here, we addressed the catalytic advantages of oligomerization of the *Sulfolobus solfataricus* exosome. Two features that appear in the enzyme upon oligomerization of the Rrp41 and Rrp42 proteins into a hexameric barrel are the entrance pore (the neck region) and a lumen that contains a very high concentration of active sites. We find that these two aspects are fundamental to the efficiency of the exosome complex (Figure 8).

Using binding measurements, we show that the neck region, where the RNA enters the exosome lumen, strongly interacts with unstructured RNA. Interestingly, this interaction involves substrate nucleotides that are located 10 bases upstream of the degradation site. This very strong interaction ensures that substrates can be recruited efficiently, and at the same time prevents that substrates are released from the enzyme complex before complete degradation (Figures 2 and 3). Experimentally we have shown this using an exosome complex that contains a single point mutation in the neck region. This mutation results in a 1000-fold decrease of the enzyme substrate interaction strength (Figure 3B). Importantly, also the processivity that is observed in the WT complex is lost upon weakening the neck-substrate interactions (Figure 3A) as substrates can no longer be retained to the enzyme complex during turnover. The importance of the neck region for RNA degradation appears to be

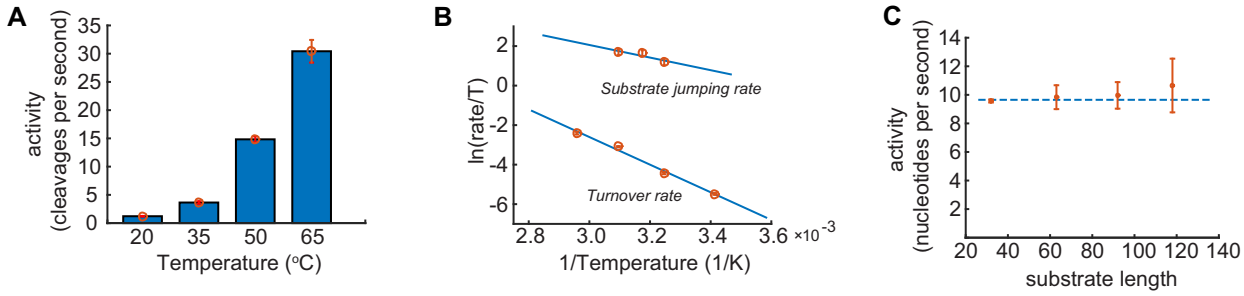


Figure 7. (**A**) Effect of the temperature on the activity of the exosome. (**B**) Eyring analysis of the temperature dependence of the degradation rates (lower line) and NMR jumping rates (upper line). The lines follow the same trend, indicating that the substrate jumping rate is much faster than the degradation rate at all biologically relevant temperatures. (**C**) The number of nucleotides cleaved by the exosome is independent of the length of the RNA. This shows that substrate recruitments or product release are not rate limiting in the reaction. The larger errors for the 118A RNA substrates might result from some inhomogeneity in the exact length of the substrate that resulted from slippage or stalling of the T7 polymerase that was used to prepare this long and highly repetitive sequence.

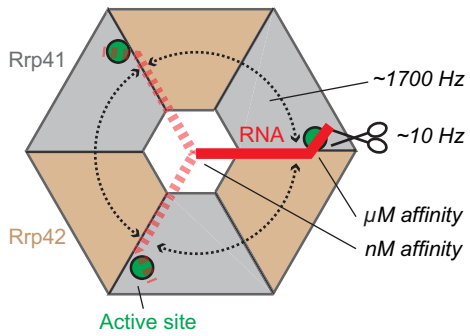


Figure 8. Cartoon that summarizes our findings. The RNA (red, dashed red) enters the exosome barrel (gray, sand) through the neck and can interact with one of three active sites (green circles). The substrate jumps between these three active sites with frequencies of ∼1700 s^{-1} (rounded dashed arrows). The catalytic cleavage rates are in the order of 10 s^{-1} (scissor symbol), two orders of magnitude slower than the RNA jumping rate. Only 1 in 100 RNA:exosome encounters results in a catalytic reaction. The exosome:RNA nM affinity interactions through the neck region keep the 3′ end of the RNA in the exosome barrel. The high concentration of active sites and substrate in the barrel of the exosome ensures that the 3′ end of the RNA is always in contact with one of the active sites.

conserved in exosome and exosome-like complexes (Supplementary Figure S1). First, it has been shown that the hexameric structure of the RNAse PH complex is essential for activity as a dimeric form of the enzyme that does not form a neck region is unable to interact with RNA (44). Secondly, mutations of basic residues in the entrance channel of the PNPase complex result in a loss of activity and processivity (45). Finally, mutations within the Rrp41 neck region of the yeast exosome have been shown to be important for the channeling of substrates through the catalytically inactive Exo-9 complex toward the catalytically active Rrp44 subunit (18).

To address the importance of the high concentration of active sites in the lumen of the exosome, we determined the kinetics of the substrate in the proximity of the active sites. Using methyl TROSY NMR methods we show that the 3′ end of the RNA is highly mobile and that it jumps between the active sites with a frequency of around 1700 Hz at 50°C (Figure 5). This high mobility will allow for a rapid dissociation of the reaction products, thereby facilitating turnover. Based on fluorescence anisotropy experiments that use a mutant where the neck interactions are impaired, the affinity between the active sites and the 3′ end of the RNA is in the μM regime (Figure 3B). This affinity is in line with previous ITC measurements that used a very small RNA substrate (46). Based on this μM affinity and the determined exchange rate of the RNA in the barrel it can be concluded that the on-rate of the binding process is fast (in the order of 10^9 $M^{-1}s^{-1}$). This fast on-rate is probably a direct consequence of the small volume of the lumen of the exosome and the resulting high substrate concentration. Using a series of RNA degradation experiments (Figures 6 and 7) we show that the motions of the RNA allow that all active sites in the exosome lumen equally participate in RNA degradation. Interestingly, the number of nucleotide cleavage events (∼10 s^{-1}) is two orders of magnitude lower than the number of active-site: substrate encounters (∼1700 s^{-1}). This shows that around 100 encounter complex formation events are required for one cleavage event.

When one assumes that the exosome has a spherical lumen with a diameter of 32 Å, the concentration of active sites in the enclosed volume can be estimated to be around 300 mM whereas the local substrate concentration is one-third of that (one RNA can enter the lumen). Based on the μM affinity of the 3′ end of the RNA for the active sites, the active-site occupancy can be predicted to be essentially 100%. As soon as RNA enters the lumen of the exosome it will thus be bound to one of the three active sites. In that light, the oligomerization of the exosome enhances catalytic efficiency by ensuring that the enzyme interacts with the substrate in a highly efficient manner at all times. In the theoretical case, where the archaeal exosome would only contain a single active site, the RNA would still be in contact with this site for more than 98% of the time. The reduction of the number of active sites in exosome(-like) complexes (6 active sites in RNase PH, 3 active sites in the PNPase and archaeal exosome, one in the plant exosome; Supplementary Figure S1) does not result in a reduction of the catalytic activity; the RNA will always be in full contact with an active site. After oligomerization of the exosome-like complexes into a barrel-like quaternary structure, the diversification of the subunits and the removal of active sites thus posed no catalytic disadvantages. It should

be noted that in our degradation experiments (Figure 6) we intentionally removed activity through a single-point mutation that did not alter the interaction with RNA. In that situation, the 3′ end of the RNA partitions between active and inactive sites, which leads to the observed reduction in activity. When the removal of active sites is accompanied by the removal of the substrate interaction, the activity should stay at its maximum, independent of the number of active sites. In that light, it should be noted that the eukaryotic exosome indeed makes no contacts with the substrate RNA as all contact points have been removed (47).

In summary, we show that the oligomerizaiton of the exosome complex into a barrel like structure provides novel catalytic advantages. The basis for this lies in the fact that the enzyme interacts very strongly with the substrate close to base N10, whereas the 3′ end of the RNA close to base N1 remains highly flexible (Figure 8). The evolution of the hexameric exosome complex from single protein chains might have benefited from the increase in catalytic efficiency that is associated with the formation of the quaternary structure.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We acknowledge all members of the laboratory for discussions. We thank Silke Wiesner for helpful suggestions and Lewis E. Kay (University of Toronto) for sharing the code for the CEST NMR experiment with us.

FUNDING

M.A.C. acknowledges funding from the IMPRS ‘From Molecules to Organisms’. This work was supported by the Max Planck Society and the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013), ERC grant agreement no. 616052.
Conflict of interest statement. None declared.

REFERENCES

1. Houseley,J., LaCava,J. and Tollervey,D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.
2. Deutscher,M.P., Marshall,G.T. and Cudny,H. (1988) RNase PH: an Escherichia coli phosphate-dependent nuclease distinct from polynucleotide phosphorylase. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 4710–4714.
3. Ishii,R., Nureki,O. and Yokoyama,S. (2003) Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus. *J. Biol. Chem.*, **278**, 32397–32404.
4. Symmons,M.F., Jones,G.H. and Luisi,B.F. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, **8**, 1215–1226.
5. Koonin,E.V., Wolf,Y.I. and Aravind,L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.*, **11**, 240–252.
6. Mitchell,P., Petfalski,E., Shevchenko,A., Mann,M. and Tollervey,D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′–>5′ exoribonucleases. *Cell*, **91**, 457–466.
7. Evguenieva-Hackenberg,E., Walter,P., Hochleitner,E., Lottspeich,F. and Klug,G. (2003) An exosome-like complex in Sulfolobus solfataricus. *EMBO Rep.*, **4**, 889–893.
8. Lorentzen,E., Walter,P., Fribourg,S., Evguenieva-Hackenberg,E., Klug,G. and Conti,E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.*, **12**, 575–581.
9. Buttner,K., Wenig,K. and Hopfner,K.P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. *Mol. Cell*, **20**, 461–471.
10. Lorentzen,E., Dziembowski,A., Lindner,D., Seraphin,B. and Conti,E. (2007) RNA channelling by the archaeal exosome. *EMBO Rep.*, **8**, 470–476.
11. Evguenieva-Hackenberg,E., Roppelt,V., Finsterseifer,P. and Klug,G. (2008) Rrp4 and Csl4 are needed for efficient degradation but not for polyadenylation of synthetic and natural RNA by the archaeal exosome. *Biochemistry*, **47**, 13158–13168.
12. Roppelt,V., Klug,G. and Evguenieva-Hackenberg,E. (2010) The evolutionarily conserved subunits Rrp4 and Csl4 confer different substrate specificities to the archaeal exosome. *FEBS Lett.*, **584**, 2931–2936.
13. Chekanova,J.A., Shaw,R.J., Wills,M.A. and Belostotsky,D.A. (2000) Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells. *J. Biol. Chem.*, **275**, 33158–33166.
14. Januszyk,K. and Lima,C.D. (2014) The eukaryotic RNA exosome. *Curr. Opin. Struct. Biol.*, **24**, 132–140.
15. Liu,Q., Greimann,J.C. and Lima,C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, **127**, 1223–1237.
16. Dziembowski,A., Lorentzen,E., Conti,E. and Seraphin,B. (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat. Struct. Mol. Biol.*, **14**, 15–22.
17. Wahle,E. (2007) Wrong PH for RNA degradation. *Nat. Struct. Mol. Biol.*, **14**, 5–7.
18. Bonneau,F., Basquin,J., Ebert,J., Lorentzen,E. and Conti,E. (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell*, **139**, 547–559.
19. Walter,P., Klein,F., Lorentzen,E., Ilchmann,A., Klug,G. and Evguenieva-Hackenberg,E. (2006) Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol. Microbiol.*, **62**, 1076–1089.
20. Mund,M., Overbeck,J.H., Ullmann,J. and Sprangers,R. (2013) LEGO-NMR spectroscopy: a method to visualize individual subunits in large heteromeric complexes. *Angew. Chem. Int. Ed. Engl.* , **52**, 11401–11405.
21. Easton,L.E., Shibata,Y. and Lukavsky,P.J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA*, **16**, 647–653.
22. Ramos,A. and Varani,G. (1998) A new method to detect long-range protein-RNA contacts: NMR detection of electron-proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.*, **120**, 10992–10993.
23. Johnson,P.E., Tomme,P., Joshi,M.D. and McIntosh,L.P. (1996) Interaction of soluble cellooligosaccharides with the N-terminal cellulose-binding domain of Cellulomonas fimi CenC 2. NMR and ultraviolet absorption spectroscopy. *Biochemistry*, **35**, 13895–13906.
24. Korzhnev,D.M., Kloiber,K. and Kay,L.E. (2004) Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: theory and application. *J. Am. Chem. Soc.*, **126**, 7320–7329.
25. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
26. Kuzmic,P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**, 260–273.
27. Lorentzen,E. and Conti,E. (2005) Structural basis of 3′ end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. *Mol. Cell*, **20**, 473–481.
28. Tugarinov,V., Hwang,P.M., Ollerenshaw,J.E. and Kay,L.E. (2003) Cross-correlated relaxation enhanced 1H[bond]13C NMR

spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J. Am. Chem. Soc.*, **125**, 10420–10428.
29. Sprangers,R. and Kay,L.E. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*, **445**, 618–622.
30. Wiesner,S. and Sprangers,R. (2015) Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.*, **35**, 60–67.
31. Audin,M.J., Dorn,G., Fromm,S.A., Reiss,K., Schutz,S., Vorlander,M.K. and Sprangers,R. (2013) The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding. *Angew. Chem. Int. Ed. Engl.*, **52**, 8312–8316.
32. Lorentzen,E. and Conti,E. (2012) Crystal structure of a 9-subunit archaeal exosome in pre-catalytic states of the phosphorolytic reaction. *Archaea*, 721869.
33. Sprangers,R., Gribun,A., Hwang,P.M., Houry,W.A. and Kay,L.E. (2005) Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 16678–16683.
34. Amero,C., Asuncion Dura,M., Noirclerc-Savoye,M., Perollier,A., Gallet,B., Plevin,M.J., Vernet,T., Franzetti,B. and Boisbouvier,J. (2011) A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J. Biomol. NMR*, **50**, 229–236.
35. Navarro,M.V., Oliveira,C.C., Zanchin,N.I. and Guimaraes,B.G. (2008) Insights into the mechanism of progressive RNA degradation by the archaeal exosome. *J. Biol. Chem.*, **283**, 14120–14131.
36. Korzhnev,D.M., Kloiber,K., Kanelis,V., Tugarinov,V. and Kay,L.E. (2004) Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J. Am. Chem. Soc.*, **126**, 3964–3973.
37. Lundstrom,P., Vallurupalli,P., Religa,T.L., Dahlquist,F.W. and Kay,L.E. (2007) A single-quantum methyl 13C-relaxation dispersion experiment with improved sensitivity. *J. Biomol. NMR*, **38**, 79–88.
38. Bouvignies,G. and Kay,L.E. (2012) A 2D (1)(3)C-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding. *J. Biomol. NMR*, **53**, 303–310.
39. Fazal,F.M., Koslover,D.J., Luisi,B.F. and Block,S.M. (2015) Direct observation of processive exoribonuclease motion using optical tweezers. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 15101–15106.
40. Xie,Y. and Reeve,J.N. (2004) Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome. *J. Bacteriol.*, **186**, 3492–3498.
41. Hartung,S., Niederberger,T., Hartung,M., Tresch,A. and Hopfner,K.P. (2010) Quantitative analysis of processive RNA degradation by the archaeal RNA exosome. *Nucleic Acids Res.*, **38**, 5166–5176.
42. Makino,D.L., Halbach,F. and Conti,E. (2013) The RNA exosome and proteasome: common principles of degradation control. *Nat. Rev. Mol. Cell Biol.*, **14**, 654–660.
43. Halbach,F., Reichelt,P., Rode,M. and Conti,E. (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, **154**, 814–826.
44. Choi,J.M., Park,E.Y., Kim,J.H., Chang,S.K. and Cho,Y. (2004) Probing the functional importance of the hexameric ring structure of RNase PH. *J. Biol. Chem.*, **279**, 755–764.
45. Shi,Z., Yang,W.Z., Lin-Chao,S., Chak,K.F. and Yuan,H.S. (2008) Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA*, **14**, 2361–2371.
46. Oddone,A., Lorentzen,E., Basquin,J., Gasch,A., Rybin,V., Conti,E. and Sattler,M. (2007) Structural and biochemical characterization of the yeast exosome component Rrp40. *EMBO Rep.*, **8**, 63–69.
47. Makino,D.L., Schuch,B., Stegmann,E., Baumgartner,M., Basquin,C. and Conti,E. (2015) RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature*, **524**, 54–58.

The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation

Maxime J. Audin, Jan Philip Wurm, Milos A. Cvetkovic and Remco Sprangers*

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

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

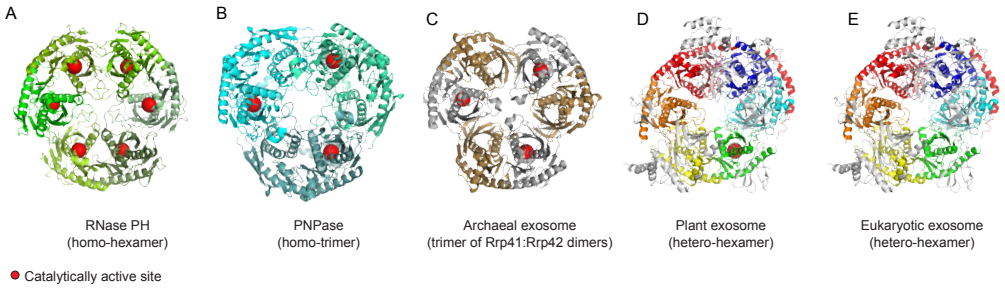


Figure S1. Structures of the exosome and exosome-like complexes.

(A) The homo-hexameric RNase PH (PDB: 1UDN) (1) that harbors six active sites (red spheres). The six identical protein chains are colored in different shades of green.

(B) The home-trimeric PNPase (PDB: 1E3H) (2) that harbors three active sites (red spheres). The three identical protein chains are colored in different shades of cyan.

(C) The archaeal exosome complex (PDB: 2BR2) (3) that contains three Rrp41 (grey) Rrp42 (sand) dimers. Only Rrp41 contains a catalytically active site.

(D) The core of the eukaryotic exosome structure (PDB: 2NN6) (4) that contains six different protein chains. The plant exosome appears to have a single catalytic active site in Rrp41. The three different proteins that form the cap of the complex are colored grey.

(E) As in (D), eukaryotic exosomes from other species have lost all catalytically active sites and act as a scaffolding complex.

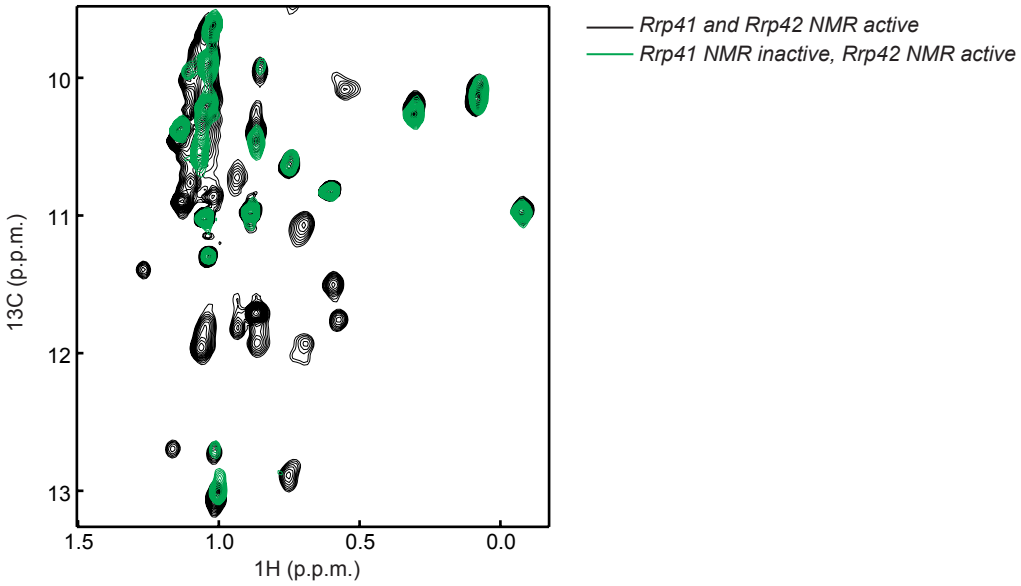


Figure S2.

Overlay of methyl TROSY NMR spectra of the exosome complex that contains NMR active isoleucine residues in both Rrp41 and Rrp42 (black) or only in Rrp42 (green). Resonances that result from residues in Rrp41 can be readlily identified.

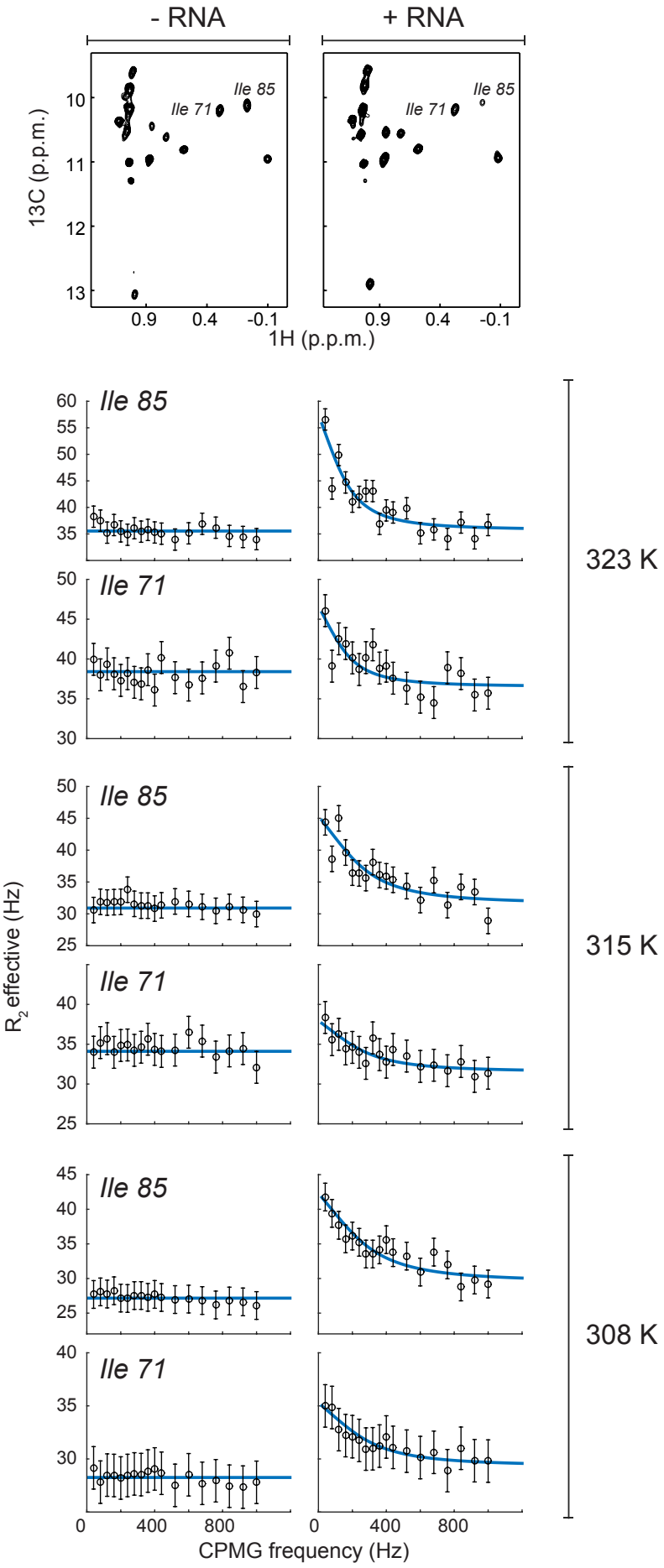

- RNA
+ RNA
Ile 85
Ile 71
13C (p.p.m.)
1H (p.p.m.)
323 K
315 K
308 K
R2 effective (Hz)
CPMG frequency (Hz)

Figure S3.

Top part: methyl TROSY NMR spectra of Rrp42 in the exosome complex in the absence (left) and presence (right) of RNA substrate. Bottom part: Isoleucine 71 and 85 show flat dispersion profiles in the absence of RNA (left column). In the presence of substrate RNA, these residues show motion on the ms timescale. The motions that we observe in the exosome:RNA complex are thus a direct result of the RNA. The shown profiles have been recorded at 800 MHz.

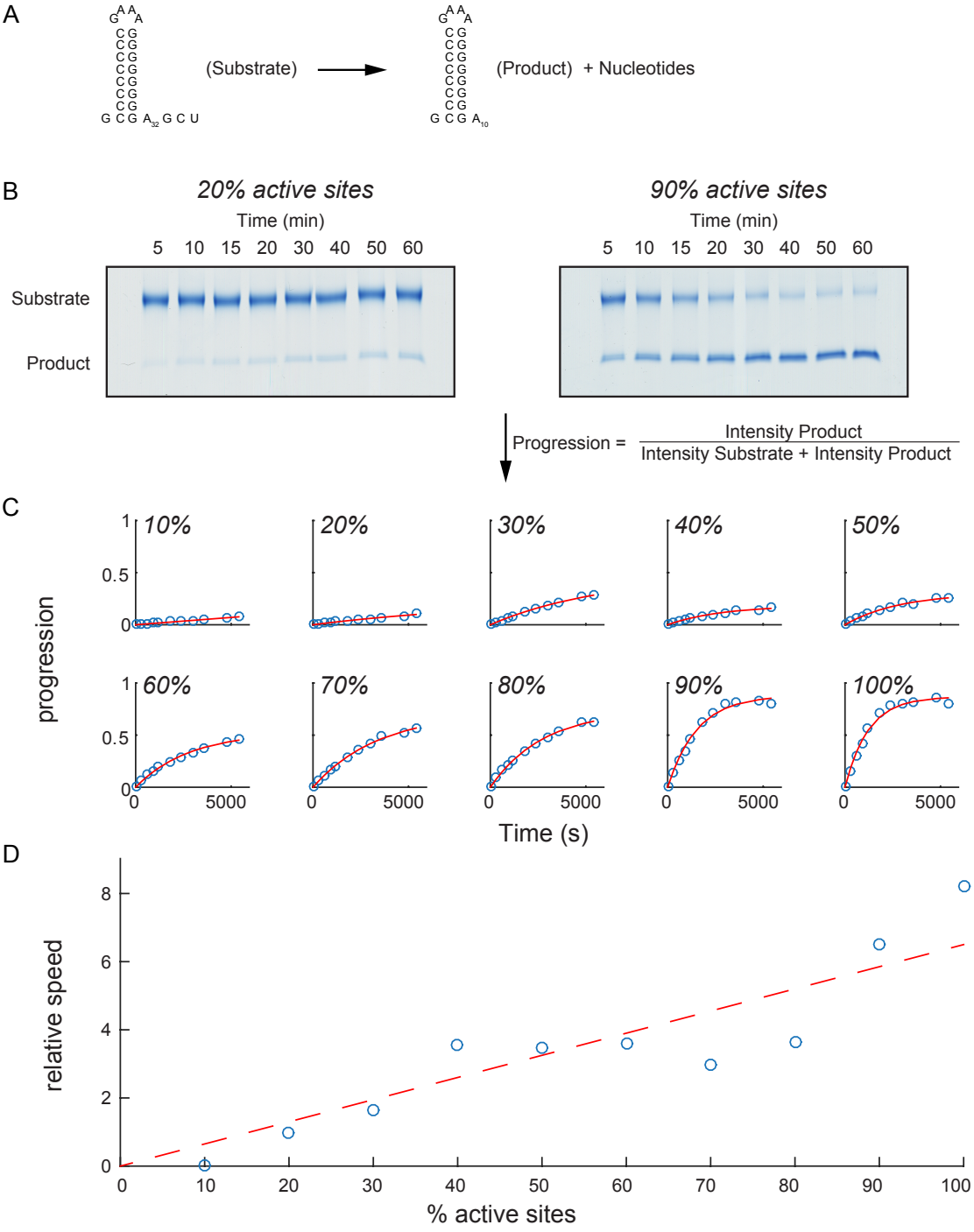


Figure S4.

(A) In our experiments the RNA substrate (a GC hairpin followed by 32 adenines, where the 3' GCU nucleotides result from the linearization of the DNA template used for *in vitro* transcription) is processively degraded into the product RNA (a GC hairpin followed by 10 adenines) and nucleotides.

(B) 30 μM substrate (see A) was incubated with 70 nM exosome (that contained an average number of active sites between 10 and 100%) at 50°C. Aliquots of the reaction were taken at different time points and mixed with 8M urea to quench the reaction. The samples of the different time-points were analyzed using Urea PAGE, after which the RNA was visualized

with methylene blue. The amount of substrate and product was quantified using ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2015). The Urea PAGE band intensities were scaled by the length of the RNA to correct for differences in staining efficiency of the substrate and product. Two exemplary gels of the reaction in the presence of the exosome that contains on average 20% (left) or 90% (right) active sites are shown. The time-points at which the reaction was quenched are indicated on top.

(C) The progression of the reaction was determined at each time point by dividing the band corrected band intensity of the product by the sum of the corrected band intensities of the product and the substrate. The rate of the reaction was subsequently determined by fitting the progression of the reaction to the exponential function B*(1-exp(-A*t)), where A is the rate of the degradation reaction and B is a scaling factor.

(D) The relative speeds of the reaction were plotted versus the average number of active sites to determine if the reaction linearly depends on the number of active sites (See Fig 6). The determined catalytic rates show larger spreads, indicative for inaccuracies in the extracted parameters. These inaccuracies can have multiple causes, including inaccuracies in the determination of Urea PAGE band intensities, different efficiencies in the methylene blue staining of the substrate and product, pipetting errors during the loading of the gel and variations in the quality of different urea-page gels.

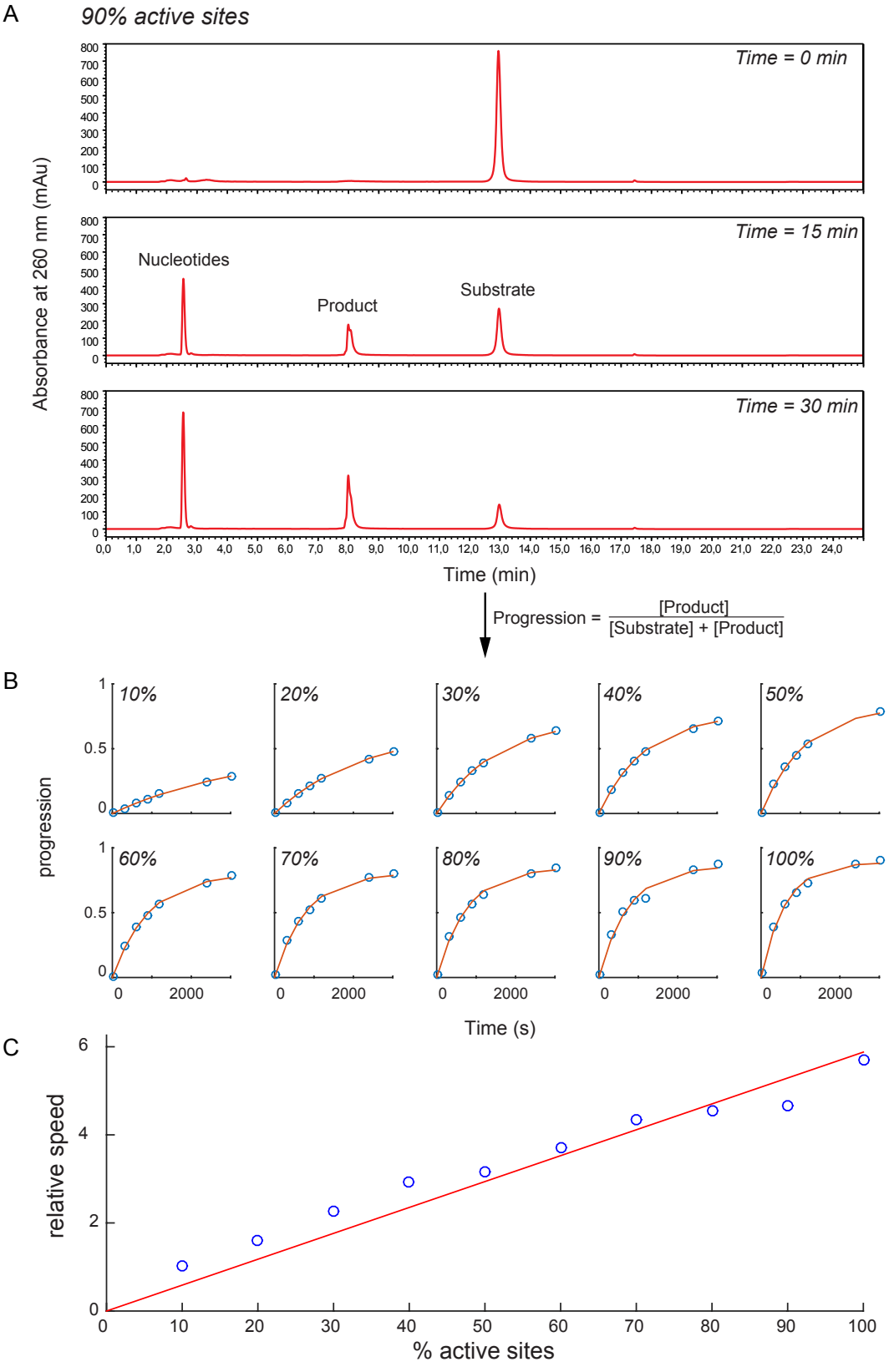


Figure S5.

(A) Exosome and substrate were mixed as described in Figure S4A and S4B, but using 25 μM substrate RNA and 60 nM exosome. The reaction time-points were analyzed on a Dionex PA-100 column that was heated to 80 °C and coupled to an HPLC system that was equipped

with an auto sampler for the precise and reproducible injection of the sample onto the column. This allowed for the accurate quantification of the levels of substrate and product in the reaction mixture, where the peak areas were converted into absolute concentrations based on determined calibration curves (Figure S6). Note that we did not use the concentration of the nucleotides in the analysis as the peak that contained the nucleotides overlapped with small peaks that resulted from the buffer in the reaction mixture (see e.g. top chromatogram taken at time=0 and where no product and nucleotides are present. The progression of the reaction was assessed by dividing the product concentration by the total RNA concentration.

(B) As in Fig S4C, the rate of the reaction was determined at each time point by fitting the progression of the reaction to the exponential function B*(1-exp(-A*t)), see Fig S4. Note that each data point in panel B is extracted from a single HPLC run.

(C) Plot that correlates the activity of the exosome complex versus the average number of active sites in the complex. The spread in the data is significantly reduced compared to the Urea PAGE analysis (See Fig S4). Based on that we decided to analyze all degradation experiments using the HPLC method.

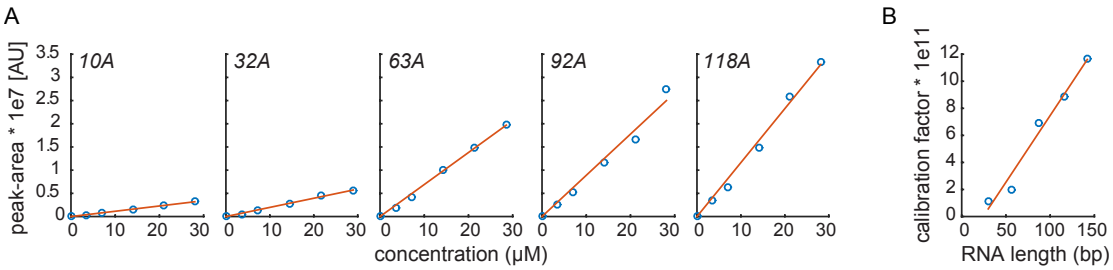


Figure S6.

(A) Calibration of the detector response using known amounts of product (a GC hairpin followed by 10 adenines) and substrates (a GC hairpin followed by 32, 63, 92 or 118 adenines).

(B) Global calibration of the product and substrate amounts that was used to convert all HPLC peak areas into absolute concentrations.

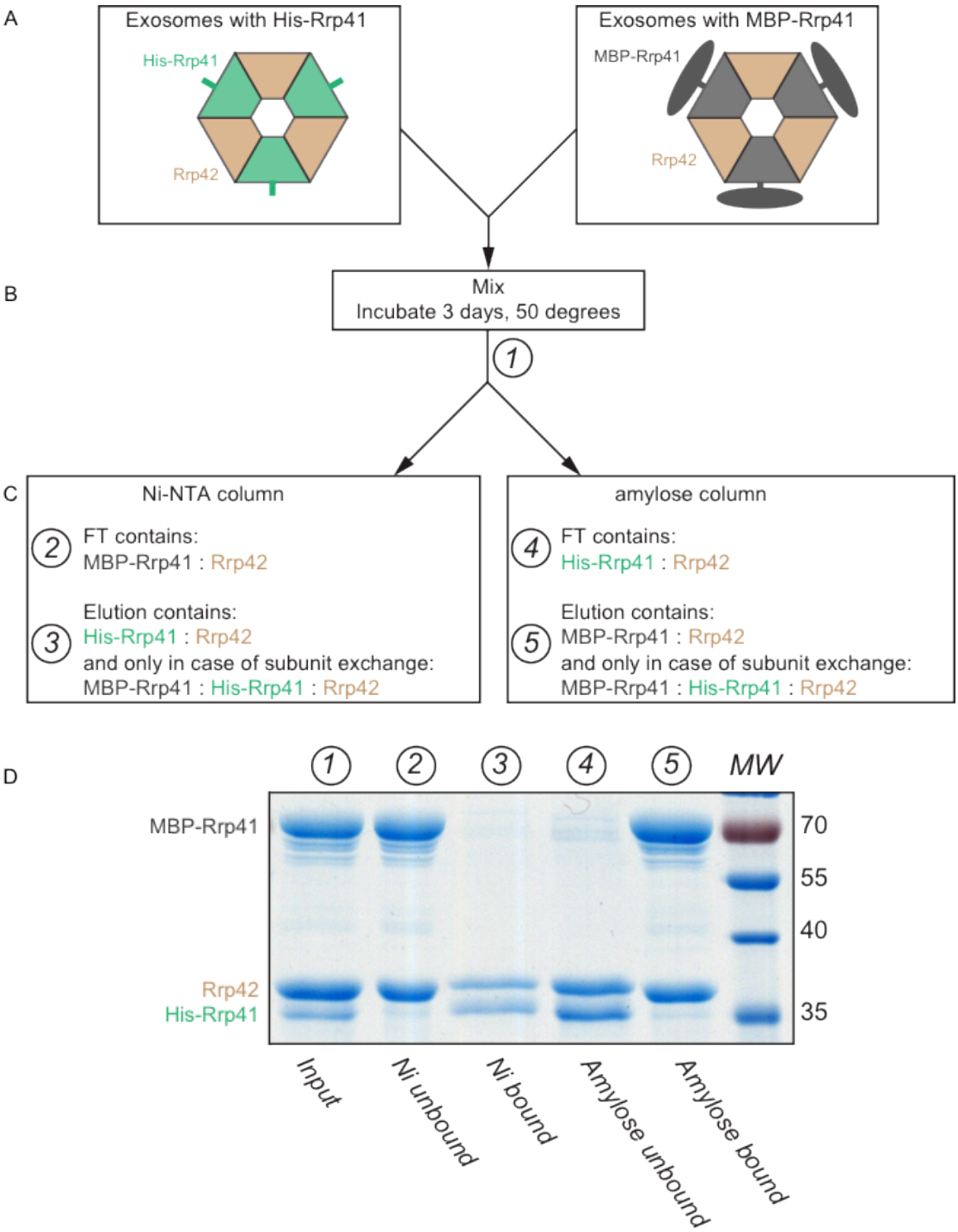


Figure S7.

(A) Exosome complexes that contain only a His-tagged version of Rrp41 (left, green) and exosome complexes that only contain an MBP-tagged version of Rrp41 (right, dark grey) were separately prepared.

(B) The His- and MBP-tagged version of the exosome complex were mixed in a 1:1 ratio and incubated at 50°C for 3 days. In case of subunit exchange between the His and MPB tagged exosome complexes a MBP-Rrp41:His-Rrp41:Rr42 complex would form.

(C) The sample was split into two parts. One part was applied to Ni-NTA resin. Exosome complexes that don't contain any His-Rrp41 subunits (only MBP-Rrp41:Rrp42) will not bind to

the resin. The elution of the Ni-NTA column contains all complexes that contain one or more histidine tags and this fraction thus contains His-Rrp41:Rrp42 complexes. In addition complexes that result from subunit exchange and that thus contain His-Rrp41:MBP-Rrp41:Rrp42 will be found in this fraction. The second half of the incubated sample was applied to amylose resin that specifically interact with the MPB tag. The fraction that does not bind to amylose resin thus only contains His-Rrp41:Rrp42 exosome complexes. The fraction that binds to the amylose resin contains the MBP-Rrp41:Rrp42 complex and potentially the His-Rrp41:MBP-Rrp41:Rrp42 complex that can have formed as a result of subunit exchange during the incubation time.

(D) SDS page analysis. (1): the exosome complexes after incubation at 50°C for 3 days that contains His-Rrp41, MBP-Rrp41 and Rrp42. (2): The FT of the Ni-column, which contains, as expected, only MBP-Rrp41 and Rrp42. (3): The elution of the Ni-column, which only contains His-tagged Rrp41 and Rrp42. MBP-tagged Rrp41 is absent, showing subunit exchange did not take place. (4): The FT of the amylose column, which contains, as expected, only His-Rrp41 and Rrp42. (5): The elution of the amylose column, which contains only MBP-tagged Rrp41 and Rrp42. The His-tagged version of Rrp41 is absent from this fraction, proving that subunit exchange between different exosome complexes does not take place at rates that are significant for the experiments that we performed here.

RNA [1]	Sequence	Note	ID
20A [3]	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAA	3' end after ribozyme cleavage	12
32A	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAGCU	Run-off transcription from linearized plasmid	17
63A	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAGCU	Run-off transcription from linearized plasmid	30
92A	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAGCU	Run-off transcription from linearized plasmid	31
118A	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAGCU	Run-off transcription from linearized plasmid	32
32A FA[2]	GCCCCCCCCGAAAGGGGGGGGAAAAAAAAAAAAAA AAAAAAA-4-S-U-AAAAAAAAAAGCU	Synthesized (Dharmacon)	-

Protein [4]	Sequence	Note	ID
His-Rrp41[5]	MKHHHHHHPMSDYDIPTTENLYFQGAMGREMLQVER PKLILDDGKRTDGRKPDELRSIKIELGVLKNADGSAIFE MGNTKAIAAVYGPKEMHPRHLSLPDRAVLRVRYHMTP FSTDERKNPAPSRREIELSKVIREALESAVLVELFPRTA IDVFTEILQADAGSRLVSLMAASLALADAGIPMRDLIAG VAVGKADGVIILDLNETEDMWGEADMPIAMMPSLNQV TLFQLNGSMTPDEFRQAFDLAVKGINIIYNLEREALKSK YVEFKEEGV	pET based vector	82 422
MBP-Rrp41[5]	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTV EHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNL QEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGA KAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTI NGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLS AGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGA VALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAF WYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNN NNNNNPMSENLYFQGAMGREMLQVERPKLILDDGKR TDGRKPDELRSIKIELGVLKNADGSAIFEMGNTKAIAAV YGPKEMHPRHLSLPDRAVLRVRYHMTPFSTDERKNP APSRREIELSKVIREALESAVLVELFPRTAIDVFTEILQA DAGSRLVSLMAASLALADAGIPMRDLIAGVAVGKADG VIILDLNETEDMWGEADMPIAMMPSLNQVTLFQLNGS MTPDEFRQAFDLAVKGINIIYNLEREALKSKYVEFKEEG V	pET based vector	938 993
TwinStrep-Rrp41[5]	MKWSHPQFEKGGGSGGGSGGSSAWSHPQFEKPMS DYDIPTTENLYFQGAMGREMLQVERPKLILDDGKRTD GRKPDELRSIKIELGVLKNADGSAIFEMGNTKAIAAVYG PKEMHPRHLSLPDRAVLRVRYHMTPFSTDERKNPAPS RREIELSKVIREALESAVLVELFPRTAIDVFTEILQADAG SRLVSLMAASLALADAGIPMRDLIAGVAVGKADGVIILD LNETEDMWGEADMPIAMMPSLNQVTLFQLNGSMTPD EFRQAFDLAVKGINIIYNLEREALKSKYVEFKEEGV	pET based vector	989
Rrp42	MGMSSTPSNQNIIPIIKKESIVSLFEKGIRQDGRKLTDY RPLSITLDYAKKADGSALVKLGTTMVLAGTKLEIDKPYE DTPNQGNLIVNVELLPLAYETFEPGPPDENAIELARVV	pET based vector No affinity tags	373

	DRSLRDSKALDLTKLVIEPGKSVWTVWLDVYVLDYGG NVLDACTLASVAALYNTKVYKVEQHSNGISVNKNEVV GKLPLNYPVVTISVAKVDKYLVVDPDLDEESIMDAKISF SYTPDLKIVGIQKSGKGSMSLQDIDQAENTARSTAVKL LEELKKHLGI		
His-Rrp42	MKHHHHHHPMSDYDIPTTENLYFQGAMSSTPSNQNII PIIKKESIVSLFEKGIRQDGRKLTDYRPLSITLDYAKKAD GSALVKLGTTMVLAGTKLEIDKPYEDTPNQGNLIVNVE LLPLAYETFEPGPPDENAIELARVVDRSLRDSKALDLT KLVIEPGKSVWTVWLDVYVLDYGGNVLDACTLASVAA LYNTKVYKVEQHSNGISVNKNEVVGKLPLNYPVVTISV AKVDKYLVVDPDLDEESIMDAKISFSYTPDLKIVGIQKS GKGSMSLQDIDQAENTARSTAVKLLEELKKHLGI	pET based vector	487
His-Rrp41 Rrp42[6]	MKHHHHHHPMSDYDIPTTENLYFQGAMGREMLQVER PKLILDDGKRTDGRKPDELRSIKIELGVLKNADGSAIFE MGNTKAIAAVYGPKEMHPRHLSLPDRAVLRVRYHMTP FSTDERKNPAPSRREIELSKVIREALESAVLVELFPRTA IDVFTEILQADAGSRLVSLMAASLALADAGIPMRDLIAG VAVGKADGVIILDLNETEAMWGEADMPIAMMPSLNQV TLFQLNGSMTPDEFRQAFDLAVKGINIIYNLEREALKSK YVEFKEEGV MGMSSTPSNQNIIPIIKKESIVSLFEKGIRQDGRKLTDY RPLSITLDYAKKADGSALVKLGTTMVLAGTKLEIDKPYE DTPNQGNLIVNVELLPLAYETFEPGPPDENAIELARVV DRSLRDSKALDLTKLVIEPGKSVWTVWLDVYVLDYGG NVLDACTLASVAALYNTKVYKVEQHSNGISVNKNEVV GKLPLNYPVVTISVAKVDKYLVVDPDLDEESIMDAKISF SYTPDLKIVGIQKSGKGSMSLQDIDQAENTARSTAVKL LEELKKHLGI	pET based vector Coexpression of Rrp41 and Rrp42 Catalytically inactive Rrp41 (as indicated with the underlined D182A mutation in the sequence)	1113

Exosome complexes with a discrete number of active sites			
# active sites	Co-expression construct design[6]	Note	ID
3	His-Rrp41 WT TwinStrep-Rrp41 WT MBP-Rrp41 WT Rrp42	pET based vector. See above for the protein sequences	1020
2	His-Rrp41 D182A TwinStrep-Rrp41 WT MBP-Rrp41 WT Rrp42	pET based vector. See above for the protein sequences	1021
1	His-Rrp41 D182A TwinStrep-Rrp41 WT MBP-Rrp41 D182A Rrp42	pET based vector. See above for the protein sequences	1038

Table S1: Used constructs for protein expression and RNA *in vitro* transcription

1 The RNA stem-loop is highlighted in green.
2 RNA used for Fluorescence anisotropy measurements. 4-S-U refers to a thio-uridine that is used to couple the RNA to 6-(Iodoacetamido)-fluorescein.
3 Used for NMR studies as this RNA is not degraded by the exosome due to the presence of a 3' cyclic phosphate that results from ribosomal cleavage.
4 Protein affinity tags are colored red (His-Tag), pink (MBP-Tag) or yellow (STREP-Tag). TEV cleavage sites are highlighted in blue, where the TEV protease cleaves between the Q and G.
5 The inactive version of the Rrp41 construct contains the D182A mutation.
6 Co-expression vectors were constructed as described (5). All four proteins are in one vector and simultaneously induced.

References

1. Ishii, R., Nureki, O. and Yokoyama, S. (2003) Crystal structure of the tRNA processing enzyme RNase PH from Aquifex aeolicus. *The Journal of biological chemistry*, **278**, 32397-32404.
2. Symmons, M.F., Jones, G.H. and Luisi, B.F. (2000) A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation. *Structure*, **8**, 1215-1226.
3. Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature structural & molecular biology*, **12**, 575-581.
4. Liu, Q., Greimann, J.C. and Lima, C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, **127**, 1223-1237.
5. Mund, M., Overbeck, J.H., Ullmann, J. and Sprangers, R. (2013) LEGO-NMR spectroscopy: a method to visualize individual subunits in large heteromeric complexes. *Angewandte Chemie*, **52**, 11401-11405.

Methyl TROSY Spectroscopy to Study Large Biomolecular Complexes

Milos A. Cvetkovic and Remco Sprangers

Abstract

Solution state NMR spectroscopy is a powerful technique in structural biology that can provide unique information regarding the structure, dynamics, and interactions of biomolecular complexes. For a long time, its experimental range was limited to proteins of modest size. However, in the recent decades, the applicability of the method has been extended such that assemblies with molecular weights far over 100 kDa became amenable to detailed analyses. The breakthroughs that enabled these advances include the development of TROSY-based NMR techniques and procedures to produce samples that are labeled in specific methyl groups.

Here, we discuss these novel approaches to the study of high molecular weight systems, explaining briefly the theoretical background behind the advancements and giving several recent practical examples. The major applications of methyl TROSY NMR spectroscopy are mentioned: studies of intermolecular interactions, protein dynamics, and complex biomolecular structures. With all this, we substantiate our notion that NMR spectroscopy will continue to be a highly valuable and relevant method for investigating large biomolecular complexes that is complementary to other structural techniques.

Keywords

Methyl TROSY • Large molecular machine • Methyl-group labeling • Biomolecular structure • Biomolecular interactions • Protein dynamics • Methyl-group resonance assignment • Relaxation dispersion • Longitudinal exchange • Methionine scanning • LEGO NMR • Segmental isotope labeling • RNA protein interactions

M.A. Cvetkovic • R. Sprangers (✉)
Max Planck Institute for Developmental Biology, Tübingen, Germany
e-mail: milos.cvetkovic@tuebingen.mpg.de; remco.sprangers@tuebingen.mpg.de

G.A. Webb (ed.), *Modern Magnetic Resonance*,
DOI 10.1007/978-3-319-28275-6_45-1

Contents

Introduction

Over the years, NMR spectroscopy has demonstrated its strength in studying structure, dynamics, and interactions of various biomolecules at atomic resolution and under near-physiological solution conditions. The size and complexity of biomolecules that can be studied has, however, been posing a limitation to its applicability. Two major difficulties associated with NMR studies of larger proteins and biomolecular complexes are signal overlap and line broadening. Signal overlap is the consequence of the vast number of NMR active nuclei present in large systems and results in overcrowded spectra that are impossible to analyze. Line broadening arises from the relaxation of transverse magnetization, which is faster in large systems due to longer rotational correlation times and more abundant spin-spin interactions. As a result, both sensitivity and resolution of the measurement decrease significantly for proteins larger than 20 kDa [1]

Large biomolecular complexes perform some of the most important processes in the cell and it is, therefore, important to extend the size limitations of the NMR spectroscopy. Initially, NMR studies of unlabeled proteins were conducted by ^{1}H homonuclear experiments that were limited to systems of less than 10 kDa [2, 3]. Heteronuclear experiments on proteins uniformly labeled with ^{15}N and ^{13}C isotopes pushed the size limitation to around 25 kDa [4–6]. Protein deuteration had long been known to enhance resolution and sensitivity of the NMR experiments by reducing the number of possible spin relaxation pathways [7, 8]. Combined with ^{15}N and ^{13}C isotope labeling, deuteration extended the applicability of NMR spectroscopy to proteins of up to 50 kDa [9, 10]. Today, ^{15}N, ^{13}C, and ^{2}H labeling remains a standard procedure in studies of large protein systems. A revolutionary development in protein NMR spectroscopy was the introduction of transverse relaxation optimized spectroscopy (TROSY) [11], which enabled structure determination of proteins as large as 80 kDa [12] and analysis of proteins up to 900 kDa [13]. However, the adverse effects associated with size and complexity are often too

pronounced to allow recording of useful ^{1}H-^{15}N-based spectra for very large macromolecular assemblies. Utilization of isotopically labeled methyl groups [$^{13}CH_3$] in a fully deuterated background, in concert with the corresponding methyl TROSY spectra, has proven to be a successful approach for studying high molecular weight complexes. This strategy benefits from excellent relaxation properties of methyl groups where its three symmetrically arranged protons and fast rotation around its threefold symmetry axis produce highly sensitive and well-resolved NMR signal [14]. Methyl groups are usually located in the hydrophobic interior of proteins and along binding surfaces [15], which makes them valuable reporters of structural integrity, conformational changes, dynamics, and interactions. In summary, methyl TROSY NMR spectroscopy now enables recording high-quality spectra of assemblies with a molecular weight over 1 MDa [16].

Methyl Group Labeling Schemes

Development of sophisticated isotope labeling schemes and improvements in the pulse sequence design have played pivotal roles for extending the applicability of NMR spectroscopy to large biomolecular systems. Optimal isotope labeling involves, essentially, finding the right balance between having an adequate amount, distribution, and/or localization of the NMR-active nuclei (^{15}N, ^{13}C, ^{1}H) to provide enough information about the system and, at the same time, not compromising quality and manageability of the recorded spectra with too many NMR-active nuclei, which would otherwise lead to signal overlap and broadening.

Isotope labeling can be uniform, affecting the whole protein, and selective, affecting particular residues or groups of residues. In methyl TROSY NMR spectroscopy, both approaches are applied: protein is uniformly deuterated, while labeled methyl ($^{13}CH_3$) groups are selectively introduced in particular amino acids. Hydrogens (^{1}H), naturally abundant in proteins, cause extensive relaxation of transverse magnetization through interactions with each other and with other NMR-active nuclei (^{15}N, ^{13}C), thereby diminishing sensitivity and resolution. Hence, full deuteration (perdeuteration) is essential for studying large proteins as it minimizes this effect, due to 6.7-fold lower gyromagnetic ratio of ^{2}H compared to ^{1}H. Routinely, deuteration is achieved by growing the bacterial culture and overexpressing the protein in D_2O-based minimal medium in the presence of deuterated glucose [17].

Methyl ($^{13}CH_3$) isotopes can be introduced into individual amino acids or in a combination of several amino acids. Strategies for production of such proteins depend on the metabolic pathways of the respective amino acids [18]. The simplest way to achieve methyl group labeling is to add a methyl-labeled biosynthetic precursor to the bacterial culture before inducing protein overexpression. This approach is possible if there is no subsequent crossing between the precursor and the metabolic pathways of other amino acids. The precursor 2-ketobutyrate can, hence, be used for production of Ile-δ_1 [$^{13}CH_3$] [19, 20] and 2-hydroxy-2-ethyl-3-ketobutyrate for production of Ile-γ_2 [$^{13}CH_3$] [21]. The biosynthesis of valine and leucine, on the other hand, is connected and their mutual precursor, 2-keto-3-

isovalerate, is used for combined labeling of these amino acids [19]. Coproduction of stereospecifically labeled valine and leucine can be achieved using 2-acetolactate [22], while for separate labeling of valine and leucine residues one needs to prevent the scrambling of the precursor into both amino acids. As an example, stereospecifically labeled valine can be obtained either with labeled precursor 2-acetolactate and deuterated leucine [23] or with presynthesized stereospecifically labeled valine and deuterated leucine [24]. In both cases, the deuterated leucine prevents that the precursor is used in the leucine metabolic pathway. Metabolic pathways of the methyl groups of methionine, alanine, and threonine are intermingled with other metabolites and they need to be added before induction of protein expression, in their final, methyl-labeled forms: Met-ε [$^{13}CH_3$] [25], Ala-β [$^{13}CH_3$] [26], Thr-γ_2 [$^{13}CH_3$] [27].

Some of the strategies for the production of methyl-labeled proteins can be combined, which enables incorporation of different methyl reporters within a single protein simultaneously. The metabolic pathways used to produce labeled amino acids need to be compatible and the respective methyl TROSY resonances should not overlap in the spectrum. Combined labeling schemes that have been efficiently used are, for example, ILV [16, 28], MILV [29], AILV [30], and AMILVT [31].

Studies of large multidomain proteins could benefit from combining methyl TROSY NMR spectroscopy with the segmental isotope labeling. With the use of segmental isotope labeling, a specific isotope labeling strategy can be applied to a particular protein domain, while the remaining part of the protein remains NMR invisible. This simplifies the spectrum and allows for investigation of full-length proteins that would otherwise be out of reach for NMR spectroscopy [32]. Labeled and unlabeled parts of the protein are first produced separately by expressions under different labeling conditions and are subsequently ligated so that a peptide bond is formed between them. Ligation of the two parts of the protein is a critical step and can be performed either by using inteins, internally placed protein domains that are able to self-excise from a protein [33, 34] or by using transpeptidase Sortase A [35].

Work with large asymmetric protein complexes further stresses the need to reduce the number of signals in the spectrum to prevent signal overlap. This can be achieved by applying isotope labeling to only one or to a subset of subunits within the complex. In most protein systems, individual subunits may be insufficiently stable to sustain separate expression and purification, while in others *in vitro* reconstitution of the complex may be problematic. These issues can be overcome with the LEGO-NMR approach where subunits are sequentially coexpressed using different promoters, such that complex reconstitution takes place *in vivo* while only a subset of proteins is isotopically labeled [36].

Methyl Group Resonance Assignments

A detailed analysis of NMR spectra requires that the resonances are assigned to the corresponding residues in the protein. The assignment of the methyl resonances in high molecular weight proteins and complexes poses a serious challenge. For

proteins up to ~25 kDa, methyl resonances can be assigned conventionally, using the general strategy for assigning aliphatic side chains, which relies on the direct correlation between methyl spin systems and the already assigned protein backbone [37]. For larger proteins, an advanced assignment method has been developed where unassigned methyl group resonances and assigned amide group resonances are independently correlated with Cα and Cβ, thus providing a link between the corresponding methyl and amide group [28, 38, 39]. However, both methods depend on the feasibility of backbone assignments, which are usually difficult or impossible to obtain for proteins larger than 50 kDa. Several strategies can be used to overcome this obstacle.

Divide and Conquer

The "divide and conquer" strategy is based on dissecting the high molecular weight system into smaller building blocks [16, 29]. Since smaller building blocks usually show better spectral quality than the whole system, their methyl group resonances can be assigned using previously described standard approaches and afterwards linked to the corresponding methyl group signals of the whole system. This strategy has been successfully applied for protein complexes, where smaller building blocks represent individual subunits [16], as well as for large proteins, where smaller building blocks represent specific protein domains [29, 40].

Mutations to Assign Methyl Groups

If backbone-based assignment methods and the divide and conquer approach fail, methyl group resonances can be assigned by mutagenesis. Individual amino acid residues containing a labeled methyl group are substituted with alternative residues. By comparing the respective spectra before and after mutagenesis, the signal(s) belonging to a particular methyl group can be readily identified [41, 42]. It should be noted, however, that this strategy can be complicated by secondary chemical shift perturbations, where the mutation of a single amino acid results in significant changes in the methyl TROSY spectrum.

Structure-Based Assignment Methods

Structure-based procedures for methyl group assignments require detailed knowledge regarding the structure of the complex. Methyl-methyl distance information obtained from NOE data can, in those cases, be mapped with distances that are known from the structure, thereby assigning the methyl group resonances [43]. Another structure-based approach involves introduction of a spin-label at specific sites of the protein. Residues in the vicinity of such a label will consequently experience pseudocontact shifts (PCSs) [44] or paramagnetic relaxation enhancement (PRE) [45], enabling assignment of affected methyl groups.

The Methyl TROSY Experiment

The intensity of an NMR signal directly depends on the amount of the initial magnetization and for that reason methyl groups that contain three protons are sensitive NMR probes. Furthermore, the NMR signal intensity depends on the decay rate of the magnetization during the experiment and significant improvements in spectral quality can be achieved with TROSY-based experiments.

During an NMR experiment, a large number of magnetization terms are created that all lose their magnetization with distinct relaxation rates. Due to destructive interference of relaxation mechanisms, a number of coherences can relax significantly slower than others. In TROSY-based experiments, these slow relaxing components are selected and special care is taken to prevent that they are mixed with fast relaxing ones. Such a mixing would result in an overall faster relaxation of the NMR signal and thus in NMR spectra of reduced quality.

The TROSY-based experiment was initially introduced for proton-nitrogen spectra that only detect the slow relaxing quarter of the NMR signal [11]. For large proteins, this results in NMR spectra of significantly improved quality compared to more traditional HSQC based experiments, where fast and slow relaxing coherence are mixed and added. The slow relaxation of specific coherences of H-N groups is caused by the destructive interference between the proton-nitrogen dipole-dipole and the nitrogen CSA (chemical shift anisotropy) relaxation interactions. The CSA depends on the magnetic field strength and an optimal cancellation of both relaxation mechanisms takes place at a magnetic field strength of around 1.1 GHz [11].

The carbon CSA in methyl groups is too small to induce destructive interference with the proton-carbon dipole. However, in the macromolecular limit, destructive interference occurs between the proton-carbon dipole-dipole and the multiple proton-proton dipolar relaxation interactions [14, 46]. To analyze the relaxation properties of a methyl group one needs to consider 16 different energy levels that are connected by 4 fast and 6 slowly relaxing proton transitions, 2 fast and 6 slowly relaxing carbon transitions, and 4 fast and 6 slowly relaxing ^{1}H-^{13}C double-/zero-quantum transitions. As was shown by Kay and coworkers, the rapidly and the slowly relaxing transitions are never mixed in the HMQC pulse sequence [14]. This renders the HMQC a highly efficient methyl TROSY experiment, where the final NMR signal of an isolated methyl group only results from slowly relaxing coherences. This is in strong contrast to the more popular HSQC-based experiment in which 90° proton pulses result in the mixing of fast and slow relaxing coherences and thus in NMR spectra with broader signals, especially for large proteins. To fully exploit the methyl TROSY effect, the methyl group of interest should be embedded in an otherwise fully deuterated protein (see above) as dipolar interactions with external protons cause interconversion of fast and slow relaxing coherences. Finally, it is worth noting that the methyl TROSY effect does not depend on the magnetic field strength as it results from the destructive interference between dipolar relaxation interactions [14].

Applications

As previously explained, methyl TROSY NMR spectroscopy benefits significantly from excellent spectral features of the methyl groups and offers a wide range of possibilities to study large biomolecular systems. It is able to yield the same quantitative information about a complex system as was only available for small proteins until recently [47]. The applications include studies of intermolecular interactions and protein dynamics, revealing novel mechanisms of biomolecular processes, as well as providing new insights into the structure of very large biomolecular systems.

Intermolecular Interactions

NMR spectroscopy is applicable to binding events with affinities that range from the nM to mM regime. Methyl bearing side-chains are generally well-suited probes for investigating binding surfaces as they often play an important role in biomolecular interactions. Changes in the local chemical environment, due to ligand binding, will affect methyl groups positioned within the binding interface. This causes chemical shift perturbations (CSPs) [48] that provide both qualitative and quantitative information about the interaction. Here, we describe a number of recent examples, where methyl TROSY NMR-based binding experiments have provided pivotal insights into biological function.

The molecular chaperone Hsp90, which forms a homodimer of 170 kDa, has an important role in protein folding. ATP and various cochaperones, including its binding partner p23, control its function. Hsp90 consists of three domains: an N-terminal domain, a middle domain, and a C-terminal domain. Stability of the isolated domains enabled successful assignments of Ile-δ1 methyl groups through divide and conquer strategy [49]. By observing CSPs in the methyl TROSY spectrum upon the addition of ligands, it was shown that ATP binds only to the N-terminal domain, while p23 cochaperone binds both the N-terminal and the middle domains of Hsp90 [49]. In addition, a 106 Å long interface was identified on Hsp90 that mediates the interaction with the intrinsically disordered Tau protein through many low-affinity contacts [50].

Binding studies in even larger complexes, like the 20S proteasome (670 kDa), show that methyl TROSY-based experiments are also feasible for systems of that size. The proteasome is responsible for degradation of damaged and dispensable proteins and its barrel-shaped core particle consists of four homo-heptameric stacked rings. The two outer rings, which form the entrance for substrates, can each bind the 150 kDa 11S activator, resulting in a complex with a molecular weight of 1.1 MDa. Methyl labeling of the outer subunits at Ile-δ_1, Leu-δ, and Val-γ positions, assignment of their resonances through the divide and conquer strategy, and successful reconstitution of the 20S complex enabled mapping of the 11S binding surface on the 20S and determination of the associated affinity [16].

Ile- δ_1, Leu-δ, and Val-γ methyl labeling was also used for the study of the interaction between the molecular chaperons ClpB (580 kDa) and DnaK (70 kDa) that is crucial for protein disaggregation. Methyl labeling of only one component of the complex at a time, while keeping the other parts NMR invisible, provided high quality methyl TROSY spectra that were exploited to map interaction surfaces in a quantitative manner [51].

Methionine Scanning

Methyl-containing amino acids that can be used in methyl TROSY experiments might not be present on the protein surface with the quantity and distribution to allow for a detailed mapping of interaction sites. To overcome this drawback, the methionine scanning approach has recently been introduced, which provides an increased coverage of the protein surface with methyl probes [52, 53]. It involves the strategic substitution, one-at-a-time, of solvent exposed residues with methyl-labeled methionine (Met-ε). The signal of the introduced reporter methionine appears as a novel resonance in the methyl TROSY spectrum and can, thus, be instantly assigned. After the addition of the ligand, a new methyl TROSY spectrum is recorded. If the reporter Met is located inside the binding interface, it will experience new local chemical environment, which will be manifested as CSP. If the reporter is located outside of the binding interface, there will be no CSP of the corresponding signal. If a residue that is pivotal for the interaction (a hot-spot) is substituted with a reporter methionine, the interaction will be abolished, which is noticeable in the absence of CSPs of the naturally occurring methyl groups.

Methionine scanning was recently successfully employed for the study of RNA: protein interactions within the archaeal exosome. The exosome complex is an important molecular machine responsible for RNA 3′ to 5′ processing and degradation. The archaeal complex has a molecular weight of 270 kDa and is composed of a hexameric core and a trimeric cap. The hexameric core is a trimer of Rrp41-Rrp42 heterodimers, while the cap is made of three copies of the Rrp4 protein. The Rrp4 cap forms the opening through which RNA substrate enters the catalytic interior of the assembly. Based on the methionine scanning approach, a 50 Å long RNA binding path on each Rrp4 protomer was identified. The interaction between the Rrp4 cap and the RNA substrate, which was unattainable by the crystallographic data, proved to be crucial for the efficient recruitment and channeling of the RNA substrate towards the active sites [54].

Protein Dynamics

Experiments for Millisecond Methyl Dynamics

Proteins are highly dynamic biomolecules that can adopt multiple conformations. Enzymes, in particular, sample structurally different states to perform biological functions. As enzyme turnover rates in biology are often in the range between 0.1

and 5000 per second, there is a special interest in detecting motions on these timescales. Here, we briefly discuss longitudinal exchange and CPMG (Carr–Purcell–Meiboom–Gill) relaxation dispersion experiments that are both able to detect and quantify such motions. We focus on the applicability of these methods to ($^{13}CH_3$) methyl groups in large protein complexes, as these samples are ideally suited for methyl TROSY spectroscopy [55].

Longitudinal exchange experiments are applicable to systems where the spin of interest exchanges between two states (A and B) with a rate that is slow compared to the chemical shift difference between the two states ($k_{ex} << \Delta\omega$; where k_{ex} is the exchange rate $k_{AB} + k_{BA}$ and $\Delta\omega$ is the chemical shift difference between states A and B). As a result, one single methyl group gives rise to two different signals in the NMR spectrum, one where the protein adopts conformation A and one where it adopts conformation B. Central to the HMQC-based longitudinal exchange experiment is a delay that is sandwiched between the carbon and proton chemical shift evolution times. In this delay the protein can change its conformation such that state A becomes state B and vice versa. In the NMR spectrum, this results in the appearance of resonances at the carbon chemical shift of state A (or B) and at the proton chemical shift of state B (or A). The dependence of the intensities of these "cross peaks" on the length of the delay directly reports on the kinetics behind the exchange process. Despite the fact that the methyl TROSY principle cannot be exploited during the complete NMR pulse sequence, longitudinal exchange experiments have been successfully applied to very large protein complexes [41, 56–58].

CPMG relaxation dispersion experiments [59] are applicable to systems where exchange of a protein from state A to state B results in the broadening of the NMR resonances. Such broadening is induced when the exchange takes place on a timescale that is comparable to the chemical shift difference between the two states ($k_{ex} \sim \Delta\omega$), which, in turn, results in a dephasing of the magnetization. The extent of this dephasing depends on the difference in the chemical shift between states A and B ($\Delta\omega$), on the exchange rate (k_{ex}), and on the populations of the two states. Interestingly, significant exchange broadening of the resonance of state A can also occur when state B is only sparsely populated (e.g., less than 5%) and therefore not directly observable in the NMR spectrum. In such a situation, line broadening of the resonance of state A can report on the presence of an "invisible" state B. The dephasing of the magnetization (and thus the line-broadening that is induced by the exchange process) can be suppresses by a train of refocusing pulses. The dependence of the line broadening on the frequency with which the refocusing pulses are applied is then used to extract the kinetic parameters that underlie the exchange process. CPMG relaxation dispersion experiments can be recorded in a variety of different manners, depending on the magnetization state during the time when the refocusing pulses are applied. First, ^{1}H-^{13}C multiple quantum (MQ)-based experiments, which can be recorded in a manner that is fully compatible with the methyl TROSY principle, report on the proton and carbon chemical shift differences of states A and B [60]. The information content of MQ dispersion experiments is, thus, very high, which can make the analysis of the data complicated. Second, single quantum (SQ)-based experiments can be recorded on $^{13}CH_3$ labeled methyl groups.

These experiments can be designed such that only the ^{13}C [61, 62] or only the ^{1}H [63] chemical shift difference is sensed in the relaxation dispersion profiles. Compared to the MQ experiments, the analysis of the SQ relaxation dispersion data is less complicated. These SQ experiments are, however, not as sensitive as the MQ experiment since only a part of the magnetization is selected and the methyl TROSY effect cannot be fully exploited. Nevertheless, ^{13}C SQ relaxation dispersion experiments have been successfully applied to protein complexes over 100 kDa to quantify exchange processes [64]. Finally, a ^{1}H triple quantum (TQ) relaxation dispersion experiment has been introduced recently, where the dispersion profiles depend on three times the proton chemical shift difference between states A and B [65]. Importantly, the dispersions in these experiments are by a factor of 10 larger than in the ^{1}H (SQ) relaxation dispersion experiments and thus applicable to larger protein complexes and to a wider range of exchange processes.

In general, extracting accurate exchange parameters (exchange rates, populations, and chemical shift differences) from a single measurement at a single magnetic field strength is very challenging. To improve the accuracy of the extracted parameters, the relaxation dispersion experiment can be repeated on multiple field strengths, as the chemical shift difference depends on the spectrometer field. Alternatively, different relaxation dispersion experiments (SQ/MQ/TQ) can be analyzed simultaneously to improve the robustness of the fitting of the data.

Example of Dynamics

As mentioned above, dynamic processes are often related with enzymatic function and NMR relaxation experiments are ideally suited to reveal potential correlations between dynamic processes and catalytic activity. DspS is an 80 kDa homodimeric enzyme that catalyzes the hydrolysis the 5′ cap structure from short eukaryotic mRNAs, as the final step of their degradation pathway. The enzyme consists of a smaller N-terminal domain, which is flexibly connected to a larger C-terminal domain. Substrate can be positioned between the N and C-terminal domains, on both sides of the homodimer. The two binding sites strongly influence each other and the two substrates interact with the enzyme with significantly different affinities [58]. To quantify the motions of the N-terminal domain, longitudinal exchange experiments were performed on the enzyme that was produced with methyl-labeled Ile-δ_1 and Met-ε. The study of intramolecular dynamics of this flexible system and its association with the substrate binding and activity revealed that the excess of flipping motion, induced by the binding of a second ligand, hampers the catalytic activity [58]. This finding highlights the relationship between intramolecular dynamics and activity and, hence, also the importance of understanding and quantifying intramolecular motions in enzymes.

Elucidating Complex Structures

Complex assemblies that contain several different subunits and/or co-factors perform many important biological processes. In order to understand the structural

and functional relationships between the components involved in these processes, it is necessary to combine different structural, biophysical, and biochemical methods.

The archaeal box C/D ribonucleoprotein enzyme is a highly complex system that contains the proteins L7Ae, Nop5 and fibrillarin, plus a guide sRNA. This 390 kDa complex methylates ribosomal RNA at the 2′-O-ribose, which is an important part of the pre-rRNA maturation and a necessary step for subsequent ribosome assembly. The structure of certain isolated segments of this ribonucleoprotein system was known, and these structures were assembled into the complete C/D RNP complex with the help of methyl CSPs, PRE experiments, SAXS (small angle X-ray scattering), and SANS (small angle neutron scattering) data. Furthermore, methyl CSPs observed upon addition of the RNA substrate were instrumental for understanding the elaborate mechanism of sequential site-specific methylation [66].

SecB is a molecular chaperone with a strong antifolding activity. This 70 kDa tetramer displays methyl TROSY spectra of high quality when expressed with methyl-labeled Ala-β Val-γ, Leu-δ, Met-**ε**, Thr-γ2, and Ile-δ1. These methyl groups were used for determining the structure of the complex that SecB forms with its client proteins MBP and PhoA in their unfolded state. This clearly revealed the mechanism by which the molecular chaperons are able to keep client proteins in an unfolded state. In brief, it was shown that SecB forms long, continuous hydrophobic grooves that bind multiple hydrophobic segments exposed across the unfolded client protein [31]. Multivalent binding mode of this interaction leads to a structure where the client protein is wrapped around SecB, which affects the folding kinetics of the substrate and keeps it in an unfolding state.

Conclusion

Here, we discussed the methodology that can extend the applicability of solution state NMR spectroscopy to systems that are orders of magnitude larger than those that are traditionally studied by this technique. Indeed, the combination of methyl-labeled samples and methyl TROSY experiments can provide quantitative insights in assemblies up to 1 MDa in molecular weight. Importantly, NMR spectroscopy is able to localize and quantify interactions and dynamic processes on a per residue basis. This information is highly relevant to biomolecular processes and often hidden in static structures. In that light, NMR spectroscopy is able to provide unique insights that are fully complementary to X-ray crystallography and cryo-EM methods. Indeed, recently the complementarity of methyl TROSY NMR spectroscopy and cryo-EM was impressively illustrated for a large AAA+ unfoldase [67]. We anticipate that approaches like that will be increasingly important in the future and we look forward to studies that will unravel a wide range of yet unexplored molecular mechanisms.

References

1. Goto NK, Kay LE. New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr Opin Struct Biol. 2000;10(5):585–92.
2. Wagner G, Wuthrich K. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. J Mol Biol. 1982;155(3):347–66.
3. Williamson MP, Havel TF, Wuthrich K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. J Mol Biol. 1985;182(2):295–315.
4. Fesik SW, Gampe Jr RT, Zuiderweg ER, Kohlbrenner WE, Weigl D. Heteronuclear three-dimensional NMR spectroscopy applied to CMP-KDO synthetase (27.5 kD). Biochem Biophys Res Commun. 1989;159(2):842–7.
5. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry. 1989;28(15):6150–6.
6. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 1990;29(19):4659–67.
7. Crespi HL, Rosenberg RM, Katz JJ. Proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains. Science. 1968;161(3843):795–6.
8. Markley JL, Putter I, Jardetzky O. High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. Science. 1968;161(3847):1249–51.
9. Garrett DS, Seok YJ, Liao DI, Peterkofsky A, Gronenborn AM, Clore GM. Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate: sugar phosphotransferase system by multidimensional NMR. Biochemistry. 1997;36(9):2517–30.
10. Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, Kay LE. Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. J Mol Biol. 2000;300(1):197–212.
11. Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 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 U S A. 1997;94 (23):12366–71.
12. Tugarinov V, Choy WY, Orekhov VY, Kay LE. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci U S A. 2005;102(3):622–7.
13. Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K. NMR analysis of a 900K GroEL GroES complex. Nature. 2002;418(6894):207–11.
14. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced 1H [bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc. 2003;125(34):10420–8.
15. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. J Mol Biol. 1988;204(1):155–64.
16. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007;445(7128):618–22.
17. Gardner KH, Kay LE. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct. 1998;27:357–406.
18. Kerfah R, Plevin MJ, Sounier R, Gans P, Boisbouvier J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. Curr Opin Struct Biol. 2015;32:113–22.
19. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. J Biomol NMR. 1999;13(4):369–74.

20. Gardner KH, Kay LE. Production and incorporation of N-15, C-13, H-2 (H-1-delta 1 methyl) isoleucine into proteins for multidimensional NMR studies. J Am Chem Soc. 1997;119 (32):7599–600.
21. Ruschak AM, Velyvis A, Kay LE. A simple strategy for (1)(3)C, (1)H labeling at the Ile-gamma2 methyl position in highly deuterated proteins. J Biomol NMR. 2010;48(3):129–35.
22. Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, et al. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem. 2010;49(11):1958–62.
23. Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J. Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. J Biomol NMR. 2013;57(3):251–62.
24. Miyanoiri Y, Takeda M, Okuma K, Ono AM, Terauchi T, Kainosho M. Differential isotope-labeling for Leu and Val residues in a protein by *E. coli* cellular expression using stereo-specifically methyl labeled amino acids. J Biomol NMR. 2013;57(3):237–49.
25. Fischer M, Kloiber K, Hausler J, Ledolter K, Konrat R, Schmid W. Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. Chembiochem. 2007;8(6):610–2.
26. Ayala I, Sounier R, Use N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J Biomol NMR. 2009;43 (2):111–9.
27. Velyvis A, Ruschak AM, Kay LE. An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. PLoS One. 2012;7(9):e43725.
28. Tugarinov V, Kay LE. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc. 2003;125 (45):13868–78.
29. Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, et al. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell. 2007;131(4):756–69.
30. Godoy-Ruiz R, Guo C, Tugarinov V. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc. 2010;132 (51):18340–50.
31. Huang C, Rossi P, Saio T, Kalodimos CG. Structural basis for the antifolding activity of a molecular chaperone. Nature. 2016;537(7619):202–6.
32. Rosenzweig R, Farber P, Velyvis A, Rennella E, Latham MP, Kay LE. ClpB N-terminal domain plays a regulatory role in protein disaggregation. Proc Natl Acad Sci U S A. 2015;112(50): E6872–81.
33. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW. Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. J Am Chem Soc. 2003;125 (35):10561–9.
34. Berrade L, Camarero JA. Expressed protein ligation: a resourceful tool to study protein structure and function. Cell Mol Life Sci. 2009;66(24):3909–22.
35. Freiburger L, Sonntag M, Hennig J, Li J, Zou P, Sattler M. Efficient segmental isotope labeling of multi-domain proteins using Sortase A. J Biomol NMR. 2015;63(1):1–8.
36. Mund M, Overbeck JH, Ullmann J, Sprangers R. LEGO-NMR spectroscopy: a method to visualize individual subunits in large heteromeric complexes. Angew Chem. 2013;52 (43):11401–5.
37. Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J Biomol NMR. 1993;3(2):185–204.
38. Tugarinov V, Kay LE. Side chain assignments of Ile delta 1 methyl groups in high molecular weight proteins: an application to a 46 ns tumbling molecule. J Am Chem Soc. 2003;125 (19):5701–6.

39. Mishra SH, Frueh DP. Assignment of methyl NMR resonances of a 52 kDa protein with residue-specific 4D correlation maps. J Biomol NMR. 2015;62(3):281–90.
40. Ogunjimi AA, Wiesner S, Briant DJ, Varelas X, Sicheri F, Forman-Kay J, et al. The ubiquitin binding region of the Smurf HECT domain facilitates polyubiquitylation and binding of ubiquitylated substrates. J Biol Chem. 2010;285(9):6308–15.
41. Sprangers R, Gribun A, Hwang PM, Houry WA, Kay LE. Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. Proc Natl Acad Sci U S A. 2005;102(46):16678–83.
42. Amero C, Asuncion Dura M, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, et al. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR. 2011;50(3):229–36.
43. Xiao Y, Warner LR, Latham MP, Ahn NG, Pardi A. Structure-based assignment of Ile, Leu, and Val methyl groups in the active and inactive forms of the mitogen-activated protein kinase extracellular signal-regulated kinase 2. Biochemistry. 2015;54(28):4307–19.
44. John M, Schmitz C, Park AY, Dixon NE, Huber T, Otting G. Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. J Am Chem Soc. 2007;129 (44):13749–57.
45. Venditti V, Fawzi NL, Clore GM. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. J Biomol NMR. 2011;51(3):319–28.
46. Ollerenshaw JE, Tugarinov V, Kay LE. Methyl TROSY: explanation and experimental verification. Magn Reson Chem. 2003;41(10):843–52.
47. Rosenzweig R, Kay LE. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. Annu Rev Biochem. 2014;83:291–315.
48. Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013;73:1–16.
49. Karagoz GE, Duarte AM, Ippel H, Uetrecht C, Sinnige T, van Rosmalen M, et al. N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. Proc Natl Acad Sci U S A. 2011;108(2):580–5.
50. Karagoz GE, Duarte AM, Akoury E, Ippel H, Biernat J, Moran Luengo T, et al. Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. Cell. 2014;156(5):963–74.
51. Rosenzweig R, Moradi S, Zarrine-Afsar A, Glover JR, Kay LE. Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. Science. 2013;339(6123):1080–3.
52. Stoffregen MC, Schwer MM, Renschler FA, Wiesner S. Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. Structure. 2012;20(4):573–81.
53. Mari S, Ruetalo N, Maspero E, Stoffregen MC, Pasqualato S, Polo S, et al. Structural and functional framework for the autoinhibition of Nedd4-family ubiquitin ligases. Structure. 2014;22(11):1639–49.
54. Cvetkovic MA, Wurm JP, Audin MJ, Schutz S, Sprangers R. The Rrp4-exosome complex recruits and channels substrate RNA by a unique mechanism. Nat Chem Biol. 2017. doi:10.1038/nchembio.2328.
55. Religa TL, Kay LE. Optimal methyl labeling for studies of supra-molecular systems. J Biomol NMR. 2010;47(3):163–9.
56. Audin MJ, Dorn G, Fromm SA, Reiss K, Schutz S, Vorlander MK, et al. The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding. Angew Chem. 2013;52(32):8312–6.
57. Religa TL, Sprangers R, Kay LE. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science. 2010;328(5974):98–102.
58. Neu A, Neu U, Fuchs AL, Schlager B, Sprangers R. An excess of catalytically required motions inhibits the scavenger decapping enzyme. Nat Chem Biol. 2015;11(9):697–704.
59. Palmer 3rd AG, Kroenke CD, Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzymol. 2001;339:204–38.

60. Korzhnev DM, Kloiber K, Kanelis V, Tugarinov V, Kay LE. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. J Am Chem Soc. 2004;126(12):3964–73.
61. Skrynnikov NR, Mulder FA, Hon B, Dahlquist FW, Kay LE. Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion NMR measurements: application to methionine residues in a cavity mutant of T4 lysozyme. J Am Chem Soc. 2001;123(19):4556–66.
62. Lundstrom P, Vallurupalli P, Religa TL, Dahlquist FW, Kay LE. A single-quantum methyl 13C-relaxation dispersion experiment with improved sensitivity. J Biomol NMR. 2007;38 (1):79–88.
63. Tugarinov V, Kay LE. Separating degenerate (1)H transitions in methyl group probes for single-quantum (1)H-CPMG relaxation dispersion NMR spectroscopy. J Am Chem Soc. 2007;129 (30):9514–21.
64. Audin MJ, Wurm JP, Cvetkovic MA, Sprangers R. The oligomeric architecture of the archaeal exosome is important for processive and efficient RNA degradation. Nucleic Acids Res. 2016;44(6):2962–73.
65. Yuwen T, Vallurupalli P, Kay LE. Enhancing the sensitivity of CPMG relaxation dispersion to conformational exchange processes by multiple-quantum spectroscopy. Angew Chem. 2016;55 (38):11490–4.
66. Lapinaite A, Simon B, Skjaerven L, Rakwalska-Bange M, Gabel F, Carlomagno T. The structure of the box C/D enzyme reveals regulation of RNA methylation. Nature. 2013;502 (7472):519–23.
67. Huang R, Ripstein ZA, Augustyniak R, Lazniewski M, Ginalski K, Kay LE, et al. Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. Proc Natl Acad Sci U S A. 2016;113(29):E4190–9.