Molecular mechanisms of messenger RNA decapping

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Sowndarya Muthukumar

aus Bengaluru, India

Tübingen

2019

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

Tag der mündlichen Qualifikation:	24.07.2019
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Thilo Stehle
2. Berichterstatter:	Jun. Prof. Dr. Jennifer Ewald

DECLARATION

This thesis describes my work conducted in the laboratory of Prof. Dr. Elisa Izaurralde in the Department of Biochemistry at the Max Planck Institute for Developmental Biology, Tübingen, Germany, from August 2014 to May 2019. The work was supervised by Dr. Eugene Valkov and co-supervised by Prof. Dr. Thilo Stehle at the Eberhard Karls University of Tübingen, Germany, and was supported by a fellowship from the International PhD Program in the Biological Sciences, Tübingen. I declare that this thesis is the product of my own work. The parts that have been published or where other sources have been used were cited accordingly.

ACKNOWLEDGMENTS

I would like to thank the late Prof. Dr. Elisa Izaurralde for giving me an opportunity to carry out my doctoral thesis work in her lab, for giving me the opportunity to work on a challenging topic and for her mentorship. She will always remain an inspiration.

The completion of my doctorate would not have been possible without the support of my supervisor, Dr. Eugene Valkov, who has been there with me from the beginning. Thank you for believing in me, for always giving me a nudge in the right direction, for encouraging my ideas, and for always being there to troubleshoot. You have played a major role in all aspects of my development as a scientist and, for that, I will be forever grateful. It has been a pleasure working with you.

I am very grateful to Prof. Dr. Thilo Stehle and Dr. John Weir for their very valuable input over these last years, serving on my thesis advisory committee and evaluating my thesis. I would like to thank Dr. Cátia Igreja for many helpful suggestions throughout the years. I am also grateful to Jun. Prof. Dr. Jennifer Ewald and Prof. Dr. Dirk Schwarzer for accepting an invitation to serve on my thesis defense committee.

I thank Prof. Dr. Ralf J. Sommer for his support over the last year.

Many thanks to the Ph.D. program coordinator, Dr. Dagmar Sigurdardottir, for her support and for the many courses she organized as a part of the Max Planck International Ph.D. program.

In the lab, I cannot begin to express my thanks to Catrin Weiler for her great technical support and help with various aspects of my project. It was a lot of fun working with you and I learned a lot from your exceptional organizational skills. Thanks also to Gabi Wagner for assistance with the insect cell culture, Sigrun Helms and Maria Fauser for a lot of help over the years.

My thanks go to all members of the department who I have known in my time here. We have had some great times and I will always cherish the memories of cheering for the RNA Devils and the great hike we had. Special thanks to Tobi, Yevgen, Chung-Te, Lara, Dipankar, Praveen, Csilla, Maria, Felix R., Michelle,

Ramona, Simone, Aoife, Annamaria, Felix S., Ying, Daniel, and Stefan. I would also like to thank Max, Ida, Marie and Nadine for laboratory support.

I would like to say thank you to some very special people whose company I have really enjoyed: Andrea, Chatu, Cataleya, Enzo, and Dani. Thank you guys for making me feel at home, for all those stress-free hours of laughter, great food and multiple bottles of wine and limoncello.

Last but not least, I would like to thank my family, especially my mother, who taught me to work hard and who constantly supported and encouraged me to follow my heart. My significant other, Girish, who has been very accommodating, put up with my many eccentricities and stayed patient, and whose support has made this journey much easier and a very memorable one.

TABLE OF CONTENTS

1. SUMMARY

The work presented in this doctoral thesis describes biochemical and structural investigations of the molecular mechanisms of messenger RNA decapping—a critical step in one of the major cytoplasmic messenger RNA decay pathways. Decapping, mediated by the DCP2 enzyme, is the removal of the protective methylguanosine cap modification from the 5′ end of messenger RNA. Decapping is a tightly regulated process because it acts as a 'trigger' event for the complete degradation of transcripts by the exoribonuclease XRN1 in the 5′–3′ decay pathway.

The initial focus of my doctoral work was on the core decapping complex in fission yeast comprising the Dcp2 enzyme and its obligate coactivator Dcp1, which form a stable binary complex. I crystallized this core complex with a peptide motif from an intrinsically disordered activator Edc1. The structural analysis supported by biochemical validation indicated that Edc1 may function by trapping the dynamic and flexible decapping complex in a conformation where the critical RNA-binding interfaces are optimally positioned for substrate binding.

The interactions between decapping regulatory factors have diverged in evolution. Unlike in yeast, metazoan decapping factors form a rather loose network of interactions. Many factors have a modular architecture with conserved structured domains connected by flexible unstructured regions, which have a crucial function because they mediate interactions and complex assembly via embedded short linear motifs. I set up an insect cell expression system optimized for the recombinant production of proteins with extended low-complexity regions. Using this approach, I was able to reconstitute an almost entire human decapping interactome as a manipulable and biochemically tractable *in vitro* system. The full-length human DCP2 is a highly active enzyme and its C-terminal region contributes directly to RNA binding. A systematic biochemical screen to investigate the effect of various purified activators on DCP2 activity *in vitro* revealed that the intrinsically disordered protein PNRC2, together with DCP1, is a principal coactivator of DCP2. This provided the first comprehensive biochemical framework for the study of metazoan decapping regulation *in vitro* and provides a solid platform for future mechanistic studies.

1. 1. ZUSAMMENFASSUNG

Im Rahmen meines Dissertationsprojektes habe ich mit Hilfe biochemischer und strukturbiologischer Methoden die molekularen Mechanismen des mRNA Decappings, einem kritischen Schritt beim Abbau zytoplasmatischer mRNAs, untersucht. Unter „Decapping'' versteht man das Entfernen der schützenden Methylguanosin-Modifikation vom 5′-Ende der mRNA. Der Decapping-Prozess ist streng reguliert, da dieser Schritt die vollständige 5′-3′ Degradation der mRNA durch die Exonuklease XRN1 auslöst.

Der Fokus meiner Arbeit lag zunächst auf dem Decapping-Kernkomplex in Hefe. Dieser besteht aus dem Enzym Dcp2 und dem erforderlichen Coaktivator Dcp1, welche einen stabilen Komplex bilden. Diesen Komplex konnte ich in Verbindung mit einem Peptidmotiv des unstrukturierten Aktivators Edc1 kristallisieren. Die Strukturanalyse und die unterstützenden biochemischen Untersuchungen dieses Komplexes legen nahe, dass Edc1 den dynamischen und flexiblen Decapping-Komplex in einer Konformation fixieren kann, welche eine optimale Ausrichtung der kritischen RNA-Bindestellen für die Substratbindung gewährleistet.

Das Interaktionsmuster der regulatorischen Faktoren des Decappings hat sich im Laufe der Evolution verändert. Im Gegensatz zu Hefe sind die Decapping-Faktoren bei den Metazoa vergleichsweise lose vernetzt. Viele Faktoren haben eine modulare Architektur mit konservierten strukturierten Domänen, die über flexible, nicht strukturierte Regionen miteinander verbunden sind. Letztere sind von großer Bedeutung, da sie sogenannte „Short Linear Motifs'' enthalten, die für Protein-Interaktionen und die Ausbildung von Komplexen notwendig sind. Um derartige Proteine mit ausgedehnten Regionen geringer Sequenz-Komplexität optimal zu exprimieren, habe ich zunächst ein auf Insektenzellen basiertes Expressionssystem etabliert. Mithilfe dieses Expressionsystems gelang es mir nahezu das komplette humane Decapping-Interaktom in vitro zu rekonstituieren. Das hoch aktive humane Enzym DCP2 trägt über seinen C-Terminus direkt zur RNA Bindung bei. In einer systematischen biochemischen Analyse untersuchte ich die Auswirkung verschiedener aufgereinigter Aktivatoren auf die DCP2 Aktivität. Dabei identifizierte ich PNRC2, ein weitgehend unstrukturiertes Protein, zusammen mit DCP1, als

wichtigste Koaktivatoren für DCP2. Meine Arbeit lieferte erstmals einen umfassenden biochemischen Rahmen für die in vitro Studie der Regulation des Decapping-Mechanismus in Metazoa und legt ein solides Fundament für zukünftige mechanistische Studien.

2. INTRODUCTION

2.1 The life of a messenger RNA: from birth to death

Messenger RNA (mRNA) has a crucial role in gene expression and its regulation. In eukaryotes, an mRNA is synthesized in the nucleus where the genetic information encoded in DNA is transferred to mRNA in the process of transcription mediated by RNA polymerase II **(Fig. 1)**. Following transcription, the precursor mRNA (pre-mRNA) undergo processing and maturation into a mature mRNA (Moore 2005). The first processing event that occurs is the addition of a 7-methlyguanylate (m^7G) 'cap' structure at the 5' terminus of the newly synthesized pre-mRNA **(Fig. 1)**. This occurs co-transcriptionally via the capping enzyme complex (CEC) which associates with the RNA Polymerase II and caps the nascent pre-mRNA (Fabrega et al. 2003; Hirose and Manley 2000). Capping acts not only as a unique identifier for the subsequent steps of splicing, polyadenylation, and nuclear export, but also as a protective structure to prevent the mRNA from aberrant degradation by exonucleases, and it is important for cap-dependent translation initiation of protein synthesis (Ramanathan et al. 2016).

Following capping, the non-coding intron sequences are excised in a complex series of events known as splicing eventually resulting in a continuous coding sequence of exons **(Fig. 1)**. Splicing is carried out by the spliceosome, a large macromolecular RNA-protein complex comprised of five small nuclear RNAs (snRNAs) and many associated proteins (Hang et al. 2015). Splicing is followed by a two-step cleavage and polyadenylation reaction that releases the mRNA from RNA polymerase II **(Fig. 1)**. The assembly of the 3′-end processing machinery is closely interconnected with the factors involved in capping and pre-mRNA splicing. In metazoans, a 'tail' of around 200 adenosines is then added at the 3′ terminus in a process termed polyadenylation (Neve et al. 2017).

During the process of maturation, mRNAs are coated by a multitude of proteins and then continue to exist as messenger ribonuclear particles (mRNPs) throughout their lifespan. Some of these factors are specific for common mRNA structural elements such as the cap structure or the poly(A) tail. Others associate in a sequence-independent manner recognizing secondary or tertiary sequence elements. In the nucleus, the mRNP assembly begins with the recruitment of the cap-binding complex

(CBC) subsequently followed by exon junction complexes (EJCs), which mark spliced exons and provide a platform for various other protein-protein interactions. The transcription/export (TREX) complex and serine/arginine-rich (SR) proteins, which function in mRNA export are then recruited. The 3′ end of an mRNA is protected by the nuclear poly(A)-binding proteins (PAPBNs). The mature mRNPs are exported via the nuclear pore complexes (NPCs) into the cytosol. Nuclear export of mRNPs is carried out by adaptor proteins such as NXF1/Mex67, which usually recognize the completion of pre-mRNA processing by binding to one or more of the late processing factors (Kelly and Corbett 2009; Katahira 2015) thus closely coupling the various steps of processing and export.

In the cytoplasm, the mRNP is remodeled and the CBC and PABPN are exchanged with the cap-binding protein eIF4E and the cytoplasmic poly(A)-binding proteins (PABPCs) to protect the transcript from premature degradation by exonucleases such as the XRN1. eIF4E then associates with eIF4G and eIF4A into the eIF4F complex, which allows the initiation of translation. During translation, the mRNP is actively stripped of all associated factors within the coding region to expose the mRNA to the ribosome (Singh et al. 2015) **(Fig. 1)**. Apart from translational regulation, the mRNAs are also subject to various surveillance pathways. The best-characterized surveillance pathway in cells is the nonsense-mediated mRNA decay (NMD) pathway, which is mediated by the Upf family of proteins and serves to recognize premature stop codons (Losson and Lacroute 1979; Kervestin and Jacobson 2012). In no-go decay (NGD), the ribosome is stalled typically by strong inhibitory secondary structures (Doma and Parker 2006). The third surveillance pathway is non-stop decay (NSD) where a stop codon is absent and the ribosome stalls after encountering the poly(A) tail (Frischmeyer et al. 2002). NGD is mediated by the Dom/Hbs family of proteins whereas the NSD is mediated by the Ski complex **(Fig. 1)**. However, there is a significant overlap between these pathways, which remains poorly understood (Fourati and Graille 2014). The expression of mRNAs can also be ‘silenced’ by small interfering RNAs (siRNAs) and microRNAs (miRNAs), which are non-coding RNAs and associate with the Argonaute family of proteins. They lead to the regulation of fully or partially complementary target mRNAs (Eulalio et al. 2008).

Although there are multiple fates of the mRNA in the cytoplasm, ultimately all mRNAs are degraded **(Fig. 1)**. The mRNP is first disassembled by the action of helicases such as Upf1, which also has a major role in NMD. Another protein involved in this process is the DEAD-box RNA helicase DDX6, which is known to function as a translational repressor. This helicase plays an important role in evicting itself and other proteins from mRNAs destined for destruction (Fourati and Graille 2014).

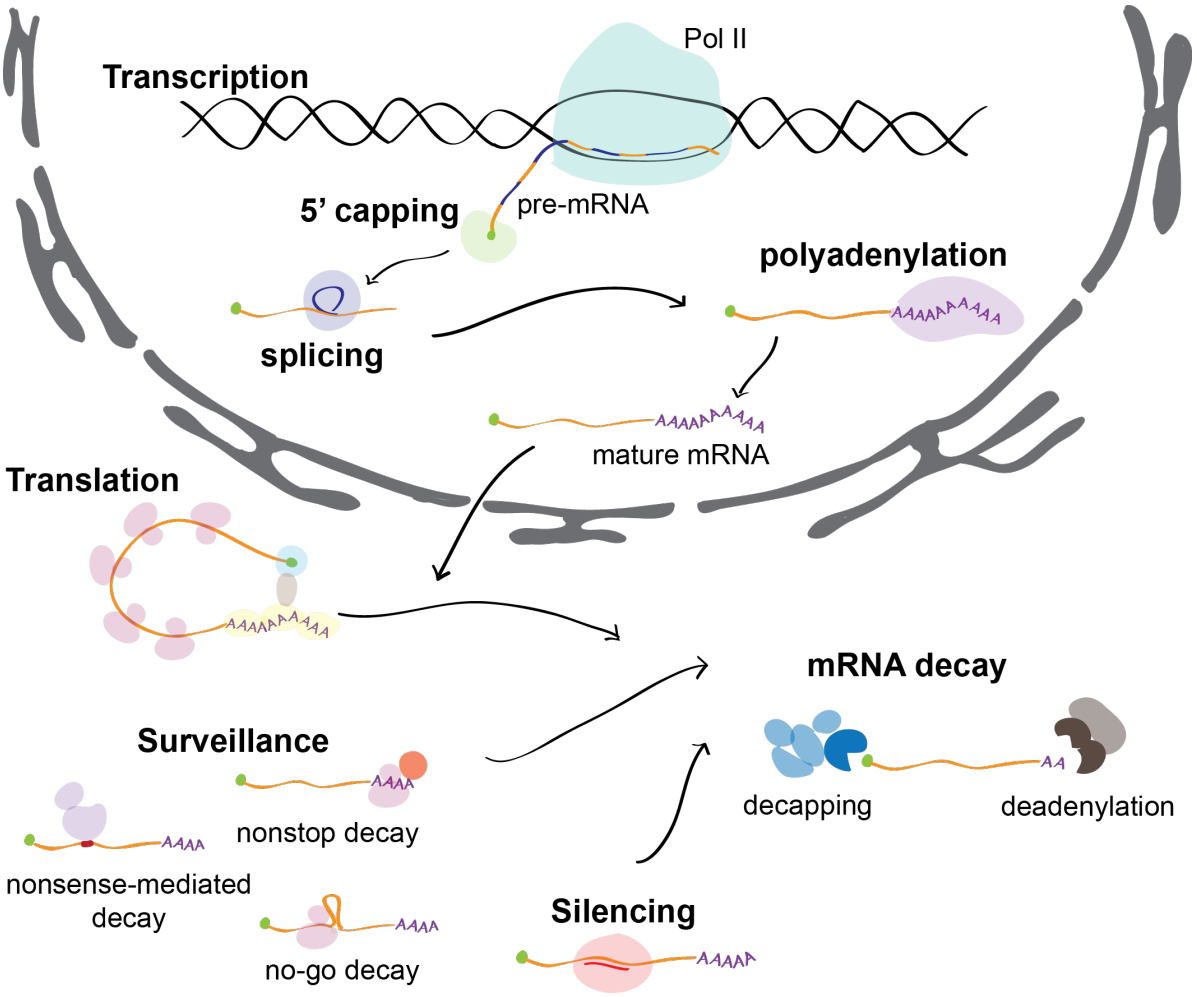


Figure 1. Gene expression in eukaryotes. An mRNA is transcribed in the nucleus as precursor mRNA. It is capped co-transcriptionally followed by splicing and polyadenylation to form a mature mRNA. This mature mRNA is exported to the cytoplasm, where it either undergoes active translation, is subject to various surveillance pathways or silenced. Ultimately all mRNAs are degraded.

2.2 Cytoplasmic mRNA decay

For the proper regulation of gene expression, it is important to maintain mRNA levels in a cell by balancing the rates of transcription and degradation. Gene expression is integral to various biological processes so any imbalance of mRNA levels can lead to developmental and neurological disorders or diseases like muscular dystrophy and cancer (Cooper et al. 2009). To initiate mRNA decay, the protective structures namely the 3′ poly(A) tail and the 5′ cap structure must be removed and subsequently, the mRNA is subject to exonucleolytic cleavage (Garneau et al. 2007) **(Fig. 2)**.

Poly(A) tail shortening is the precursor for two major cytoplasmic mRNA degradation pathways (Chen and Shyu 2011). Deadenylation is mainly mediated by the CCR4-NOT deadenylase complex (Wahle and Winkler 2013). Following deadenylation, mRNAs may be targeted for 5′–3′ decay, mediated by removal of the 5′ cap structure followed by exonucleolytic cleavage by XRN1 (Decker and Parker 1993; Dunckley and Parker 1999; Hsu and Stevens 1993). Deadenylated mRNAs can also be degraded by the exosome-mediated 3′–5′ decay pathway (Franks and Lykke-Andersen 2008; Januszyk and Lima 2014; Mitchell et al. 1997) and the remaining cap structure is hydrolyzed by the scavenger decapping enzyme DcpS (Liu et al. 2002). These two pathways are conserved amongst eukaryotes and are tightly regulated **(Fig. 2)**.

2.2.1 Shortening of the poly(A) tail is a signal for degradation

The poly(A) tail of mature mRNA in the cytoplasm is bound by multiple copies of PABPC1 (Mangus et al. 2004; Muhlrad et al. 1994), which must be displaced for deadenylation to occur. Deadenylation is a processive, rate-limiting stepwise process in which the initial catalytic cycles are accomplished by the PAN2-PAN3 complex that starts shortening of the poly(A) tail **(Fig. 2)** (Brown and Sachs 1998). The catalysis is eventually taken over by the CCR4-NOT, which is the major deadenylase complex and leaves an oligo(A) tail on mRNAs **(Fig. 2)** (Tucker et al. 2001; Yamashita et al. 2005; Zheng et al. 2008). The CCR4-NOT complex is highly conserved and consists of catalytic and various non-catalytic modules docked on a central scaffolding protein NOT1. The catalytic center—the nuclease module—is formed by two catalytically active exonucleases named CAF1 and CCR4 (Tucker et al. 2001). CAF1 has a DEDD nuclease domain (Thore et al. 2003; Bianchin et al. 2005) and binds to NOT1, while

CCR4 consists of a nuclease domain of the EEP (endonuclease exonuclease phosphatase) family (Wang et al. 2010) and docks onto CAF1 via an N-terminal LRR (leucine-rich repeat) domain (Basquin et al. 2012; Dupressoir et al. 2001). Recently, it was shown that the yeast Pab1 (PABPC1 in higher eukarya) interacts with the Ccr4-Not complex to stimulate deadenylation and differentiate the activity of the two nucleases (Webster et al. 2018; Yi et al. 2018). The shortening of the poly(A) tail via the coordinated action of PABPC1 and the deadenylases marks the start of mRNA decay.

2.2.2 Exosome-mediated 3′–5′ decay

The release of PABP during the process of deadenylation allows for direct degradation of the 3′ end of mRNAs by 3′–5′ exonucleases, mainly the RNA exosome complex **(Fig. 2)**. The eukaryotic exosome is a multiprotein ribonuclease complex, which is essential for 3′–5′ mRNA degradation (Allmang et al. 1999; Mitchell et al. 1997). Structurally, the core of exosome is conserved in eukaryotes and is formed of nine proteins, six RNase PH-like proteins forming a central channel, and three other RNA-binding proteins containing S1 and KH domains (Januszyk and Lima 2014). The core is catalytically inactive and requires a catalytic subunit for activity. The second group of exosome components comprises two ribonuclease subunits, Dis3 and Rrp6, which associate with the core in different combinations, resulting in different complexes.

The exosome is also a crucial effector in NSD and plays a role in NMD and NGD by interacting with many co-factors. These co-factors modulate the exosome activity by providing structural roles and also by interaction and processing of different RNA substrates. In the cytoplasm, the exosome interacts with Ski protein complex via the conserved Ski7 protein (Halbach et al. 2012, 2013; Wang et al. 2005), which links NSD targets to the 3′–5′ decay pathway. Following the 3′–5′ catalytic activity of the exosome, the short mRNAs with remaining cap structure are metabolized by the scavenger-decapping enzyme (DCPS) **(Fig. 2)** (Liu et al. 2002). DCPS can hydrolyze capped RNA substrates less than 10 nt long and can also play a role in scavenging the m^7GDP product following 5′–3′ decay.

2.2.3 Decapping-mediated 5′–3′ decay

The removal of the poly(A) tail causes the displacement of PABPC1 remodeling the mRNP to remove factors associated with the cap namely the eIF4F complex responsible for efficient translation initiation and to recruit factors responsible for removal of the 5′ cap structure followed by 5′–3′ exonucleolytic degradation. The decapping of mRNA by DCP2 along with co-activator and enhancer proteins generates 5′-monophosphorylated mRNAs **(Fig. 2)**. These are the substrates of 5′ exonucleases like XRN1 that readily degrades the mRNA transcripts **(Fig. 2)** (Coller and Parker 2005).

Events at the 3′ end of mRNA are linked to the 5′ end in multiple ways. One is via recruitment of the like Sm (LSm) 1-7 complex to the oligo(A) tail, which interacts with PATL1. PATL1 is known to interact with multiple decapping factors and could be the molecular ‘trigger’ for the mRNA to enter the 5′–3′ decay pathway. Another is through the interactions of the CCR4-NOT complex and decapping factors with an RNA helicase such as DDX6. This interaction could be direct or mediated via a translational repressor protein such as 4E-T, which binds to the eIF4E and disrupts binding to the eIF4G thus blocking translation initiation and promoting decapping.

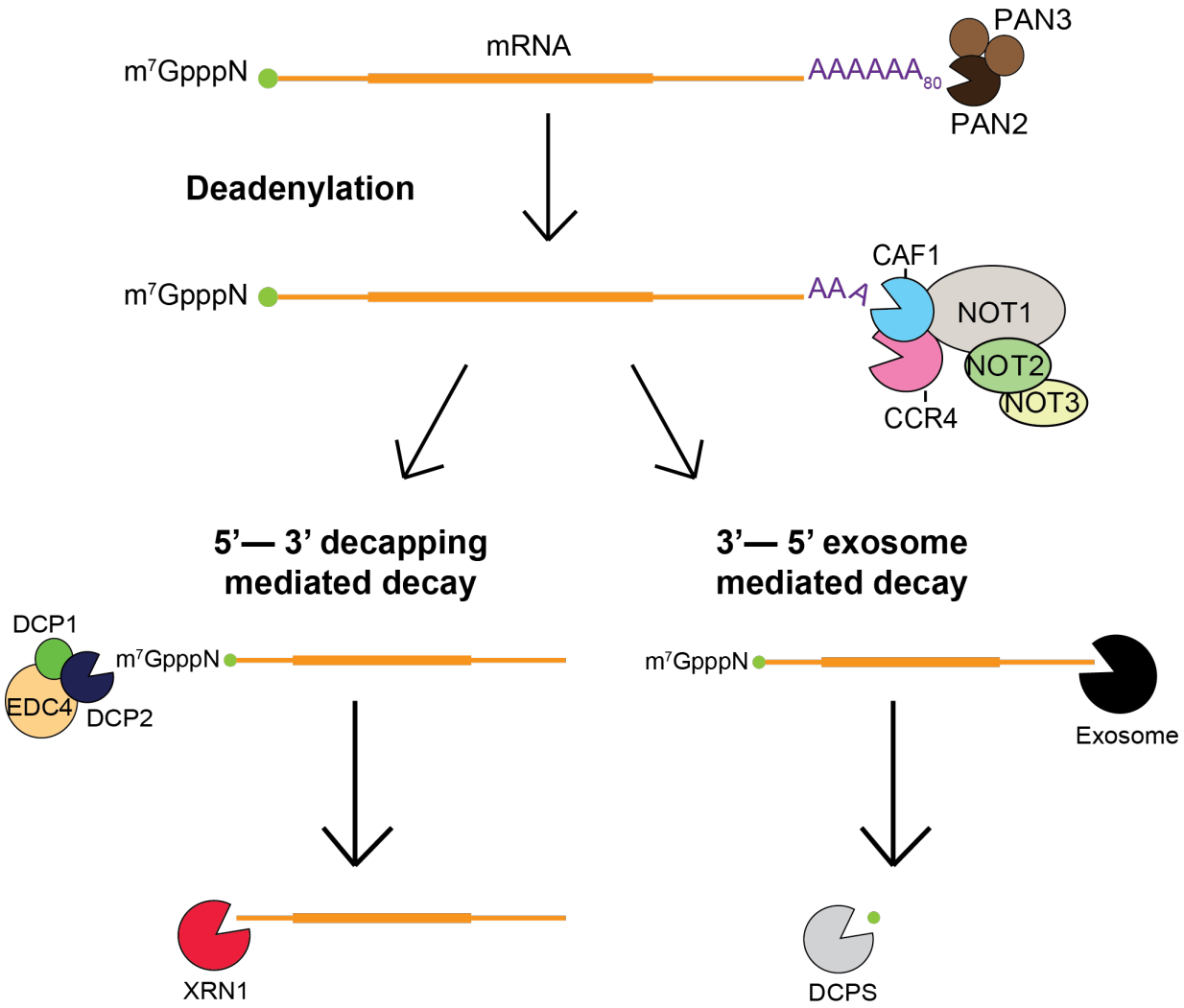


Figure 2. Cytoplasmic mRNA decay. The first step of mRNA decay is the removal of poly (A) tails by the process of deadenylation which is initiated by the PAN2-PAN3 complex and taken over by the CCR4-NOT complex. Once deadenylation is complete, the mRNA can be degraded by the 5′–3′ pathway which involves the removal of the cap structure by the decapping enzyme DCP2, followed by degradation from the 5′ end by the exoribonuclease XRN1. The other pathway is the 3′–5′ decay pathway which is exosome-mediated and starts at the 3′ end following poly (A) tail removal. Here, the cap structure at the 5′ end is degraded by the decapping scavenger enzyme DCPS, which recognizes free cap structures.

2.3 mRNA decapping

2.3.1 DCP2 — the decapping enzyme

DCP2 belongs to the Nudix family of pyrophosphatases and catalyzes the hydrolysis of the m^7GpppN cap structure at the 5′ end of an mRNA releasing m^7GDP and a monophosphorylated mRNA. The Nudix family of proteins is highly conserved and is marked by the presence of a consensus sequence consisting of $GX_5EX_7REUXEEXGU$ (where X represents any residue and U is a hydrophobic residue), which forms the catalytic site for hydrolysis of diphosphates (Bessman et al. 1996; McLennan 2006). Twenty-two Nudix hydrolases have been identified to date yet DCP2 and Nudt16 are the only ones exhibiting decapping activity (Song et al. 2010).

The DCP2 enzyme consists of an N-terminal regulatory domain (NRD) followed by a Nudix domain, which harbors the residues responsible for catalysis and a Box-B helix, which contains several lysine residues that were shown to be involved in RNA binding. The C-terminal region is highly divergent in sequence and—in yeast—contains short linear motifs (SLiMs) such as helical leucine-rich motifs (HLMs), which are responsible for mediating complex assembly via interaction with folded domains of their binding partners (Lykke-Andersen 2002; van Dijk et al. 2002; Wang et al. 2002; She et al. 2004; Fromm et al. 2012) **(Fig. 3A)**.

In yeast, the knockout of Dcp2 leads to severe decapping defects and this led to the hypothesis that Dcp2 is the major cytoplasmic decapping enzyme (Dunckley and Parker 1999). However, in a more recent study, DCP2 was shown to be differentially expressed in various tissues and, interestingly, mice depleted of DCP2 survive with no obvious phenotype. This indicates that multiple Nudix hydrolases in the cell can regulate various subsets of mRNAs.

2.3.2 Organisation of the yeast decapping machinery

2.3.2.1 The Dcp2:Dcp1 core decapping complex

In yeast, Dcp2 acts in concert with an obligate activator Dcp1 to form the conserved core of the decapping complex (Lykke-Andersen 2002; van Dijk et al. 2002; Wang et al. 2002). For a long time, Dcp1 was believed to be the decapping enzyme up until Dcp2 was characterized. Dcp1 consists primarily of an EVH1 domain **(Fig. 3A)** which

differs from typical Ena/Vasp homology 1 (EVH1) domains in that it contains an additional N-terminal α-helix (She et al. 2004, 2008), which along with a conserved loop in the EVH1 domain is responsible for interacting with the NRD of Dcp2. EVH1 domains typically have a hydrophobic groove that binds proline-rich sequences (PRSs) in protein ligands.

The first structural insight of the core complex came from a crystal structure of the Schizosaccharomyces pombe (*Sp*) Dcp2:Dcp1 decapping complex which was crystallized in two states termed ‘open’ and ‘closed’ (She et al. 2008) **(Fig. 3B)**. The NRD and Nudix domains of Dcp2 are connected by a flexible hinge region, which permits the conformational flexibility of the two domains with respect to each other **(Fig. 3B)**. It was proposed that the closed form represents the more active form of the enzyme (Deshmukh et al. 2008; She et al. 2008) by the formation of a composite active site between the NRD and Nudix domains. NMR studies showed that the transition between the open and closed states is necessary for efficient decapping. It was also shown biochemically that Dcp1 along with the NRD of Dcp2 enhanced the decapping activity of the Nudix domain by ~1,000 fold (Floor et al. 2010). However, the residues involved in RNA binding are not clustered on the same face of the decapping complex (Floor et al. 2012). Collectively, these observations suggest that there must exist an additional, active state of the decapping complex, which would likely be conformationally distinct from the previously observed open and closed states.

2.3.2.2 Enhancers of decapping promote the activity of the core complex

The process of decapping also involves various enhancers of decapping (EDCs). Some enhancers are species-specific such as Edc1 and Edc2 in yeast and EDC4 in metazoans. These enhancers work directly by interacting with either Dcp1 and/or Dcp2 and stimulate the basal decapping activity (Arribas-Layton et al. 2013). However, the molecular basis of the mechanism of action of how these enhancers act remains poorly understood. Edc1 and Edc2 are intrinsically disordered proteins and have PRSs **(Fig. 3A)** which interact directly with the EVH1 domain of Dcp1. In *Saccharomyces cerevisiae* (*Sc*) it was shown that the PRS and a short region just N-terminal to it are necessary and sufficient to enhance the decapping activity of the Dcp2:Dcp1 core complex (Dunckley et al. 2001; Schwartz et al. 2003; Borja et al. 2011). Edc3 is another enhancer of decapping protein which in yeast, binds directly to the first HLM

in Dcp2 and enhances activity. Edc3 contains an N-terminal divergent Lsm (Sm-like) domain, which binds Dcp2 (Rajyaguru et al. 2012), a central FDF domain, and a C-terminal YjeF-N domain, which self-interacts and promotes Edc3 dimerization **(Fig. 3A)** (Decker et al. 2007; Ling et al. 2008). It is a component of the decapping machinery that can both generally affect the activity of the decapping enzyme and be recruited for the regulation of specific mRNAs (Harigaya et al. 2010). Another similar protein Scd6, shown to be involved in translation repression in yeast (Nissan et al. 2010) also binds to the HLMs in Dcp2 in competition with Edc3. This points to a network of competing interactions between enhancers that can modulate Dcp2:Dcp1 decapping activity and cellular localization (Fromm et al. 2012).

More recently it was shown that the C-terminal region of Dcp2 contains both positive and negative regulatory elements. Positive regulation of the decapping enzyme is mediated by the binding of specific decapping activators to the Dcp2 C-terminal domain and the likely formation of distinct complexes that control the substrate specificity and final activation of the decapping enzyme. Negative regulation of the decapping enzyme by a cis-inhibitory element in Dcp2 is important as it was shown that elimination of the inhibitory element results in constitutive activation of the decapping enzyme and loss of the enzyme's specificity for substrate mRNAs (He and Jacobson 2015; Paquette et al. 2018). Over the past three years, multiple crystal structures of the core decapping complex in yeast with one or more enhancers with or without cap analogs have been reported, improving our understanding of cap recognition and activation of the core complex by its enhancers (Mugridge et al. 2018; Valkov et al. 2016; Wurm et al. 2016; Charenton et al. 2016). A structure of the complex with an RNA substrate is needed to further advance our understanding of the decapping machinery in yeast.

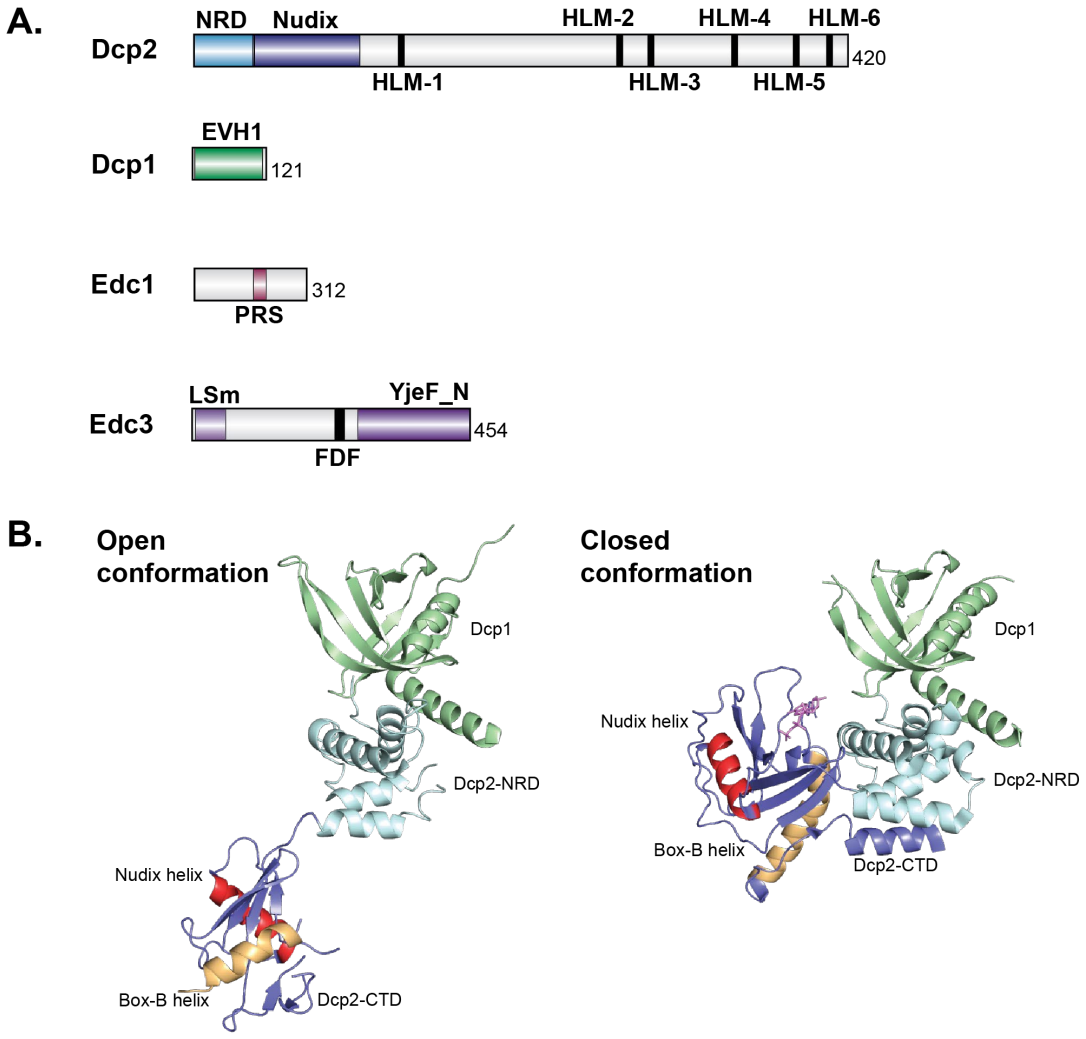


Figure 3. (A) Domain organization of *Schizosaccharomyces pombe* (*Sp*) proteins. The numbers at the ends of the schematic represent the last amino acid. Colored regions represent structured domains. Dcp2 has a conserved N-terminal regulatory domain (NRD) and a catalytic Nudix domain. The Dcp2 C-terminus is highly divergent and interspersed with six helical leucine-rich motifs (HLMs). Dcp1 contains an Ena/Vasp homology domain1 (EVH1). Edc1 has a proline-rich sequence (PRS) and interacts with Dcp1. Edc3 contains an N-terminal divergent Lsm (Sm-like) domain, which binds Dcp2, a central FDF domain, and a C-terminal YjeF-N domain. **(B)** Cartoon representations of the first structures of the decapping complexes (PDB: 2QKM). The Dcp1 (green), Dcp2 N-terminal regulatory domain (NRD) (cyan) and Dcp2 catalytic Nudix domain (blue) adopt different relative orientations. The Nudix helix and the Box-B helix are highlighted in red and yellow respectively.

2.3.3 The human decapping interactome

2.3.3.1 Reorganisation of the human decapping network

In metazoans, DCP1 and DCP2 do not assemble into a stable complex, unlike in yeast, and their direct interaction was proposed to be stabilized by an additional metazoan-specific factor EDC4. EDC4 contains an N-terminal WD40 domain and a C-terminal α-helical domain connected by a serine-rich linker (Fenger-Grøn et al. 2005; Yu et al. 2005; Xu et al. 2006; Jinek et al. 2008). The C-terminal region is divided into a proximal and distal region. The proximal region is predicted to mediate self-association and the distal part is structurally related to ARM/HEAT repeat proteins (Xu et al. 2006; Jinek et al. 2008).

DCP1 is known to interact with three of the enhancers of decapping: EDC3, EDC4, and PNRC2. DCP1 interacts with the LSm domain of EDC3 via its HLM motif. The FDF motif on EDC3 is known to interact with the C-terminal Rec-A domain of DDX6. The other interacting partner of DCP1 is PNRC2. PNRC2 is an intrinsically disordered protein. It has a proline-rich sequence (PRS), which binds to EVH1 domains typically present on DCP1 (Lai et al. 2012). The EVH1 domain of DCP1 interacts with the PRS of PNRC2. DCP1 is highly phosphorylated and it was proposed that phosphorylation is important for its localization in cells *in vivo* (Aizer et al. 2013; Rzeczkowski et al. 2011).

In the human decapping network, EDC3 interacts with the HLM on DCP1 whereas in yeast it interacts with multiple HLMs on Dcp2 (Fromm et al. 2012; Tritschler et al. 2008). The yeast Edc3 acting in concert with Edc1 fully activates the core Dcp2:Dcp1 complex (Mugridge et al. 2018). It was proposed that human EDC3 may provide a binding platform for DCP1, DCP2, and EDC4. Self-association via its C-terminal domain in metazoa has the potential to bring in close proximity many factors necessary for efficient decapping. EDC3 also interacts with DDX6, an indirect enhancer of decapping, which acts by repressing translation. In order to derive a comprehensive understanding of decapping, it is necessary to study it in several species. Metazoan-specific adaptations such as the requirement for the EDC4 scaffold, the acquisition of the trimerization domain in DCP1, and the evolution of the C-terminal extension of DCP2 points to a complete rewiring of this network of interactions and a significant mechanistic divergence from yeast.

2.3.3.2 General decapping activators

General enhancers such as the RNA helicase DDX6, the LSm 1-7 complex, PATL1 are thought to facilitate decapping by inhibiting translation and/or promoting conformational rearrangements that increase the accessibility of DCP2 to the mRNA cap structure (Kshirsagar and Parker 2004; Jonas and Izaurralde 2013).

DDX6 is an ATP dependent RNA helicase with two RecA domains belonging to the DEAD-box family of proteins. These proteins are characterized by having nine distinct motifs that are involved in RNA binding, ATP-binding and its hydrolysis (Linder and Jankowsky 2011). In addition to helicase activity, it was proposed that DDX6 is responsible for promoting mRNP rearrangements to inhibit translation and promote decapping (Tritschler et al. 2007). DDX6 binds to motifs embedded in low complexity regions of other decapping factors (Tritschler et al. 2007, 2009a). DDX6 can directly interact with various decapping factors like EDC3, PATL1, and Scd6 in yeast. It also interacts in a mutually exclusive manner with NOT1 which indicates that it can be involved in a hand-off from deadenylation to decapping. Another interaction partner of DDX6 is 4E-T, a translational repressor which is also a potential link between deadenylation and triggering decapping (Ozgur et al. 2015).

Another protein complex that plays a central role in connecting decapping and deadenylation is the PATL1:LSm1-7 complex. In yeast, Pat binds to Dcp1, Dcp2, Dhh1 (DDX6 in human) and Xrn1. It has also been shown to interact with the Ccr4-Not complex in Drosophila (Haas et al. 2010; Pilkington and Parker 2008; Ozgur et al. 2010; Sharif and Conti 2013; Sharif et al. 2013; Bouveret et al. 2000). Pat proteins contain a conserved N-terminal sequence, a proline-rich region, a Mid domain and a C-terminal domain (Pat-C). Pat-C along with the mid and P-rich regions mediate interactions with mRNA decapping factors. LSm 1-7 has seven Sm-like proteins and binds to the mid-region of Pat (Wilusz and Wilusz 2005).

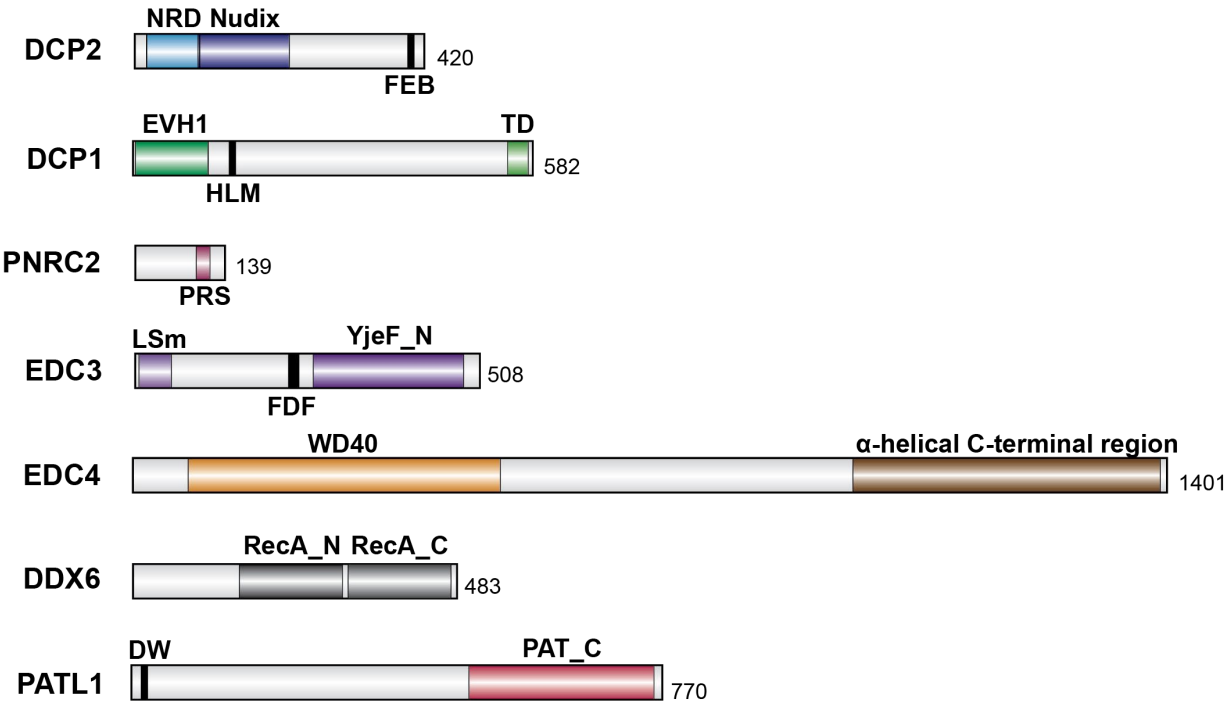


Figure 4. Domain organization of human proteins. The numbers at the ends of the schematic represent the last amino acid. Colored regions represent structured domains. DCP2 has an NRD and a catalytic Nudix domain similar to yeast. The C-terminal region is shorter than in yeast and binds to EDC4 via the FEB (Phenylalanine rich EDC4 binding motif). DCP1 contains an EVH1 domain and a long unstructured C-terminal region with a trimerization domain (TD) at its end through which it binds to EDC4. PNRC2 has a proline-rich sequence (PRS) and interacts with DCP1. EDC3 contains an N-terminal divergent Lsm (Sm-like) domain, which binds to the HLM in DCP1, a central FDF domain, and a C-terminal YjeF-N domain. EDC4 has a WD40 domain at its N-terminus which binds to DCP1 and an α-helical C-terminal region which binds to DCP2. DDX6 has two RecA domains: RecA_N and RecA_C. It binds to the FDF motif of EDC3 via its RecA_N domain. PATL1 has an N-terminal DW motif followed by a P/Q rich region. The C-terminal region is called PAT_C and mediates interactions with DCP2, EDC4 and LSm 1-7.

2.4 P-bodies

Processing bodies (PBs) are a class of ribonucleoprotein (RNP) granules that persist during cellular homeostasis and are enriched for RNA processing and degradation factors (Eulalio et al. 2007; Parker and Sheth 2007). These granules exhibit characteristics of liquid droplets, and their formation follows the principles of classical liquid-liquid phase separations (Brangwynne et al. 2009; Li et al. 2012b; Weber and Brangwynne 2012). Low-affinity interactions and multivalency of interactions are sufficient to induce phase separation into liquid droplets. These properties are common in RNA-binding proteins (Kato et al. 2012; Li et al. 2012a). Furthermore, disordered, low-complexity protein regions are capable of inducing phase transitions, forming hydrogels (Han et al. 2012; Kato et al. 2012). PBs are enriched in various proteins related to RNA metabolism and various proteins were found associated with PBs using immunostaining or fluorescence-tagging approaches. These broadly fall into the categories of mRNA decay, translational control, and RNA interference. Interestingly, while the 5′–3′ decay pathway is well represented, components of the 3′–5′ exosome pathway have not been identified in PBs.

3. AIMS AND OBJECTIVES

The objective of my doctoral studies was to improve our mechanistic understanding of mRNA decapping. The first structural insights were obtained from crystal structures of the fission yeast decapping complex comprising Dcp2 and its obligate activator Dcp1 (She et al. 2008). However, the mechanism of activation of Dcp2 by Dcp1 remained unclear. How various enhancers acted in concert with Dcp2 and/or Dcp1 and further stimulated decapping was also unknown. To address this, I set out to crystallize the core decapping complex with one or more of its enhancers to elucidate the mechanism of activation. This is described in **Section 4.1**.

In the subsequent phase of my research, I applied structural and biochemical methods to investigate decapping regulation in higher eukaryotes. The principal aim was to biochemically reconstitute and purify an intact metazoan decapping complex in quantities sufficient to pursue structural studies using crystallography or single-particle analysis by electron microscopy.

The network of interactions between DCP2 and the other factors diverged from yeast to human. Self-association is ubiquitous among many decapping factors in higher eukaryotes and this creates an additional layer of complexity. Our understanding of the mechanistic impact of this oligomerization on the regulation of DCP2 activity is very fragmented and at the outset, my motivation was to investigate the complexity of decapping networks with critical regions intact and in the context of full-length proteins.

Human DCP1 and DCP2 interact directly but they do not assemble into stable complexes, unlike in yeast, and their direct interaction requires stabilization by an additional metazoan-specific enhancer EDC4, which was proposed to serve as a scaffold for complex assembly (Chang et al. 2014). Structural information is only available for the distal part of the C-terminal region of EDC4 from Drosophila (also called Ge-1) (Jinek et al. 2008). Other factors include EDC3, known to interact with DCP1 in metazoans, and had been shown to enhance decapping activity in yeast (Kshirsagar and Parker 2004). PNRC2 is an intrinsically disordered protein and is a functionally equivalent protein of Edc1 in yeast. It has been shown that a proline-rich sequence and a region just N-terminal to it is necessary and sufficient to enhance

decapping (Lai et al. 2012). These enhancers are known to work directly by interacting with either DCP1 and/or DCP2 and stimulate the basal decapping activity (Arribas-Layton et al. 2013). Other enhancers such as the RNA helicase DDX6, the LSm 1-7 complex, Lsm14A, PATL1 are thought to facilitate decapping by inhibiting translation and/or promoting conformational rearrangements that increase the accessibility of DCP2 to the mRNA cap structure (Kshirsagar and Parker 2004; Jonas and Izaurralde 2013). The interactions between these factors are between structured domains and short linear motifs embedded in low complexity regions. However, the molecular basis of the mechanism of action of how these enhancers act remains poorly understood. Very little information is available so far despite considerable efforts to determine the networking and assembly of the human decapping factors. This limitation is due to the fact that these proteins have never been purified and reconstituted *in vitro*, with their unstructured regions intact.

I have established production and purification strategies for all principal human decapping factors as full-length proteins with key regulatory low-complexity regions intact. I have also reconstituted several multiprotein complexes in quantities and with purity to enable biochemical and structural studies. I then performed an *in vitro* biochemical screen of DCP2 decapping activity in the presence of one or several activators. The results of these studies are described in **Sections 4.2 and 4.3**.

4. RESULTS

During the course of my doctoral studies, I investigated mechanisms of messenger RNA decapping in yeast and metazoa using structural and biochemical approaches. The initial focus of my research was on the characterization of the mechanism of activation of the fission yeast Dcp2:Dcp1 decapping complex by an intrinsically disordered protein Edc1. I then purified and characterized decapping factors from *C. elegans,* which gave us important initial insights into the metazoan decapping network. The principal focus of the second half of my doctoral work was directed toward the *in vitro* characterization of human decapping factors and the reconstitution of multiprotein regulatory complexes. To facilitate this, I deployed a baculovirus-based insect cell expression system for producing complex proteins with extended unstructured regions. I have established and optimized the production and purification conditions of all principal human decapping factors as full-length proteins and a number of multisubunit complexes with crucial regulatory disordered regions fully intact. I then conducted an unbiased and comprehensive biochemical screen of regulation of the activity of full-length human DCP2 enzyme by all key human decapping factors. Collectively, the data presented in this thesis provides a comprehensive biochemical *in vitro* framework that has improved our mechanistic understanding of the regulation of decapping machinery across species.

4.1 Structural studies of decapping activation in yeast

Experimental data and detailed experimental procedures described in this chapter are available in the attached manuscript Valkov, Muthukumar, Chang *et al*., (2016).

4.1.1 Crystallization of the core decapping complex with Edc1

Yeast Dcp2 is not intrinsically active and requires an obligate coactivator Dcp1. Together these assemble into a stable binary complex. Other enhancer factors further stimulate the activity of this core Dcp2:Dcp1 complex. The first question I addressed was how the enhancer Edc1 acts in concert with the core complex and stimulates decapping. The Dcp2:Dcp1 complex from fission yeast was crystallized in two distinct conformations: an open conformation where Dcp1 is positioned well away from the Nudix domain of Dcp2; and, a closed conformation, which is a more compacted form (She et al. 2008). Edc1 is an intrinsically disordered protein and has been shown in

Saccharomyces cerevisiae that a proline-rich sequence (PRS), which binds to Dcp1 together with a small region just N-terminal to it, is necessary and sufficient to enhance decapping activity (Borja et al. 2011). However, the molecular mechanism of stimulation remained unknown though it was proposed that the closed conformation of Dcp2:Dcp1 is stabilized by Edc1 and results in synergistic activation of the enzyme (Floor et al. 2012).

To reconstitute the Dcp2:Dcp1 complex with an Edc1 region comprising residues necessary for decapping activation, I first identified an Edc1 ortholog in *Schizosaccharomyces pombe* through the analysis of sequence conservation. I delineated a motif of just 26 residues (S155–S180) comprising the proline-rich sequence and a region just N-terminal to it and termed it the decapping activating motif (DAM) **(Fig. 5A)**. I then reconstituted the Dcp2:Dcp1 decapping complex with a synthetic peptide corresponding to the Edc1 DAM and obtained crystals using a high-throughput crystallization screening approach. Following some optimization, the crystals diffracted X-rays to 1.6 Å and the phases were obtained by molecular replacement.

Following rebuilding and refinement of the crystallographic model, I could observe convincing electron density for the entire Edc1 DAM peptide. The N-terminal residues of Edc1 DAM were observed to fold into a β-strand to augment and extend the β-sheet of the Dcp2 Nudix domain **(Fig. 5B)**. The structure of the Dcp2:Dcp1:Edc1 ternary complex revealed two striking features. First, the Nudix domain of Dcp2 was observed to undergo a 120° rotation as compared to the structure in the absence of the Edc1 peptide **(Fig. 5B and C)**. Second, this rotation positions the highly basic Box-B helix at the opening of a newly formed cleft between the Nudix and the Dcp1 EVH1 domains. This results in the clustering of positively charged residues forming a continuous charged surface leading to the active site which can potentially bind RNA. **(Fig. 5D)**.

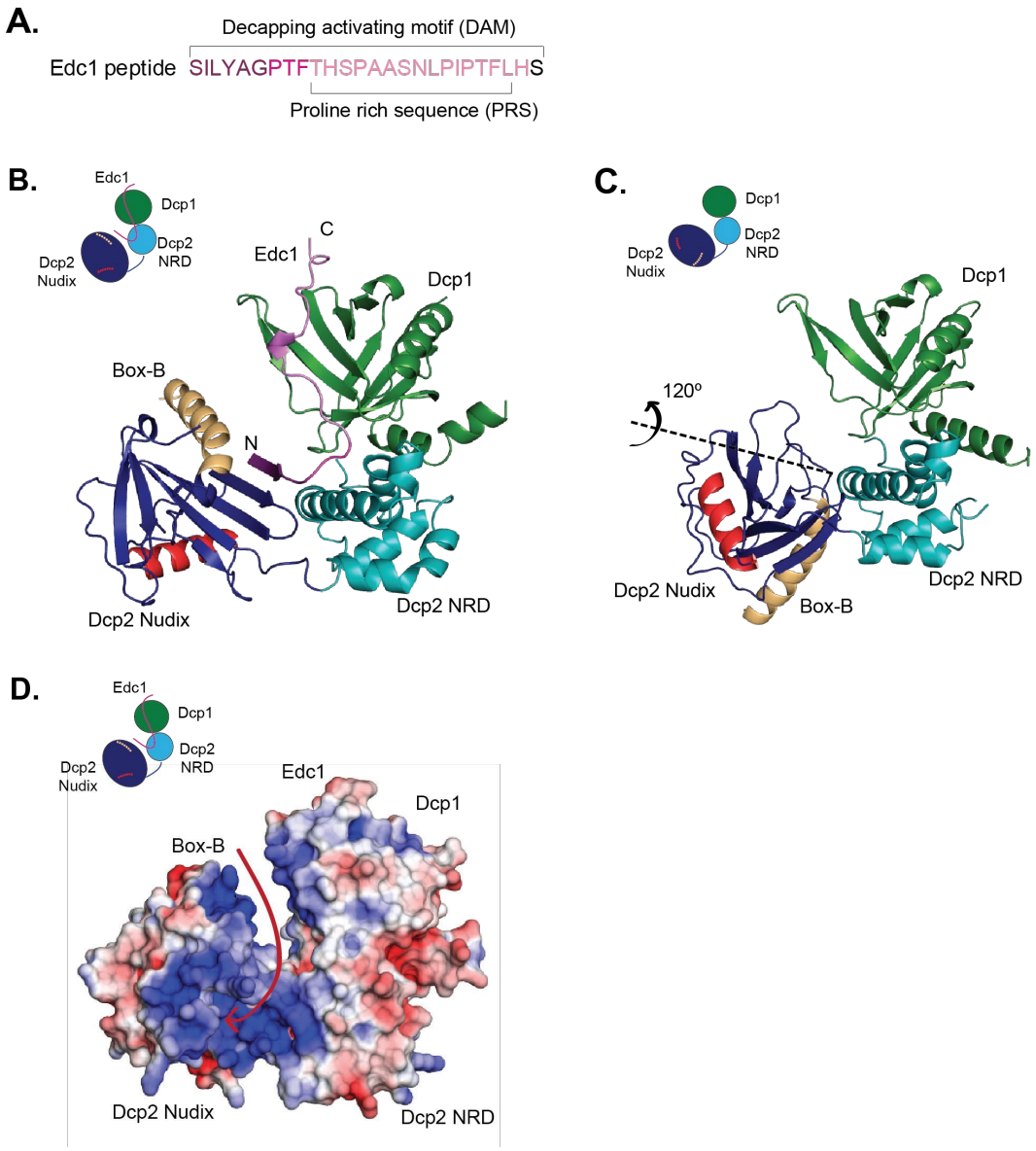


Figure 5. **(A)** The *Sp* Edc1 peptide showing the PRS which interacts with Dcp1 and the DAM which includes the Dcp2 binding region. **(B)** Structure of the *Sp* Dcp2:Dcp1:Edc1 complex showing the Edc1 peptide contacting all three structured domains of the core complex. **(C)** Structure of the *Sp* Dcp2:Dcp1 complex indicating the repositioning of the Nudix domain to bring it to its active state as seen in (B). **(D)** Surface charge distribution of the *Sp* Dcp2:Dcp1:Edc1 complex shows a potential RNA-binding channel leading to the active site. (Figures adapted from Valkov et. al., 2016)

4.1.2 Increased RNA binding increases the activity of the trimeric complex

The biased distribution of electrostatic charge observed in the conformation stabilized by Edc1 indicated that one explanation for the stimulation of Dcp2:Dcp1 activity may be that Edc1 traps the complex in a conformation that is more optimal for RNA binding and positioning mRNA more optimally for decapping.

The N-terminal intrinsically disordered region of Edc1 can directly bind to RNA (Schwartz et al. 2003). Using electrophoretic mobility shift assays (EMSAs), I observed that in the presence of the Edc1 DAM, the Dcp2:Dcp1 complex interacts stably with RNA **(Fig. 6A, lanes 6-8)**. The PRS motif alone is not sufficient to promote stable binding of Dcp2:Dcp1 to the RNA **(Fig. 6A, lane 2 vs lane 5)**. To validate the proposed RNA-binding surface observed in the structure of the ternary complex **(Fig. 5D)**, I mutated the residues on the Box-B helix of Dcp2, which have been previously shown to be important for substrate binding and confirmed that RNA binding is reduced. I then tested two previously uncharacterized conserved Dcp1 lysines (K44 and K47) and observed that these Dcp1 mutants significantly reduce the affinity of Dcp2:Dcp1 complex toward RNA **(Fig. 6B)**. These Dcp1 mutants still assembled into a complex with Dcp2 as verified by pulldowns indicating that the Dcp1 fold was not affected **(Fig. 6D)**.

Next, I tested whether the Edc1 DAM used in this study is sufficient to enhance the activity of the core Dcp2:Dcp1 complex. In *S. cerevisiae* it has been shown that Edc1 does not have intrinsic decapping activity but can enhance the activity of the Dcp2:Dcp1 complex *in vitro*. I observed that the Edc1 DAM peptide significantly enhances the activity of the decapping complex **(Fig. 6C lanes 5-7 vs lanes 11-13)** whereas the PRS alone does not **(Fig. 6C lanes 5-7 vs lanes 8-10)**. I then mutated residues on Dcp1 (Y36 and W107), which disrupt Edc1 binding and tested this in pulldowns with MBP-tagged Edc1 DAM **(Fig. 6E)**. In decapping assays, these Dcp1 mutants in the presence of the Edc1 DAM did not activate Dcp2:Dcp1 **(Fig. 6C lanes 5-7 vs lanes 14-19)**. Taken together, the structure of the Dcp2:Dcp1:Edc1 ternary complex reveals mechanistic insights into how an enhancer such as Edc1 stimulates the activity of the core decapping complex by stabilizing it in a conformation that improves interactions with the RNA. This study also provided the first explanation for the role of Dcp1 in enhancing Dcp2 activity.

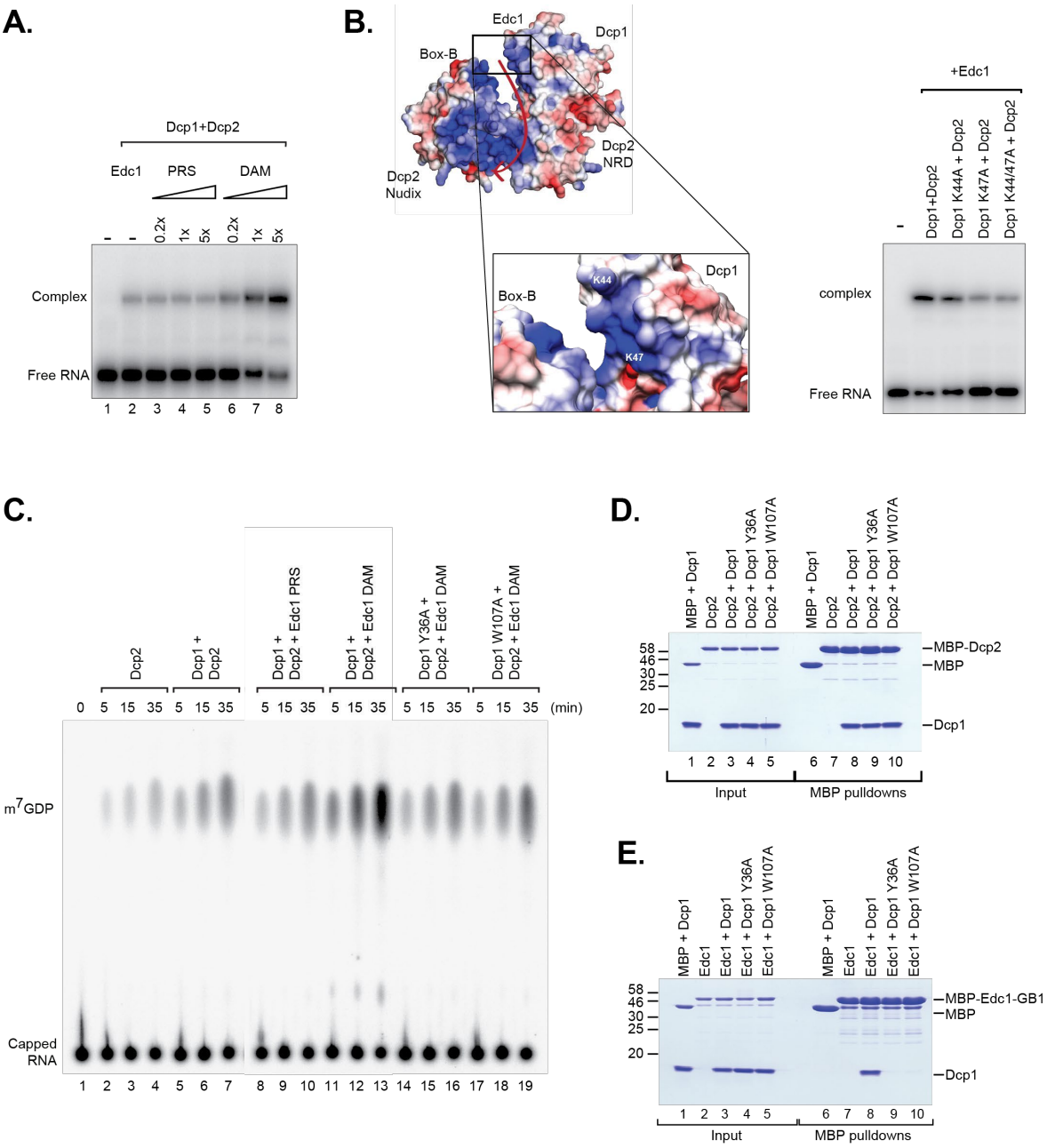

A.
Dcp1+Dcp2
Edc1
PRS
DAM
-
-
0.2x
1x
5x
0.2x
1x
5x
Complex
Free RNA
1 2 3 4 5 6 7 8
B.
Edc1
Dcp1
Box-B
Dcp2
NRD
Dcp2
Nudix
Box-B
K44
Dcp1
K47
+Edc1
-
Dcp1+Dcp2
Dcp1 K44A + Dcp2
Dcp1 K47A + Dcp2
Dcp1 K44/47A + Dcp2
complex
Free RNA
C.
Dcp2
Dcp1 + Dcp2
Dcp1 + Dcp2 + Edc1 PRS
Dcp1 + Dcp2 + Edc1 DAM
Dcp1 Y36A + Dcp2 + Edc1 DAM
Dcp1 W107A + Dcp2 + Edc1 DAM
0 5 15 35 5 15 35 5 15 35 5 15 35 5 15 35 5 15 35 (min)
m7GDP
Capped RNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
D.
MBP + Dcp1
Dcp2
Dcp2 + Dcp1
Dcp2 + Dcp1 Y36A
Dcp2 + Dcp1 W107A
58
46
30
25
20
MBP-Dcp2
MBP
Dcp1
1 2 3 4 5 6 7 8 9 10
Input
MBP pulldowns
E.
MBP + Dcp1
Edc1
Edc1 + Dcp1
Edc1 + Dcp1 Y36A
Edc1 + Dcp1 W107A
58
46
30
25
20
MBP-Edc1-GB1
MBP
Dcp1
1 2 3 4 5 6 7 8 9 10
Input
MBP pulldowns

Figure 6. **(A)** Electrophoretic mobility shift assay of the *Sp* Dcp2:Dcp1 complex with increasing concentrations of the Edc1 PRS or DAM peptides. **(B)** A close-up view of the position of Dcp1 residues involved in RNA binding and an EMSA with single and double mutants of Dcp1 in complex with Dcp2. **(C)** Decapping assay of the core decapping complex with the PRS and DAM of Edc1 indicating that the DAM is necessary for decapping activation. Decapping assays of the Dcp2:Dcp1 complex were also carried out with mutations in Dcp1 which disrupt binding to Edc1 DAM. These show no increase in decapping activity and is comparable to Dcp2:Dcp1 alone. **(D)** MBP pulldowns to show that the Dcp1 mutants described in panel (C) still form a complex with Dcp2. **(E)** MBP pulldowns to show that the Dcp1 mutants described in panel (C) can no longer form a complex with Edc1 DAM.

4.2 Characterization of decapping factors in *Caenorhabditis elegans*

Experimental procedures described in this chapter are available in the **Appendix**.

In metazoa, DCP1 and DCP2 do not form a stable complex and the interaction was proposed to be stabilized by an additional metazoan-specific enhancer EDC4. Coimmunoprecipitation experiments using human cells revealed that the trimerization domain of DCP1 interacts with the WD40 domain of EDC4 while the FEB motif of DCP2 binds to the C-terminal α-helical region of EDC4 (Chang et al. 2014). Based on this data, it was proposed that in metazoa DCP1 and DCP2 assemble into a ternary complex with EDC4, which is then the active decapping complex. However, this was not verified with recombinant purified proteins.

In order to test this model of the assembly of the metazoan decapping complex, I chose the decapping factors from *C. elegans*. Sequence conservation analysis indicated that in this organism EDC4 has a significantly shorter low-complexity region compared to the human ortholog, which I hypothesized would facilitate the production of this protein in an intact, recombinant form **(Fig. 7A vs 4A)**. EDC4 isolated from human cells using immonopurification was shown to exhibit a molecular weight of over 2 MDa in complex with other factors (Fenger-Grøn et al. 2005). Analysis of the protein sequence also suggested the presence of a coiled-coil motif preceding the C-terminal α-helical region.

I then delineated a minimal fragment of the predicted coiled-coil in *Ce* EDC4 and then crystallized a synthetic peptide comprising the 33 residues (520–552) in this motif. The crystals diffracted X-rays to 1.4 Å resolution and the phases were obtained by molecular replacement using a polyalanine helix as a search model. Following rebuilding and refinement of the model, I could observe 27 residues in the electron density (523-549) and assign a high-symmetry cubic space group $P4_132$. The structure revealed a tetrameric arrangement of the helices in parallel orientation **(Fig. 7B)**. Given that the crystallized coiled-coil motif is likely the principal mediator of the oligomerization of EDC4, it seems likely that EDC4 may oligomerize as a dimer of dimers or a tetramer in a parallel orientation. The functional significance of this remains unknown.

Next, I set out to map the minimal interacting regions between DCP2 and EDC4. It was shown in cells that DCP2 associates stably with EDC4 via its FEB motif and the interaction with DCP1 can only be detected in the presence of EDC4. Using sequence conservation analysis, I identified an analogous motif in *C. elegans* and then purified it while fused to a small structured domain, termed macrodomain (Wild and Hothorn 2017), to enhance its stability in solution. The principal aim was to map the motif to the minimal region of the EDC4 C-terminal region necessary for binding to reconstitute a complex for crystallization. I made several truncation constructs from the N and the C terminal regions of EDC4 and by *in vitro* pulldown assays could then map the fragments of EDC4 required for stable interaction with DCP2 **(Fig. 7C)**. The pulldown analysis further indicated that the coil-coiled region of EDC4 is not required for DCP2 binding. A synthetic peptide of the DCP2 FEB motif was highly insoluble and crystallization attempts with purified proteins were unsuccessful, however.

Nonetheless, these preliminary results with the *C. elegans* factors provided some interesting insights into the assembly of the metazoan decapping complexes, which, in turn, were invaluable in developing effective strategies to reconstitute and study the human decapping factors.

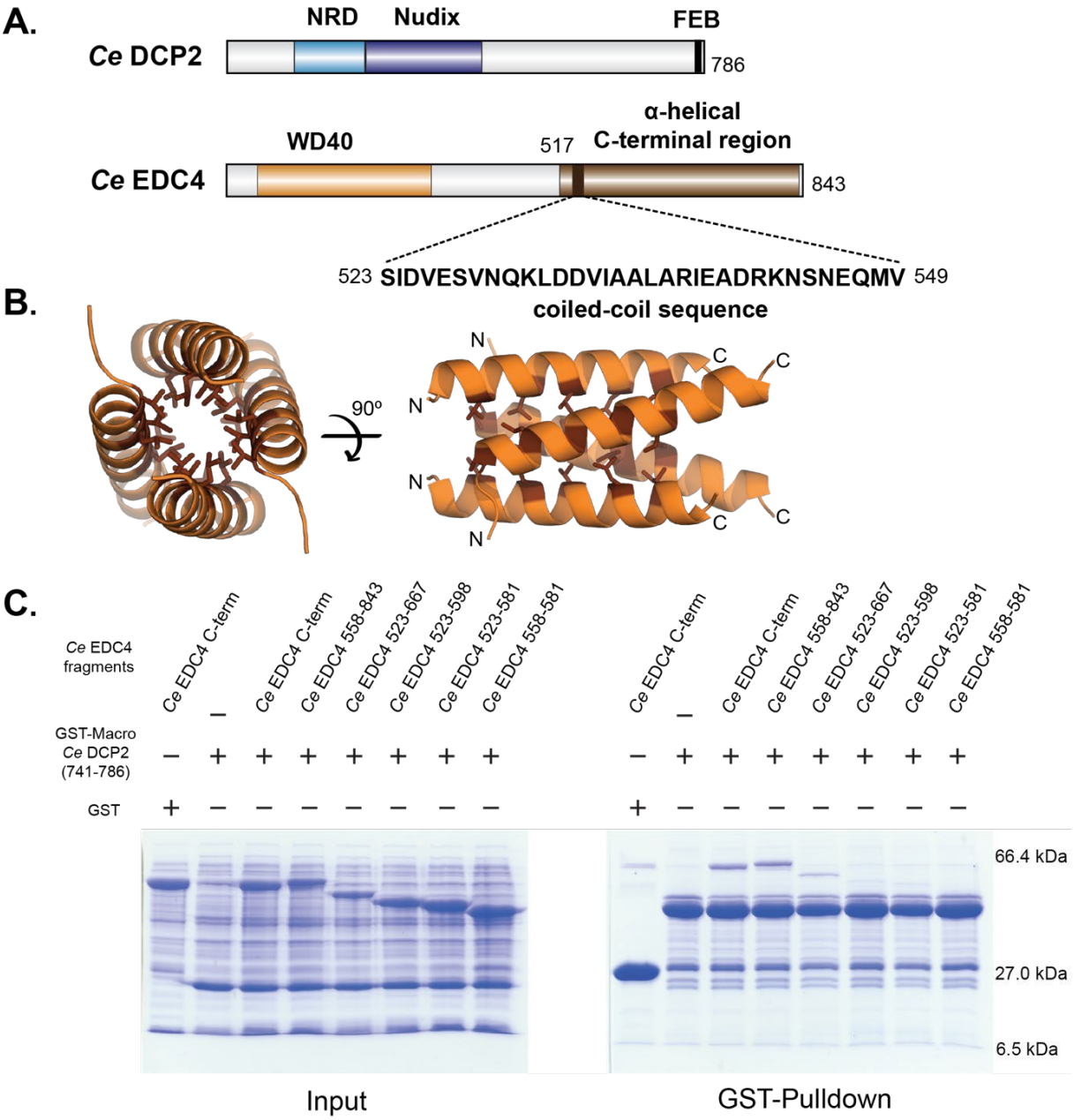


Figure 7. (A) Domain organization of *C. elegans* DCP2 and EDC4. Colored regions represent structured domains. The numbers represent amino acid positions. The sequence of the coiled-coil motif used for crystallization is indicated. **(B)** Crystal structure of the coiled-coil motif of *Ce* EDC4 (front and side views). The hydrophobic residues and the N and C termini are indicated. **(C)** GST pulldown of various C-terminal fragments of *Ce* EDC4 with a peptide of *Ce* DCP2 to map the shortest interacting region.

4.3 Reconstitution and characterization of human decapping interactome

Experimental data described in this chapter will form the basis for a manuscript, currently in preparation (Muthukumar and Valkov).

4.3.1 Production of recombinant, full-length human decapping factors

Biochemical characterization of the human decapping interactome with purified full-length proteins has never previously been accomplished due to the significant challenges in producing many of these factors with their low-complexity regions intact. Most of our current knowledge about the molecular interactions between the human decapping factors is derived from crystal structures of domains with interacting peptides/motifs and validated in immunoprecipitation experiments in cells (Lai et al. 2012; Tritschler et al. 2009a). For a comprehensive biochemical study, it is, however, necessary to produce the full-length human decapping factors as the short linear motifs, which crucial for mediating protein-protein interactions, are interspersed in their low-complexity regions (Jonas and Izaurralde 2013). I developed a recombinant approach to produce an almost complete repertoire of human decapping factors as full-length proteins and stable multiprotein complexes. For this, I employed the MultiBac (Bieniossek et al. 2012; Sari et al. 2016) and the biGBac (Weissmann et al. 2016; Weissmann and Peters 2018) baculovirus-based insect cell expression systems.

The initial efforts were directed toward the production and purification of the full-length DCP2:DCP1:EDC4 as a ternary complex. Following affinity purification of the StrepII-tagged EDC4, I observed that DCP1 and DCP2, although produced in significant quantities in cells, copurified in very low, substoichiometric amounts suggesting that the three factors do not assemble into a stable complex (data not shown). When these factors were produced in isolation, EDC4 could be purified in a single affinity capture step using a C-terminal StrepII tag. Strikingly, and in contrast to expectations, the full-length EDC4 ($EDC4_{FL}$) was observed to be remarkably stable in solution and I did not observe significant degradation despite the presence of extensive low-complexity regions **(Fig. 8A)**.

Full-length DCP1 ($DCP1_{FL}$) was also produced in significant quantities in a soluble form and I purified this protein using an N-terminal polyhistidine tag. $DCP1_{FL}$

has several predicted phosphorylation sites in its extensive C-terminal region. I could observe blurred rather than typically sharp bands by SDS-PAGE during purification consistent with the presence of substantial post-translational modifications **(Fig. 8A)**.

DCP2 could not be overexpressed in insect cells for reasons that were not entirely clear, possibly because of toxicity as its a very active enzyme, which mediates mRNA degradation in eukaryotes. I then switched to a bacterial expression system and using this approach I could then obtain the full-length DCP2 ($DCP2_{FL}$) as well as truncated versions ($DCP2_{\Delta C}$) for further study.

Full-length EDC3, DDX6, and PATL1 ($EDC3_{FL}$, $DDX6_{FL}$, $PATL1_{FL}$) were all produced in insect cells. $EDC3_{FL}$ and PAT_{FL} were purified in a single step using a C-terminal StrepII tag. $DDX6_{FL}$ was purified using an N-terminal polyhistidine tag by nickel affinity chromatography. Full-length MBP-tagged PNRC2 ($PNRC2_{FL}$) was produced in bacteria **(Fig. 8A)**.

The DCP1-containing multiprotein complexes were all produced in insect cells and purified using the affinity tag on DCP1. Interestingly, $DCP1_{FL}$:$PNRC2_{FL}$ and $DCP1_{FL}$:$EDC3_{FL}$:$PNRC2_{FL}$ required a high salt concentration during the purification procedure for stability. The DCP1 complexes in which PNRC2 was not included, ($DCP1_{FL}$:$EDC3_{FL}$ and $DCP1_{FL}$:$EDC3_{FL}$:$DDX6_{FL}$) additionally require the inclusion of a non-ionic detergent as well as additional sucrose gradient separation following size exclusion chromatography **(Fig. 8A)**.

Taken together, the production and purification of these recombinant human decapping factors then provided the necessary reagents to conduct unbiased biochemical activity screen.

4.3.2 Full-length human DCP2 is highly active *in vitro*

In the first series of experiments, I characterized the basal activity of the human DCP2 in its full-length form as well as a truncated version comprising the folded domains only. Strikingly, and in contrast to expectations, I observed that $DCP2_{FL}$ is a highly active enzyme *in vitro* and decaps around 95% of capped mRNA substrate over a time course of 60 min **(Fig 8C)**.

It was shown the decapping activity of human DCP2 resides in the NRD and Nudix domains along with the Box-B helix characterized to bind to RNA. To test this, I

made a construct comprising only these folded domains ($DCP2_{\Delta C}$) and assayed its decapping activity under the same conditions as $DCP2_{FL}$. I observed that $DCP2_{\Delta C}$ is very weakly active reaching approximately 25% of decapped substrate in a time course of 60 min **(Fig. 8C)**. This suggested that the C-terminal low-complexity region of DCP2 contributes to the stimulation of the DCP2 enzymatic activity.

Sequence analysis of the C-terminal region indicated that it is highly enriched in positively charged residues. I hypothesized that the enhanced activity of $DCP2_{FL}$ may be attributed to increased RNA binding mediated by these positively charged residues. To test this, I employed an electrophoretic mobility shift assay (EMSA) and compared $DCP2_{FL}$ vs $DCP2_{\Delta C}$. $DCP2_{FL}$ formed a very stable complex with RNA under the EMSA conditions whereas $DCP2_{\Delta C}$ did not **(Fig. 8B)**.

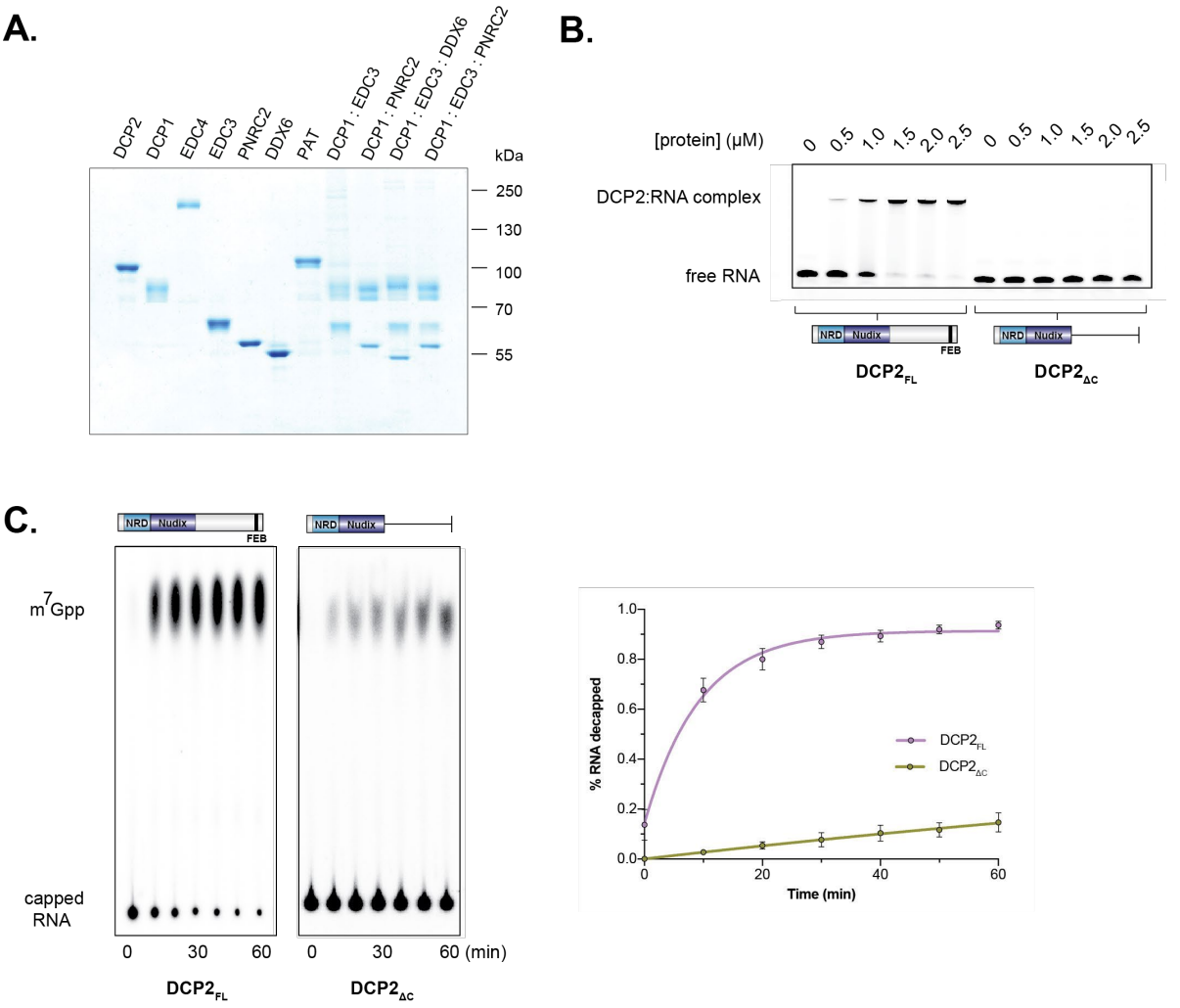


Figure 8. **(A)** SDS-PAGE gel of all recombinant, purified full-length human decapping factors. **(B)** EMSA of $DCP2_{FL}$ and $DCP2_{\Delta C}$ under the same conditions showing that the C-terminal unstructured region is involved in RNA binding. **(C)** Decapping assay of $DCP2_{FL}$ and $DCP2_{\Delta C}$ over a time course of 60 min. The % of RNA decapped was quantified and plotted against time in both cases. The assays were done as triplicates and the error bars represent standard deviation.

4.3.3 Human DCP1 slows down the activity of DCP2

When recombinant $DCP1_{FL}$ was assayed together with $DCP2_{FL}$ in an equimolar ratio, I observed that decapping was initially reduced although it reached the same level of activity as $DCP2_{FL}$ alone at the 60 min end of the time course **(Fig. 9A)**.

Next, I asked if this effect of DCP1 on DCP2 activity could be mapped to a specific region $DCP1_{FL}$. Deletion of the EVH1 domain or even the entire C-terminal region of DCP1 produced no visible effect. The C-terminal trimerization domain (TD) of DCP1 was previously shown to be important for localization in cells (Tritschler et al. 2009b). Introduction of specific mutations in the TD in the context of $DCP1_{FL}$ to reduce its propensity to self-associate also produced no discernable impact compared to $DCP1_{FL}$. I then dephosphorylated the $DCP1_{FL}$ using λ-phosphatase to test the effect of post-translational modifications but the dephosphorylated $DCP1_{FL}$ still reduced the activity of DCP2.

Based on this analysis, I conclude that DCP1 does not directly activate DCP2 in the human system. Instead, it appears to reduce the activity of DCP2 and affect the kinetics of decapping.

4.3.4 Effect of various enhancers on DCP2 activity

In previous studies, it was suggested that human EDC3 may provide a binding platform for DCP1, DCP2, and EDC4 and, as a result, of self-association mediated by the C-terminal domain bring these factors together in spatial proximity to enhance decapping (Ling et al. 2008). EDC3 also interacts with DDX6, an indirect enhancer of decapping, which acts by repressing translation (Tritschler et al. 2009a). I asked if this scenario of activation may be recreated with purified full-length factors. Interestingly, full-length EDC3 ($EDC3_{FL}$) is able to relieve the repressive effect exerted by $DCP1_{FL}$ on the activity of $DCP2_{FL}$ when I assayed the $EDC3_{FL}$:$DCP1_{FL}$ complex, but it does not enhance the activity of $DCP2_{FL}$ beyond the basal level **(Fig. 9B)**.

The effect of $EDC4_{FL}$ on $DCP2_{FL}$ activity has never been tested *in vitro* before with recombinant proteins. I asked if $DCP1_{FL}$ can enhance the $DCP2_{FL}$ activity in the presence of $EDC4_{FL}$ as has been reported using immunopurified proteins from cells. Surprisingly, however, $EDC4_{FL}$ dramatically reduced $DCP2_{FL}$ activity in the presence of $DCP1_{FL}$ by almost 30% indicating that there is no direct stimulation **(Fig. 9C)**.

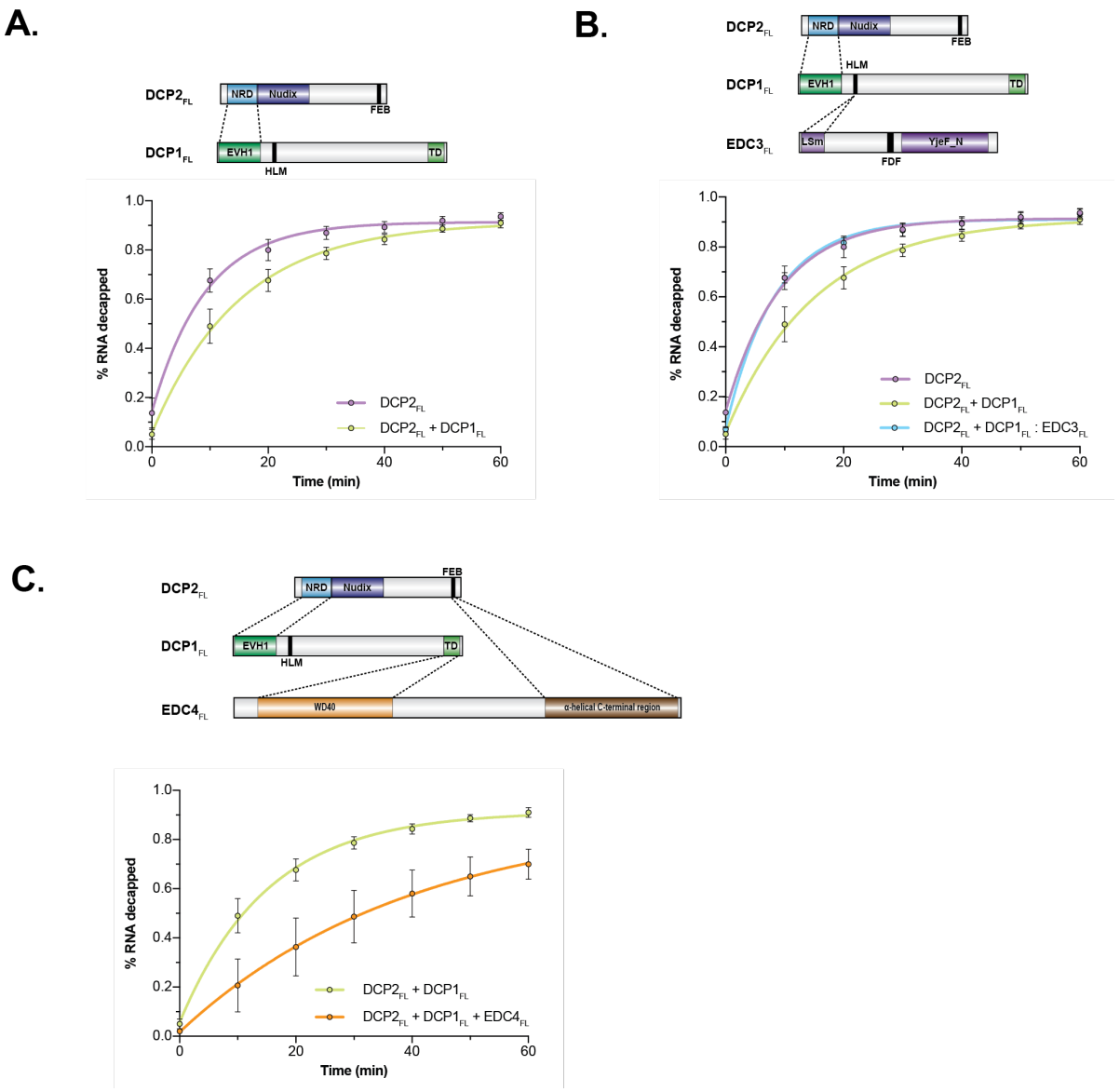


Figure 9. **(A)** Decapping assay of DCP2$_{FL}$ vs DCP2$_{FL}$ + DCP1$_{FL}$ over a time course of 60 min showing the repressive effect of DCP1. **(B)** Decapping assay of DCP2$_{FL}$ vs DCP2$_{FL}$ + DCP1$_{FL}$ and DCP2$_{FL}$ + DCP1$_{FL}$:EDC3$_{FL}$ over a time course of 60 min showing the release of the repressive effect of DCP1. **(C)** Decapping assay of DCP2$_{FL}$ + DCP1$_{FL}$ vs DCP2$_{FL}$ + DCP1$_{FL}$ + EDC4$_{FL}$ over a time course of 60 min showing inhibition of decapping activity. The % of RNA decapped was quantified and plotted against time in all cases. The assays were done as triplicates and the error bars represent standard deviation.

4.3.5 PNRC2 activates the decapping complex

PNRC2 is a vertebrate-specific intrinsically disordered protein, which interacts with the EVH1 domain of DCP1 via a proline-rich sequence (PRS) (Lai et al. 2012). I asked if PNRC2, in complex with DCP1, can enhance the activity of DCP2. To test this, in the context of full-length proteins, I assayed $PNRC2_{FL}$ in complex with $DCP1_{FL}$. Strikingly, $PNRC2_{FL}$ in complex with $DCP1_{FL}$ enhances $DCP2_{FL}$ activity as observed at 90% of decapped substrate after just 10 min in a time course **(Fig. 10A)**. I then tested if mutations in the PRS affect this stimulation. Based on the crystal structure of the PNRC2 PRS with the EVH1 domain of DCP1, I mutated the W114A which has been tested before and was shown to reduce activity and abrogate binding to DCP1, respectively (Lai et al. 2012). $PNRC2_{FL}$ W114A in complex with $DCP1_{FL}$ was not able to activate $DCP2_{FL}$ **(Fig. 10B)**. However, when $PNRC2_{FL}$ was assayed with $DCP2_{FL}$ alone it was observed to reduce its activity. PNRC2 can bind RNA directly (data not shown) and it is quite likely that it sequesters the RNA substrate in the assay. These results indicate that PNRC2 requires assembly together with DCP1 in order to act as an activator.

It was shown previously in the yeast Edc1 that the PRS alone is not sufficient for stimulation but requires the region N-terminal to the PRS, which are collectively termed the decapping activation motif (DAM). To test if this is also the case in the human PNRC2, I assayed synthetic peptides corresponding to the human PNRC2 DAM (residues 87–118) and just the PRS (residues 99–113) with $DCP1_{FL}$. I observed that the DAM of PNRC2, but not the PRS alone, is necessary and sufficient for decapping stimulation **(Fig. 10C vs 10D)**.

Taken together, PNRC2 is the only activator which I observed to enhance the activity of $DCP2_{FL}$ *in vitro*. $EDC3_{FL}$ in complex with $DCP1_{FL}$ and MBP-tagged $PNRC2_{FL}$ does not provide additional enhancement **(Fig. 10E)**.

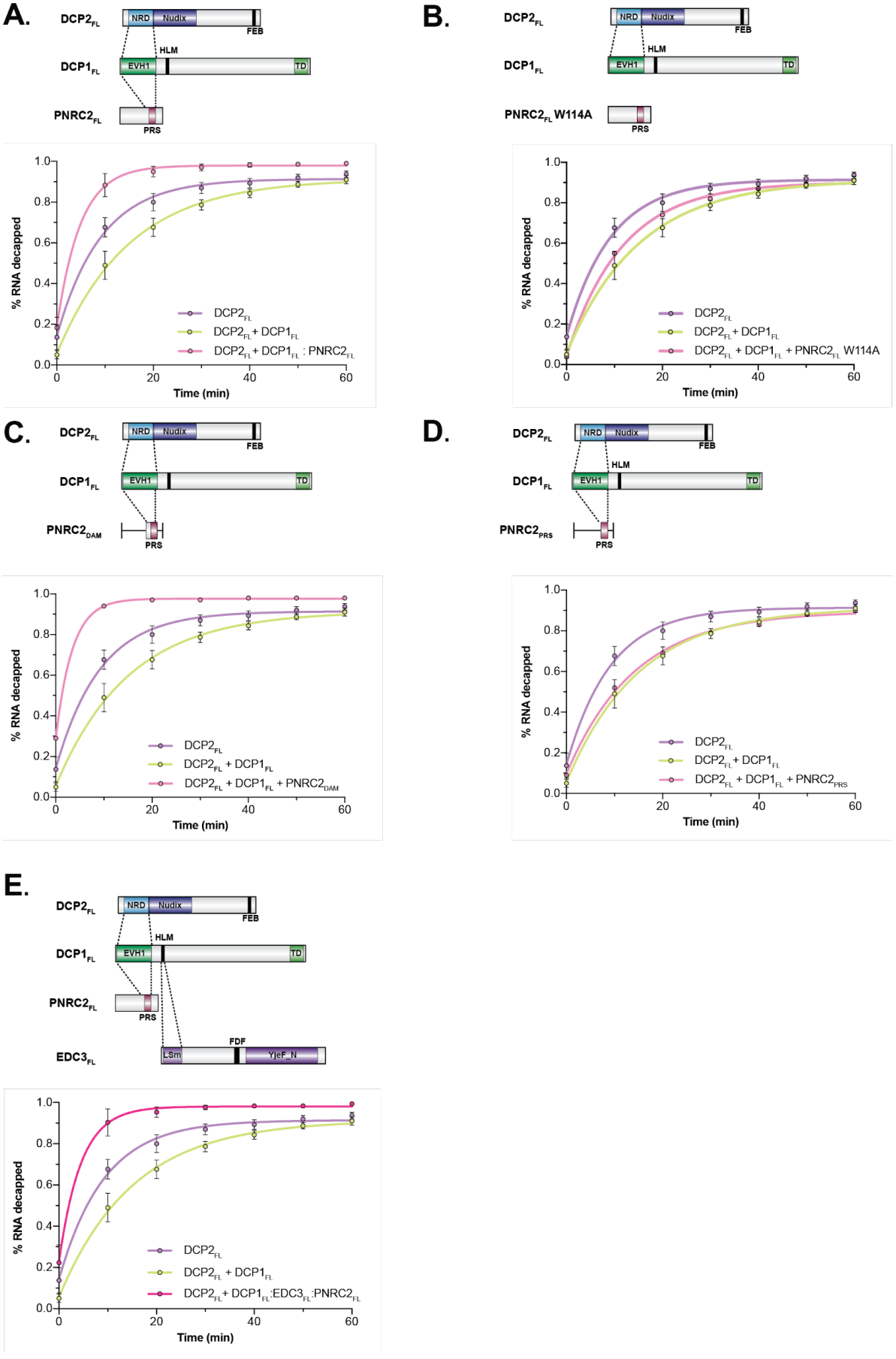

A.
DCP2FL
NRD
Nudix
FEB
HLM
DCP1FL
EVH1
TD
PNRC2FL
PRS
% RNA decapped
Time (min)
DCP2FL
DCP2FL + DCP1FL
DCP2FL + DCP1FL : PNRC2FL
B.
DCP2FL
DCP1FL
PNRC2FL W114A
PRS
DCP2FL + DCP1FL + PNRC2FL W114A
C.
PNRC2DAM
DCP2FL + DCP1FL + PNRC2DAM
D.
PNRC2PRS
DCP2FL + DCP1FL + PNRC2PRS
E.
EDC3FL
FDF
LSm
YjeF_N
DCP2FL + DCP1FL:EDC3FL:PNRC2FL

Figure 10. **(A)** Decapping assay of $DCP2_{FL}$ vs $DCP2_{FL}$ + $DCP1_{FL}$ and $DCP2_{FL}$ + $DCP1_{FL}$:$PNRC2_{FL}$over a time course of 60 min showing the activation of $DCP2_{FL}$. **(B)** Decapping assay of $DCP2_{FL}$ vs $DCP2_{FL}$ + $DCP1_{FL}$ and $DCP2_{FL}$ + $DCP1_{FL}$ + $PNRC2_{FL}$W114A which cannot bind to DCP1 over a time course of 60 min showing that PNRC2 needs to be bound to DCP1 for activation of DCP2. **(C)** Decapping assay of $DCP2_{FL}$ vs $DCP2_{FL}$ + $DCP1_{FL}$ and $DCP2_{FL}$ + $DCP1_{FL}$ + $PNRC2_{DAM}$ over a time course of 60 min which shows that the DAM is sufficient to activate DCP2. **(D)** Decapping assay of $DCP2_{FL}$ vs $DCP2_{FL}$ + $DCP1_{FL}$ and $DCP2_{FL}$ + $DCP1_{FL}$ + $PNRC2_{PRS}$ over a time course of 60 min where the PRS alone cannot activate DCP2. **(E)** Decapping assay of $DCP2_{FL}$ vs $DCP2_{FL}$ + $DCP1_{FL}$ and $DCP2_{FL}$ + $DCP1_{FL}$:$EDC3_{FL}$: $PNRC2_{FL}$ over a time course of 60 min where the presence of EDC3 does not enhance DCP2 activity beyond PNRC2 activation. The % of RNA decapped was quantified and plotted against time in all cases. The assays were done as triplicates and the error bars represent standard deviation.

5. DISCUSSION

5.1 Rewiring of the decapping interactome in evolution

The biochemical and structural characterization of the decapping 'interactome' has provided important mechanistic insights into the regulation of the decapping network and how key regulatory interactions have diverged in evolution. In yeast, Dcp2 is at the nexus of the regulatory network and its activity is coordinated and regulated by direct, high-affinity interactions with its enhancers. This is necessary because Dcp2 has poor basal activity and requires its obligate activator Dcp1 along with other enhancers to reach maximal activity. The C-terminal unstructured region of yeast Dcp2 has also been shown to have an autoinhibitory function, which is relieved by enhancers such as Edc3 (He et al. 2018; Zeidan et al. 2018; Paquette et al. 2018). In contrast, human DCP2 is a highly active enzyme that does not require significant stimulation *in vitro*. The C-terminal unstructured region of DCP2 has an important role in RNA binding thus directly influencing the enzymatic properties of DCP2 by stabilizing an enzyme-substrate complex. When this unstructured region is not present, the structured NRD and Nudix domains of DCP2 cannot interact stably with the substrate, which leads to low overall enzymatic activity despite the presence of elements such as the Box-B helix in the Nudix domain, which have an important role in substrate binding in the yeast Dcp2 (She et al. 2008; Aglietti et al. 2013; Deshmukh et al. 2008; Valkov et al. 2016). Taken together. these results point to a significant evolutionary divergence in the mechanistic aspects of autoregulation and activation of the decapping enzyme.

In yeast, Dcp1 is an obligate coactivator of Dcp2 and enhances the activity of Dcp2 by an order of magnitude (Borja et al. 2011). I showed, however, that human DCP1 does not activate DCP2 directly *in vitro* but rather slows down its activity. Whilst a compelling mechanistic explanation for this effect is still to be determined, one possible explanation is the presence of an extended unstructured C-terminal region in human DCP1, in addition to the EVH1 domain. This unstructured region has an embedded helical motif, which serves to mediate protein-protein interactions in similarity to what was observed in the C-terminal unstructured region of the yeast Dcp2. Since in the yeast Dcp2 this motif serves an autoinhibitory function, it is tempting

to speculate that during the course of evolution this motif, together with its inhibitory function, was transferred to DCP1 from Dcp2.

The DCP1 C-terminal region is also extensively phosphorylated (Aizer et al. 2013; Chiang et al. 2013; Tenekeci et al. 2016; Ma et al. 2013). I observed that the phosphorylation of this region does not appear to influence its function in so far DCP2 activation is concerned as the dephosphorylated protein had an identical activity profile together with DCP2 *in vitro* compared to the phosphorylated one. Whether phosphorylation and additional posttranslational modifications of DCP1 have other roles such as in regulating the localization of DCP1 or connecting various signaling processes to decapping still remains to be elucidated.

Another important distinction between yeast and human DCP1 is the presence of the trimerization domain in the human ortholog. The functional relevance of DCP1 self-association remains unclear, but the removal of this trimerization domain and mutagenesis of key residues mediating trimerization results in some observable decapping defects in human cells, and monomeric DCP1 does not localize to P-bodies (Tritschler et al. 2009b). Although evidence is accumulating that suggests that P-bodies are sites of mRNA storage rather than decay, there remains a possibility that the oligomerization of DCP1 is important for its localization in the cytoplasm. Another possibility is that DCP1 serves a function in the direct recruitment of target mRNAs to DCP2 for decapping. DCP1 is the only decay-associated factor that shuttles in and out of P-bodies whereas DCP2 is the least mobile (Coller and Parker 2005; Parker and Sheth 2007). Since it is increasingly appreciated that P-bodies are not directly associated with mRNA decay it may be that translationally repressed mRNAs need to exit P-bodies prerequisite to decay and DCP1 could serve an important a role in mediating this process.

Characterized enhancers of decapping in yeast such as Edc3 bind directly to the HLM in the C-terminal region of Dcp2 (Fromm et al. 2012; Harigaya et al. 2010). However, human EDC3 interacts with the HLM embedded in the C-terminal region of human DCP1. In decapping assays, I showed that a binary DCP1:EDC3 complex can relieve the repressive effect of DCP1 but does not enhance the activity of DCP2. This suggests that mechanistic aspects of the stimulation by EDC3 are conserved in

evolution even though this activator acts directly on different factors: Dcp2 in yeast and human DCP1.

PNRC2 is an intrinsically disordered protein with a proline-rich sequence (PRS) and is very similar to the yeast Edc1. The PRS interacts directly and with high affinity with the EVH1 domain in DCP1 (She et al. 2004; Borja et al. 2011). I showed that PNRC2 is the only activator of DCP2 among all the purified factors I tested. I established that PNRC2 stimulates DCP2 activity only in the presence of DCP1, which provides the high-affinity binding surface. The requirement for DCP1 is absolute as PNRC2 cannot activate DCP2 in the context of mutations that disrupt binding to DCP1. Taken together, I propose that the mechanism of activation of DCP2 activity by PNRC2 is very likely to be highly similar to that of Edc1 in yeast as described in Valkov et. al. 2016, but structural evidence will be required to verify this.

EDC4 was proposed to serve as a scaffold for DCP1 and DCP2 assembly by facilitating their interaction (Chang et al. 2014), and a short linear motif in DCP2 C-terminal region binds directly to the structured C-terminal domain of EDC4. In assays with purified recombinant factors, I did not observe that EDC4 stimulates DCP2 activity in the presence of DCP1. In contrast, the activity of DCP2 is quite substantially inhibited in the presence of EDC4 irrespective of DCP1 presence. The mechanistic explanation for this is presently lacking. I hypothesized that EDC4 could negatively affect DCP2 activity by sequestering the RNA substrate *in vitro*. To test this, I titrated increasing amounts of cold RNA in decapping assays using DCP2 alone and DCP2 together with EDC4. PNRC2 binds directly to RNA (data not shown) and was used as a control, whilst DCP2 was previously shown to be inhibited by increasing amounts of RNA *in vitro*. I observed that EDC4 is unaffected in its inhibitory activity toward DCP2 with increasing amounts of RNA. In contrast, PNRC2 is strongly negatively affected in its ability to activate DCP2 in the presence of excess RNA. Collectively, this data suggests that EDC4 exerts its repressive activity directly on DCP2, possibly through a direct interaction involving by the activatory C-terminal region of DCP2, rather than indirectly via substrate removal.

In summary, I propose that in the human decapping network it is DCP1 rather than DCP2, which serves as a ‘hub’ for various regulatory inputs. Whether DCP1 acts as a recruitment factor for DCP2 remains unclear and requires further study.

5.2 Coordination of decay events at the 3′ and 5′ ends of mRNA

An intrinsic element called ARE (AU-rich element) at the 3′-UTR recruits factors that trigger 3′–5′ as well as 5′–3′ decay (Li and Kiledjian 2010). These AREs associate with various specific RNA binding proteins. In particular, TTP (Tristetraprolin) can specifically promote ARE-dependent DCP2 decapping *in vitro (Fenger-Grøn et al. 2005)* suggesting these factors can recruit decapping factors to the target mRNA directly to activate decapping and subsequent degradation. Another trigger for decapping is the presence of a stretch of U residues at the 3′ end, known as the 'U-tract'. A U-tract with an optimal length of 5 U's at the 3′ end of mRNA triggers DCP2-mediated decapping in mammalian cell extracts and this is directed by the LSm 1–7 complex (Song and Kiledjian 2007), which recognizes terminal oligo U stretches (Chowdhury et al. 2007). The LSm 1–7 complex interacts with PATL1, which directly interacts with decapping factors thus triggering decapping (Sharif and Conti 2013; Chowdhury et al. 2007). Endogenous mRNAs that lack a poly(A) tail, such as histone mRNAs, are also known to be oligouridylated at the 3′ end and are directed toward decapping-dependent decay (Song and Kiledjian 2007; Mullen and Marzluff 2008). An alternative is for the decapping factors to be recruited via the CCR4-NOT complex via DDX6 helicase, which interacts with both the CCR4-NOT complex and EDC3 (Tritschler et al. 2009a; Ozgur et al. 2015). Decapping at the 5′ end of mRNAs is thus well known to be coordinated with the shortening of the poly(A) tail at the 3′ end. But the precise sequence of events and mechanism remains an open and a controversial question in the field. There is evidence to support a model in which the recruitment of DCP2 to the cap can be mediated by factors that associate with the 3′ end of transcripts. Experimental evidence presented in this thesis, which shows that DCP2 can interact with RNA directly and with high affinity, suggests that additional connections via mRNA itself can be made between the decay factors but this remains speculative.

Although the primary focus of the latter part of my doctoral work was on the reconstitution of decapping interactome, I also collaborated with colleagues in the department and contributed toward the reconstitution of the entire human CCR4-NOT deadenylase complex. We characterized the intrinsic regulation of the deadenylation and showed that linear motifs from different RNA-binding factors that recruit CCR4-

NOT to specific mRNAs can also inhibit bulk deadenylation. This work is described in a manuscript (Raisch, Chang, Levdansky et. al., 2019) now accepted for publication and is attached in the **Appendix**.

Another collaborative project, to which I made a contribution, was focused on the characterization of a C-terminal low-complexity region in human XRN1 exonuclease. We showed that XRN1 interacts directly with the human CCR4-NOT complex and represses deadenylation. This inhibition is, in turn, relieved by the general decapping activator PATL1. The data is described in the manuscript Chang et. al., currently under revision.

Collectively, the development of production and purification strategies for human decapping factors and multiprotein complexes with all their principal interacting regions fully intact in addition to my contributions towards the reconstitution of the human CCR4-NOT deadenylation complex and its connection to XRN1 is an important step towards the goal of rebuilding the entire 5′–3′ decay pathway in a test tube. This offers means to map and directly investigate the sequence of interactions that coordinates the decay events at the 3′ and 5′ ends of mRNA with exquisite precision in a strictly compositionally defined system.

An important and outstanding but underappreciated question is the regulation of mRNA decay in response to external stimuli. Almost all factors described in this thesis are post-translationally modified and have connections to various signaling pathways. Recent advances in sequencing technologies in various cellular contexts in combination with the biochemical reconstitution of large macromolecular assemblies offers an unprecedented opportunity to dissect the mechanism of mRNA decay and its integration in the eukaryotic gene expression pathway.

6. REFERENCES

Aglietti RA, Floor SN, McClendon CL, Jacobson MP, Gross JD. 2013. Active site conformational dynamics are coupled to catalysis in the mRNA decapping enzyme Dcp2. *Structure* **21**: 1571–1580.

Aizer A, Kafri P, Kalo A, Shav-Tal Y. 2013. The P body protein Dcp1a is hyper-phosphorylated during mitosis. *PLoS One* **8**: e49783.

Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P. 1999. The yeast exosome and human PM-Scl are related complexes of 3′ --> 5′ exonucleases. *Genes Dev* **13**: 2148–2158.

Arribas-Layton M, Wu D, Lykke-Andersen J, Song H. 2013. Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim Biophys Acta* **1829**: 580–589.

Basquin J, Roudko VV, Rode M, Basquin C, Séraphin B, Conti E. 2012. Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. *Mol Cell* **48**: 207–218.

Bessman MJ, Frick DN, O'Handley SF. 1996. The MutT Proteins or "Nudix" Hydrolases, a Family of Versatile, Widely Distributed, "Housecleaning" Enzymes. *J Biol Chem* **271**: 25059–25062.

Bianchin C, Mauxion F, Sentis S, Séraphin B, Corbo L. 2005. Conservation of the deadenylase activity of proteins of the Caf1 family in human. *RNA* **11**: 487–494.

Bieniossek C, Imasaki T, Takagi Y, Berger I. 2012. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* **37**: 49–57.

Borja MS, Piotukh K, Freund C, Gross JD. 2011. Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. *RNA* **17**: 278–290.

Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B. 2000. A Sm-like protein complex that participates in mRNA degradation. *EMBO J* **19**: 1661–1671.

Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, Gharakhani J, Jülicher F, Hyman AA. 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**: 1729–1732.

Brown CE, Sachs AB. 1998. Poly(A) Tail Length Control inSaccharomyces cerevisiaeOccurs by Message-Specific Deadenylation. *Molecular and Cellular Biology* **18**: 6548–6559. http://dx.doi.org/10.1128/mcb.18.11.6548.

Chang C-T, Bercovich N, Loh B, Jonas S, Izaurralde E. 2014. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res* **42**: 5217–5233.

Charenton C, Taverniti V, Gaudon-Plesse C, Back R, Séraphin B, Graille M. 2016. Structure of the active form of Dcp1-Dcp2 decapping enzyme bound to m7GDP

and its Edc3 activator. *Nat Struct Mol Biol* **23**: 982–986.

Chen C-YA, Shyu A-B. 2011. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**: 167–183.

Chiang P-Y, Shen Y-F, Su Y-L, Kao C-H, Lin N-Y, Hsu P-H, Tsai M-D, Wang S-C, Chang G-D, Lee S-C, et al. 2013. Phosphorylation of mRNA decapping protein Dcp1a by the ERK signaling pathway during early differentiation of 3T3-L1 preadipocytes. *PLoS One* **8**: e61697.

Chowdhury A, Mukhopadhyay J, Tharun S. 2007. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**: 998–1016.

Coller J, Parker R. 2005. General translational repression by activators of mRNA decapping. *Cell* **122**: 875–886.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* **136**: 777–793.

Decker CJ, Parker R. 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* **7**: 1632–1643.

Decker CJ, Teixeira D, Parker R. 2007. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. *J Cell Biol* **179**: 437–449.

Deshmukh MV, Jones BN, Quang-Dang D-U, Flinders J, Floor SN, Kim C, Jemielity J, Kalek M, Darzynkiewicz E, Gross JD. 2008. mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell* **29**: 324–336.

Dunckley T, Parker R. 1999. The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J* **18**: 5411–5422.

Dunckley T, Tucker M, Parker R. 2001. Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in Saccharomyces cerevisiae. *Genetics* **157**: 27–37.

Dupressoir A, Morel AP, Barbot W, Loireau MP, Corbo L, Heidmann T. 2001. Identification of four families of yCCR4- and Mg2+-dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAF1/hPOP2 binding. *BMC Genomics* **2**: 9.

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. 2007. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* **27**: 3970–3981.

Fenger-Grøn M, Fillman C, Norrild B, Lykke-Andersen J. 2005. Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol Cell* **20**: 905–915.

Floor SN, Borja MS, Gross JD. 2012. Interdomain dynamics and coactivation of the

mRNA decapping enzyme Dcp2 are mediated by a gatekeeper tryptophan. *Proc Natl Acad Sci U S A* **109**: 2872–2877.

Floor SN, Jones BN, Hernandez GA, Gross JD. 2010. A split active site couples cap recognition by Dcp2 to activation. *Nat Struct Mol Biol* **17**: 1096–1101.

Franks TM, Lykke-Andersen J. 2008. The control of mRNA decapping and P-body formation. *Mol Cell* **32**: 605–615.

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

Garneau NL, Wilusz J, Wilusz CJ. 2007. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**: 113–126.

Haas G, Braun JE, Igreja C, Tritschler F, Nishihara T, Izaurralde E. 2010. HPat provides a link between deadenylation and decapping in metazoa. *J Cell Biol* **189**: 289–302.

Halbach F, Reichelt P, Rode M, Conti E. 2013. The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell* **154**: 814–826.

Halbach F, Rode M, Conti E. 2012. The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* **18**: 124–134.

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, et al. 2012. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**: 768–779.

Harigaya Y, Jones BN, Muhlrad D, Gross JD, Parker R. 2010. Identification and analysis of the interaction between Edc3 and Dcp2 in Saccharomyces cerevisiae. *Mol Cell Biol* **30**: 1446–1456.

He F, Celik A, Wu C, Jacobson A. 2018. General decapping activators target different subsets of inefficiently translated mRNAs. *Elife* **7**. http://dx.doi.org/10.7554/eLife.34409.

He F, Jacobson A. 2015. Control of mRNA decapping by positive and negative regulatory elements in the Dcp2 C-terminal domain. *RNA* **21**: 1633–1647.

Hsu CL, Stevens A. 1993. Yeast cells lacking 5′-->3′ exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5′ cap structure. *Mol Cell Biol* **13**: 4826–4835.

Januszyk K, Lima CD. 2014. The eukaryotic RNA exosome. *Current Opinion in Structural Biology* **24**: 132–140. http://dx.doi.org/10.1016/j.sbi.2014.01.011.

Jinek M, Eulalio A, Lingel A, Helms S, Conti E, Izaurralde E. 2008. The C-terminal region of Ge-1 presents conserved structural features required for P-body localization. *RNA* **14**: 1991–1998.

Jonas S, Izaurralde E. 2013. The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev* **27**: 2628–2641.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, et al. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**: 753–767.

Kshirsagar M, Parker R. 2004. Identification of Edc3p as an enhancer of mRNA decapping in Saccharomyces cerevisiae. *Genetics* **166**: 729–739.

Lai T, Cho H, Liu Z, Bowler MW, Piao S, Parker R, Kim YK, Song H. 2012. Structural basis of the PNRC2-mediated link between mrna surveillance and decapping. *Structure* **20**: 2025–2037.

Linder P, Jankowsky E. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* **12**: 505–516.

Ling SHM, Decker CJ, Walsh MA, She M, Parker R, Song H. 2008. Crystal structure of human Edc3 and its functional implications. *Mol Cell Biol* **28**: 5965–5976.

Li P, Banjade S, Cheng H-C, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, et al. 2012a. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**: 336–340.

Liu H, Rodgers ND, Jiao X, Kiledjian M. 2002. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J* **21**: 4699–4708.

Li Y, Dai J, Song M, Fitzgerald-Bocarsly P, Kiledjian M. 2012b. Dcp2 decapping protein modulates mRNA stability of the critical interferon regulatory factor (IRF) IRF-7. *Mol Cell Biol* **32**: 1164–1172.

Li Y, Kiledjian M. 2010. Regulation of mRNA decapping. *Wiley Interdiscip Rev RNA* **1**: 253–265.

Lykke-Andersen J. 2002. Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol Cell Biol* **22**: 8114–8121.

Ma J, Flemr M, Strnad H, Svoboda P, Schultz RM. 2013. Maternally recruited DCP1A and DCP2 contribute to messenger RNA degradation during oocyte maturation and genome activation in mouse. *Biol Reprod* **88**: 11.

Mangus DA, Evans MC, Agrin NS, Smith M. 2004. Positive and negative regulation of poly (A) nuclease. *and cellular biology*. https://mcb.asm.org/content/24/12/5521.short.

McLennan AG. 2006. The Nudix hydrolase superfamily. *Cell Mol Life Sci* **63**: 123–143.

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

Mugridge JS, Tibble RW, Ziemniak M, Jemielity J, Gross JD. 2018. Structure of the activated Edc1-Dcp1-Dcp2-Edc3 mRNA decapping complex with substrate analog poised for catalysis. *Nat Commun* **9**: 1152.

Muhlrad D, Decker CJ, Parker R. 1994. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5′--> 3′ digestion of the transcript. *Genes Dev*. http://genesdev.cshlp.org/content/8/7/855.short.

Mullen TE, Marzluff WF. 2008. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5′ to 3′ and 3′ to 5′. *Genes Dev* **22**: 50–65.

Nissan T, Rajyaguru P, She M, Song H, Parker R. 2010. Decapping activators in Saccharomyces cerevisiae act by multiple mechanisms. *Mol Cell* **39**: 773–783.

Ozgur S, Basquin J, Kamenska A, Filipowicz W, Standart N, Conti E. 2015. Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Rep* **13**: 703–711.

Ozgur S, Chekulaeva M, Stoecklin G. 2010. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. *Mol Cell Biol* **30**: 4308–4323.

Paquette DR, Tibble RW, Daifuku TS, Gross JD. 2018. Control of mRNA decapping by autoinhibition. *Nucleic Acids Res*. http://dx.doi.org/10.1093/nar/gky233.

Parker R, Sheth U. 2007. P bodies and the control of mRNA translation and degradation. *Mol Cell* **25**: 635–646.

Pilkington GR, Parker R. 2008. Pat1 contains distinct functional domains that promote P-body assembly and activation of decapping. *Mol Cell Biol* **28**: 1298–1312.

Rajyaguru P, She M, Parker R. 2012. Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. *Mol Cell* **45**: 244–254.

Rzeczkowski K, Beuerlein K, Müller H, Dittrich-Breiholz O, Schneider H, Kettner-Buhrow D, Holtmann H, Kracht M. 2011. c-Jun N-terminal kinase phosphorylates DCP1a to control formation of P bodies. *J Cell Biol* **194**: 581–596.

Sari D, Gupta K, Raj DBTG, Aubert A, Drncová P, Garzoni F, Fitzgerald D, Berger I. 2016. The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics. In *Advanced Technologies for Protein Complex Production and Characterization* (ed. M.C. Vega), pp. 199–215, Springer International Publishing, Cham.

Schwartz D, Decker CJ, Parker R. 2003. The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. *RNA* **9**: 239–251.

Sharif H, Conti E. 2013. Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. *Cell Rep* **5**: 283–291.

Sharif H, Ozgur S, Sharma K, Basquin C, Urlaub H, Conti E. 2013. Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Res* **41**: 8377–8390.

She M, Decker CJ, Sundramurthy K, Liu Y, Chen N, Parker R, Song H. 2004. Crystal structure of Dcp1p and its functional implications in mRNA decapping. *Nat Struct Mol Biol* **11**: 249–256.

She M, Decker CJ, Svergun DI, Round A, Chen N, Muhlrad D, Parker R, Song H. 2008. Structural basis of dcp2 recognition and activation by dcp1. *Mol Cell* **29**: 337–349.

Song M-G, Kiledjian M. 2007. 3' Terminal oligo U-tract-mediated stimulation of decapping. *RNA* **13**: 2356–2365.

Song M-G, Li Y, Kiledjian M. 2010. Multiple mRNA decapping enzymes in mammalian cells. *Mol Cell* **40**: 423–432.

Tenekeci U, Poppe M, Beuerlein K, Buro C, Müller H, Weiser H, Kettner-Buhrow D, Porada K, Newel D, Xu M, et al. 2016. K63-Ubiquitylation and TRAF6 Pathways Regulate Mammalian P-Body Formation and mRNA Decapping. *Mol Cell* **62**: 943–957.

Thore S, Mauxion F, Séraphin B, Suck D. 2003. X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO reports* **4**: 1150–1155. http://dx.doi.org/10.1038/sj.embor.7400020.

Tritschler F, Braun JE, Eulalio A, Truffault V, Izaurralde E, Weichenrieder O. 2009a. Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. *Mol Cell* **33**: 661–668.

Tritschler F, Braun JE, Motz C, Igreja C, Haas G, Truffault V, Izaurralde E, Weichenrieder O. 2009b. DCP1 forms asymmetric trimers to assemble into active mRNA decapping complexes in metazoa. *Proc Natl Acad Sci U S A* **106**: 21591–21596.

Tritschler F, Eulalio A, Helms S, Schmidt S, Coles M, Weichenrieder O, Izaurralde E, Truffault V. 2008. Similar modes of interaction enable Trailer Hitch and EDC3 to associate with DCP1 and Me31B in distinct protein complexes. *Mol Cell Biol* **28**: 6695–6708.

Tritschler F, Eulalio A, Truffault V, Hartmann MD, Helms S, Schmidt S, Coles M, Izaurralde E, Weichenrieder O. 2007. A divergent Sm fold in EDC3 proteins mediates DCP1 binding and P-body targeting. *Mol Cell Biol* **27**: 8600–8611.

Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell* **104**: 377–386.

Valkov E, Muthukumar S, Chang C-T, Jonas S, Weichenrieder O, Izaurralde E. 2016. Structure of the Dcp2-Dcp1 mRNA-decapping complex in the activated

conformation. *Nat Struct Mol Biol* **23**: 574–579.

van Dijk E, Cougot N, Meyer S, Babajko S, Wahle E, Seraphin A. 2002. Human Dcp2: a catalytically active mRNA decapping enzyme locatedin speci®c cytoplasmic structures. *EMBO J* **21**.

Wahle E, Winkler GS. 2013. RNA decay machines: Deadenylation by the Ccr4–Not and Pan2–Pan3 complexes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**: 561–570.

Wang H, Morita M, Yang X, Suzuki T, Yang W, Wang J, Ito K, Wang Q, Zhao C, Bartlam M, et al. 2010. Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. *EMBO J* **29**: 2566–2576.

Wang L, Lewis MS, Johnson AW. 2005. Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. *RNA* **11**: 1291–1302.

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

Weber SC, Brangwynne CP. 2012. Getting RNA and protein in phase. *Cell* **149**: 1188–1191.

Webster MW, Chen Y-H, Stowell JAW, Alhusaini N, Sweet T, Graveley BR, Coller J, Passmore LA. 2018. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol Cell* **70**: 1089–1100.e8.

Weissmann F, Peters J-M. 2018. Expressing Multi-subunit Complexes Using biGBac. In *Protein Complex Assembly: Methods and Protocols* (ed. J.A. Marsh), pp. 329–343, Springer New York, New York, NY.

Weissmann F, Petzold G, VanderLinden R, Huis In 't Veld PJ, Brown NG, Lampert F, Westermann S, Stark H, Schulman BA, Peters J-M. 2016. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc Natl Acad Sci U S A* **113**: E2564–9.

Wild R, Hothorn M. 2017. The macro domain as fusion tag for carrier-driven crystallization. *Protein Sci* **26**: 365–374.

Wilusz CJ, Wilusz J. 2005. Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* **12**: 1031–1036.

Wurm JP, Overbeck J, Sprangers R. 2016. The S. pombe mRNA decapping complex recruits cofactors and an Edc1-like activator through a single dynamic surface. *RNA* **22**: 1360–1372.

Xu J, Yang J-Y, Niu Q-W, Chua N-H. 2006. Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. *Plant Cell* **18**: 3386–3398.

Yamashita A, Chang T-C, Yamashita Y, Zhu W, Zhong Z, Chen C-YA, Shyu A-B. 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian

mRNA turnover. *Nat Struct Mol Biol* **12**: 1054–1063.

Yi H, Park J, Ha M, Lim J, Chang H, Kim VN. 2018. PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol Cell* **70**: 1081–1088.e5.

Yu JH, Yang W-H, Gulick T, Bloch KD, Bloch DB. 2005. Ge-1 is a central component of the mammalian cytoplasmic mRNA processing body. *RNA* **11**: 1795–1802.

Zeidan Q, He F, Zhang F, Zhang H, Jacobson A, Hinnebusch AG. 2018. Conserved mRNA-granule component Scd6 targets Dhh1 to repress translation initiation and activates Dcp2-mediated mRNA decay in vivo. *PLoS Genet* **14**: e1007806.

Zheng D, Ezzeddine N, Chen C-YA, Zhu W, He X, Shyu A-B. 2008. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. *J Cell Biol* **182**: 89–101.

7. ABBREVIATIONS

4E-T	eIF4E-transporter
ARE	AU-rich element
AREs	AU (Adenylate-Uridylate)-rich elements
ARM	domain consisting of armadillo-like helical repeats
ATP	Adenosine triphosphate
CAF1	CCR4-associated factor 1
CBC	Cap binding complex
CCR4-NOT	Carbon catabolite repressor protein 4-negative on TATA
Ce	Caenorhabditis elegans
CEC	Capping enzyme complex
DAM	Decapping activating motif
Dcp1	Decapping protein subunit 1 from yeast
DCP1	Decapping protein subunit 1 from metazoans
Dcp2	Decapping protein subunit 2 from yeast
DCP2	Decapping protein subunit 2 from metazoans
DcpS	Scavenger decapping enzyme
DDX6	DEAD-box helicase 6
DEAD	Aspartic Acid-Glutamic Acid-Alanine-Aspartic Acid box
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
Edc1	Enhancer of decapping protein 1 from yeast
Edc2	Enhancer of decapping protein 2 from yeast
EDC3	Enhancer of decapping protein 3 from metazoans
Edc3	Enhancer of decapping protein 3 from yeast
EDC4	Enhancer of decapping protein 4 from metazoans
EEP	Endonuclease exonuclease phosphatase domain
eIF4A	eukaryotic initiation factor 4A
eIF4E	eukaryotic initiation factor 4E
eIF4F	eukaryotic initiation factor 4F
eIF4G	eukaryotic initiation factor 4G
EJCs	Exon junction complexes
EMSA	Electrophoretic mobility shift assay

EVH1	Ena/Vasp homology 1
FDF	Phenylalanine-Aspartate-Phenylalanine
FEB	Phenylalanine rich EDC4 binding motif
HEAT	helical hairpin repeats originally found in Huntingtin, Elongation factor 3, protein phosphatase 2A and TOR1
HLM	Helical leucine rich motif
LLR	Leucine-rich repeat
LSm	Like Sm
m7G	7-methylguanosine
m7GDP	7-methylguanosine diphosphate
Mex67	mRNA export factor 67
miRNAs	microRNAs
mRNA	messenger ribonucleic acid
mRNPs	messenger ribonuclear particles
NGD	No-go decay
NMD	Nonsense mediated mRNA decay
NMR	Nuclear magnetic resonance
NPCs	Nuclear pore complexes
NRD	N-terminal regulatory domain
NSD	Non-stop decay
Nudt16	Nudix hydrolase 16
NXF1	Nuclear RNA export factor 1
PABPC1	Poly(A)-binding protein C1
PAN2-PAN3	Poly(A) nuclease 2-3
PAPBNs	Nuclear poly(A)-binding proteins
PATL1	Topoisomerase 2 associated protein homolog 1
PB	P-body
PH domain	Pleckstrin homology domain
PNRC2	Proline-rich nuclear receptor coactivator 2
poly(A)	Poly adenosine stretch at the 3′ end of mRNAs
pre-mRNA	precursor-messenger RNA
PRSs	Proline rich sequences
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II

SDS-PAGE	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
siRNAs	small interfering RNAs
SLiMs	Short linear motifs
snRNAs	small nuclear RNAs
Sp	Schizosaccharomyces pombe
SR proteins	Serine/arginine rich protein
TREX	Transcription/export complex
TTP	Tristetraprolin
Upf	Up-frameshift suppressor
WD40	structural motif of 40 amino acids terminating in a Tryptophan Aspartic acid dipeptide
XRN1	Exoribonuclease 1

8. APPENDIX

8.1 List of Publications

8.1.1 Manuscripts published/accepted for publication

1) Structure of the Dcp2-Dcp1 mRNA-decapping complex in the activated conformation.

Valkov, E.*, **Muthukumar, S.***, Chang, C-T.*, Jonas, S., Weichenreider, O., & Izaurralde, E. (2016). *Nat. Struct. Mol. Biol.* **23**: 574–579. *denotes equal contribution

Contribution: I optimized and purified all the proteins, carried out *in vitro* pulldowns and ITC experiments, set up crystallization screens, optimized crystals and with Eugene Valkov determined and analyzed the structure. I contributed towards writing and preparation of figures.

2) Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation

Raisch, T.*, Chang, C-T.*, Levdansky, Y.*, **Muthukumar, S.,** Raunser, S., & Valkov, E. (2019). *Nat. Commun.* *denotes equal contribution

Contribution: I designed and carried out all the electrophoretic mobility shift assays and UV-crosslinking experiments. I also contributed to the intial stages with the insect cell culture and optimization of protein production. I contributed towards writing and preparation of figures.

8.1.2 Manuscripts under revision

3) A low-complexity region in human XRN1 directly recruits decapping and deadenylation factors in 5′–3′ messenger RNA decay

Chang, C-T., **Muthukumar, S.,** Weber, R., Levdansky, Y., Chen, Y., Bhandari, D., Igreja, C., Wohlbold, L., Valkov, E & Izaurralde, E. (2019). *Nucleic Acids Res.*

Contribution: I purified and provided proteins used for all the biochemical assays and carried out the electrophoretic mobility shift assays. I contributed towards writing and preparation of figures.

8.1.3 Manuscripts in preparation

4) Assembly and regulation of the human decapping machinery: insights from biochemical reconstitution

Muthukumar, S. & Valkov, E.

Contribution: I designed and carried out all the experiments outlined in **Section 4.3**, analyzed all the data, wrote the first draft of the manuscript and made all the figures.

8.2 Methods for Chapter 4.2:

Characterization of decapping factors in *Caenorhabditis elegans*

Crystallization and structure determination of *Ce* EDC4 peptide

The peptide of the coiled-coil motif of *Ce* EDC4 used for crystallization (residues 520–552) was synthesized by Biomers. It was dissolved in 50 mM MES pH 8.0, 400mM NaCl and 2 mM DTT to a final concentration of 10 mg/ml. Initial crystals of the peptide appeared after 24 hr in sitting drop vapor diffusion plates at 22°C using a drop ratio of 200 nl of protein complex with 200 nl of reservoir solution. Crystallization hits were obtained in several conditions with inorganic salts as the main precipitant and crystals in 3.0 M ammonium sulfate and 10% v/v glycerol diffracted X-rays to 1.4 Å resolution. No further optimization was necessary and the crystals were directly flash frozen in liquid nitrogen. All diffraction data were recorded from a single crystal at the wavelength of 0.9999 Å on a PILATUS 6M detector (Dectris) at the PXII beamline of the Swiss Light Source (Villigen, Switzerland). Reflections were indexed in a cubic setting and integrated using *XDS*. Phases were readily obtained by the automated molecular replacement *phenix.automr* as implemented in the *PHENIX* suite using a 33 poly-alanine helix as a search model. 27 residues were visible in the electron density (523–549) and could be refined in space group $P4_132$. This was followed by manual residue fitting in *COOT* with several rounds of reciprocal-space refinement carried out using *phenix.refine*.

Pulldown assays

For the pulldown assays, GST tagged-macro *Ce* DCP2 peptide (741–786) and all *Ce* EDC4 constructs annotated were expressed in *Escherichia coli* BL21(DE3) Star cells (Life Technologies) in LB medium and induced with IPTG overnight at 20°C. Protein expression levels were checked on a gel and appropriate amounts of lysates were mixed and incubated with 50 µl (50% slurry) of protino glutathione agarose resin (Macherey Nagel). The samples were gently rotated for 1 hr at 4°C. The beads were washed three times with standard phosphate buffer saline (PBS). The proteins were eluted from the beads with PBS containing 25 mM L-glutathione. They were then resuspended in denaturing SDS sample buffer and analyzed by SDS-PAGE.

nature structural & molecular biology

Structure of the Dcp2–Dcp1 mRNA-decapping complex in the activated conformation

Eugene Valkov[1,3], Sowndarya Muthukumar[1,3], Chung-Te Chang[1,3], Stefanie Jonas[1,2], Oliver Weichenrieder[1] & Elisa Izaurralde[1]

The removal of the mRNA 5′ cap (decapping) by Dcp2 shuts down translation and commits mRNA to full degradation. Dcp2 activity is enhanced by activator proteins such as Dcp1 and Edc1. However, owing to conformational flexibility, the active conformation of Dcp2 and the mechanism of decapping activation have remained unknown. Here, we report a 1.6-Å-resolution crystal structure of the *Schizosaccharomyces pombe* Dcp2–Dcp1 heterodimer in an unprecedented conformation that is tied together by an intrinsically disordered peptide from Edc1. In this ternary complex, an unforeseen rotation of the Dcp2 catalytic domain allows residues from both Dcp2 and Dcp1 to cooperate in RNA binding, thus explaining decapping activation by increased substrate affinity. The architecture of the Dcp2–Dcp1–Edc1 complex provides a rationale for the conservation of a sequence motif in Edc1 that is also present in unrelated decapping activators, thus indicating that the presently described mechanism of decapping activation is evolutionarily conserved.

The removal of the protective cap structure from an mRNA 5′ end in the process of decapping is a critical step in the regulation of gene expression, because it is generally an irreversible process that inhibits translation initiation and commits the mRNA to full degradation[1,2]. Decapping is catalyzed by Dcp2, a bilobed hydrolase of the Nudix family[3–5], which hydrolyzes the mRNA 5′-cap structure, thereby releasing m^7GDP and a 5′-monophosphorylated mRNA. Decapped mRNAs are in turn susceptible to full degradation by 5′-to-3′ exonucleases of the XRN1 family[1,2]. Hence, decapping is a crucial step in bulk mRNA decay as well as in multiple specific pathways, including nonsense-mediated mRNA decay and microRNA-mediated gene silencing[1,2].

Although Dcp2 is catalytically active *in vitro*[3–13], its basal activity is enhanced by activator proteins termed decapping activators or enhancers of decapping (Edcs)[1,2]. An obligatory decapping activator is Dcp1, which stimulates Dcp2 activity by one order of magnitude[10,11] and together with Dcp2 forms the conserved core of the decapping complex[1,2,6–14]. Additional activators include Edc1–4, LSm14A, Pat1, the LSm1–7 complex and the RNA helicase Dhh1 (also known as DDX6 in vertebrates), which activate decapping in either a generic or a pathway-specific manner and play central roles in a variety of cellular processes[1,2].

Dcp2 consists of a C-terminal catalytic Nudix domain connected by a flexible hinge to an N-terminal regulatory domain (NRD), which binds the Dcp1 EVH1 domain[3–7,14] (**Supplementary Fig. 1a**). In solution, the Dcp2 Nudix and NRD domains sample multiple conformations relative to each other[6–11], and two alternative conformations have been visualized by X-ray crystallography in the presence of the Dcp1 EVH1 domain[6]: an open conformation, in which the two Dcp2 domains are splayed apart (**Supplementary Fig. 1b**), and a more compact, closed conformation, in which the Nudix domain is loaded with a nucleotide that faces the NRD domain[6] (**Supplementary Fig. 1b**). Thus, it has been proposed that catalysis occurs at a split active site formed between the two domains and that conformational changes between the open and closed conformation are responsible for the regulation of catalysis[6,8–11]. However, even in the closed conformation, there are no direct contacts between the Dcp1 EVH1 and the Dcp2 Nudix domains that would stabilize the closed Dcp2 conformation or explain the catalytic enhancement by Dcp1. Thus, an alternative, catalytically activated conformation of the enzyme must exist[9–11].

In *Saccharomyces cerevisiae* (*Sc*), the activity of the Dcp2–Dcp1 complex is further enhanced by the intrinsically disordered coactivators Edc1 and Edc2 (refs. 15–17) (**Supplementary Fig. 1a**). Edc1 has been proposed to activate decapping by promoting conformational changes in the Dcp2–Dcp1 complex[17], but the underlying molecular mechanism has remained obscure.

To shed light on the catalytically active conformation of Dcp2 and the molecular mechanism underlying decapping activation by Dcp1 and decapping enhancer Edc1, we used an intrinsically disordered peptide from Edc1 to trap and crystallize the *S. pombe* (*Sp*) Dcp2–Dcp1–Edc1 ternary complex in an activated conformation. In conjunction with binary Dcp2–Dcp1 and Dcp1–Edc1 subcomplexes, this structure reveals an unforeseen rotation of the catalytic domain of Dcp2 that brings its RNA binding residues from a peripheral position into a composite and positively charged cleft that forms between Dcp2 and Dcp1. This cleft probably accommodates the mRNA substrate, thus indicating that activation results mainly from increased substrate affinity.

[1]Department of Biochemistry, Max Planck Institute for Developmental Biology, Tübingen, Germany. [2]Present address: Institute of Biochemistry, ETH Zurich, Zurich, Switzerland. [3]These authors contributed equally to this work. Correspondence should be addressed to S.J. (stefanie.jonas@bc.biol.ethz.ch), O.W. (oliver.weichenrieder@tuebingen.mpg.de) or E.I. (elisa.izaurralde@tuebingen.mpg.de).

Received 25 March; accepted 20 April; published online 16 May 2016; doi:10.1038/nsmb.3232

Our structural analysis also uncovers the sequence requirements for Edc1 that bring about the domain rotation in Dcp2. Sequence alignments show that the Edc1 decapping-activator motif (DAm) is also present in metazoans in the PNRC protein family[18] (**Supplementary Fig. 1a**,**c**), thus indicating that the mechanism of decapping activation described in this study is evolutionarily conserved.

RESULTS

Conservation of a decapping-activator motif

Decapping activation by *Sc* Edc1 requires the 25 most C-terminal residues of *Sc* Edc1 (refs. 15–17), which we therefore termed the DAm. Sequence alignments revealed that the DAm is present in the uncharacterized *Sp* protein C18G6.09c, which is likely to represent the *Sp* Edc1 ortholog, as well as in the metazoan PNRC family of activators[18]. The DAm features a conserved $YAGX_2F$ motif followed by a C-terminal proline-rich sequence (PRS; **Supplementary Fig. 1a**,**c**). The PRS sequences of Edc1 and human PNRC2 mediate binding to the EVH1 domains of the corresponding Dcp1 proteins[17,18]. We therefore included the *Sp* Edc1 DAm peptide in combinatorial crystallization experiments with Dcp2 and Dcp1.

Structures of decapping complexes

To stabilize and visualize the activated conformation of Dcp2, we took a dual approach. First, we aimed at destabilizing the known closed conformation of the *Sp* Dcp2–Dcp1 decapping complex[6] by deleting a Dcp2 C-terminal α-helix known as the helical leucine-rich motif-1 (HLm-1; **Supplementary Fig. 1a**,**b** and **Supplementary Table 1**). This helix does not contribute to catalysis[10], and its position in the crystals is incompatible with binding and decapping enhancement by Edc3 (ref. 19). Second, we sought to trap the activated conformation of the complex and included the DAm peptide of *Sp* Edc1 (residues S155–S180; **Supplementary Fig. 1c**).

We obtained crystal structures of three distinct complexes (**Fig. 1a**–**c** and **Table 1**). The first structure shows the *Sp* Edc1 DAm peptide bound to the Dcp1 EVH1 domain (**Fig. 1a**) and is highly similar to the structure of the human DCP1 EVH1 domain bound to the PNRC2 PRS[18] (r.m.s. deviation of 1.57 Å over 124 equivalent residues; **Supplementary Fig. 2a**). In this structure, only the PRS portion of the DAm peptide is well ordered, a result consistent with the observation that the PRS is sufficient for Edc1 binding to the Dcp1 EVH1 domain[17]. We also obtained a new crystal form of the Dcp2–Dcp1 heterodimer in the

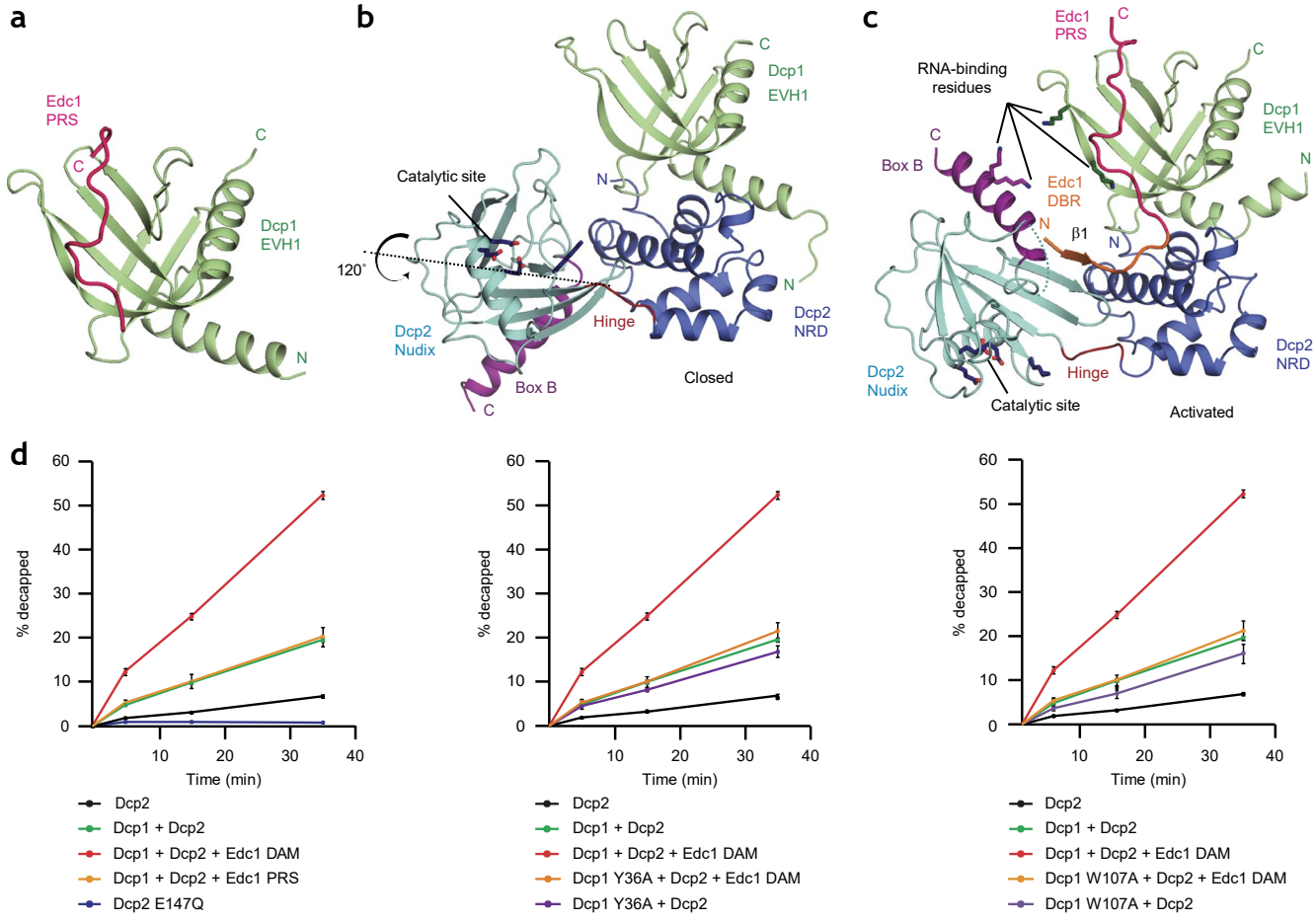


Figure 1 Structural organization and activation of decapping complexes. (**a**) Structure of the Dcp1 EVH1 domain (green) bound to the Edc1 DAM peptide comprising a C-terminal proline-rich sequence (PRS, magenta) and N-terminal Dcp2-binding residues (DBR, not visible in the structure). (**b**) Structure of the Dcp2–Dcp1 binary (closed) complex in the absence of Edc1. The Dcp2 NRD (blue) and Nudix (cyan) domains are linked by a flexible hinge (dark red). The Nudix Box-B helix is purple. (**c**) Structure of the Dcp2–Dcp1–Edc1 ternary (activated) complex. The DBR and PRS of Edc1 DAM are orange and magenta, respectively. The approximate axis of Nudix-domain rotation is indicated in **b**. Selected catalytic and RNA-binding residues are shown as sticks to delineate the catalytic site and the proposed RNA-binding groove, respectively. (**d**) Decapping assays *in vitro*. The fraction of decapped mRNA substrate, measured by the release of m^7GDP (**Supplementary Fig. 4a**), is plotted as a function of time. Error bars, s.d. from three independent experiments. The catalytically inactive Dcp2 E147Q mutant served as a negative control.

Table 1 Data collection and refinement statistics

	Sp Dcp2–Dcp1 (PDB 5J3Y)	*Sp* Dcp1–Edc1 (PDB 5J3Q)	*Sp* Dcp2–Dcp1–Edc1 (PDB 5J3T)
Data collection			
Space group	$P6_522$	$P2_12_12_1$	$C2$
Cell dimensions			
a, *b*, *c* (Å)	124.4, 124.4 367.0	41.6, 70.1, 96.7	98.8, 41.3, 93.7
α, β, γ (°)	90, 90, 120	90, 90, 90	90, 114.0, 90
Resolution (Å)	47.4–3.29 (3.47–3.29)[a]	19.5–1.87 (1.91–1.87)	45.1–1.60 (1.68–1.60)
R_{sym}	0.136 (0.859)	0.080 (0.981)	0.028 (0.653)
$I/\sigma(I)$	10.3 (2.3)	11.9 (1.7)	18.9 (1.6)
$CC_{1/2}$	99.5 (79.8)	99.9 (82.6)	100 (64.2)
Completeness (%)	99.6 (97.5)	99.8 (100)	99.1 (96.4)
Redundancy	9.6 (9.5)	6.5 (6.5)	3.3 (3.2)
Refinement			
Resolution (Å)	47.4–3.29	19.5–1.87	23.4–1.60
No. reflections	26,390	24,032	45,762
R_{work} / R_{free}	0.224 / 0.252	0.207 / 0.242	0.186 / 0.219
No. atoms			
Protein	6,021	2,310	3,239
Water	–	142	215
Ligand/ion	–	–	16[b]
B factors			
Protein	130	38.2	40.3
Water	–	44.5	48.5
Ligand/ion	–	–	36.0
R.m.s. deviations			
Bond lengths (A)	0.002	0.010	0.010
Bond angles (°)	0.399	1.050	1.050

Each data set was collected from a single crystal. [a]Values in parentheses are for highest-resolution shell. [b]Five molecules of formic acid and one magnesium ion.

absence of the Edc1 DAm peptide (**Fig. 1b** and **Table 1**), with two similar complexes per asymmetric unit. These crystals grew in a number of different conditions, in both the presence and absence of GmP or cap analog, and were morphologically identical. However, despite the lack of the HLm helix, the conformation of the Dcp2–Dcp1 complex was almost identical to the closed conformation of the previously published ATP-bound complex[6] (**Supplementary Fig. 2b**). We therefore conclude that the closed conformation represents a stabilized state of the Dcp2–Dcp1 complex, in agreement with observations in solution[6,8–11].

Rotation of the catalytic Nudix domain of Dcp2

Our third structure reveals the Dcp2–Dcp1 complex in the presence of the Edc1 DAm peptide (**Fig. 1c** and **Supplementary Fig. 1b**), which bound with an affinity of 660 nm (**Supplementary Fig. 3a**,**b**). The PRS portion of the DAm peptide binds the Dcp1 EVH1 domain in an almost identical manner as observed in the isolated Dcp1–Edc1 complex (r.m.s. deviation of 0.72 Å over 129 equivalent residues, **Supplementary Fig. 2c**). Importantly, in the ternary complex, the entire DAm peptide is visible (**Fig. 1c** and **Supplementary Fig. 1c**), including the N-terminal Dcp2-binding residues (DBR) that compose the $YAGX_2F$ motif and are required for the stimulation of decapping activity[16,17] (**Fig. 1d** and **Supplementary Fig. 4a–d**). These residues form a β-strand that augments the β-sheet of the Dcp2 Nudix domain, which is followed by a loop that contacts the Dcp2 NRD domain (**Fig. 1c**). most strikingly, these interactions stabilize the Nudix domain in an unprecedented orientation with respect to the NRD–Dcp1 module. As compared with the structure in the absence of the Edc1 peptide, in the presence of Edc1, an approximate 120° rotation of the Nudix domain moves the active site away from the NRD domain and brings originally peripheral RNA-binding residues of the so-called Box B helix and additional Nudix residues into a newly formed cleft that emerges between the repositioned Nudix domain and the NRD–Dcp1 module (comparison of **Fig. 1b** and **Fig. 1c**, and **Fig. 2b** and **Fig. 2a**). This cleft is lined by basic residues on either side and by the Edc1 β-strand at its base (**Fig. 2b**,**c**). The hydrophilic and positively charged surface of the cleft is highly suggestive of an RNA-binding channel, which may even be closed from the top by a flexible and partially disordered loop of the Nudix domain (F207–K218; **Fig. 2c** and Dcp2 alignment in **Supplementary Note**), which comprises five additional lysine residues. As a result, all previously identified RNA-binding residues (**Supplementary Table 2**) cluster in a continuous RNA-binding groove toward the active site of the Nudix domain (**Fig. 2b**,**c**). We therefore propose that the new orientation of the Nudix domain represents the activated conformation of the decapping complex and that activation primarily occurs through improved binding of the RNA substrate.

Our model rationalizes a plethora of published mutational and biochemical data. The structure of the ternary complex explains why the Edc1 DAm peptide requires its N-terminal DBR residues in addition to the PRS motif to activate decapping[16,17], why activation depends on the presence of the Dcp1 EVH1 domain[6–11,17] and how the specific interaction of the Edc1 DAm peptide with the Dcp2–Dcp1 complex results in the activated conformation.

Decapping activation requires the entire Edc1 DAM peptide

Binding of the Edc1 DAm peptide to the Dcp1 EVH1 domain is primarily mediated through hydrophobic contacts along the entire PRS (**Fig. 3a**). Highly conserved (P167, L172 and F177) and invariant residues (P173 and P175) from Edc1 PRS are accommodated in a hydrophobic groove of the EVH1 domain lined by small aliphatic and highly conserved aromatic (Y36, F38, W45, Y93 and W107) Dcp1 residues (**Fig. 3a**, **Supplementary Fig. 1c** and Dcp1 alignment in **Supplementary Note**). Additional hydrophobic interactions, which bend and orient the path of the conserved ($YAGX_2F$) motif of Edc1 on the Dcp2 NRD domain, are mediated by Edc1 Y158 and F163 (**Fig. 3b**).

Decapping activation by Edc1 strictly depends on the observed interactions with both Dcp1 and Dcp2, as demonstrated in decapping assays *in vitro* (**Fig. 1d** and **Supplementary Fig. 4a–d**). Activation was abolished when the interaction with the EVH1 domain was disrupted by mutations in Dcp1 (Y36A or W107A) or when the N-terminal DBR residues of the Edc1 peptide were deleted (**Fig. 1d** and **Supplementary Fig. 4a–d**). Furthermore, mutation of several corresponding interface residues in Edc1, Dcp1 and Dcp2 NRD of yeast and metazoan orthologs have previously been shown to reduce decapping activity *in vivo* and *in vitro*[6–8,10,11,16–18,20–22] (**Supplementary Table 2**).

The DBR stabilizes the activated conformation

The N-terminal β-strand of Edc1 extends the β-sheet of the Dcp2 Nudix domain via main chain hydrogen bonds (**Fig. 3b**,**c**), thus explaining why the N-terminal residues preceding the $YAGX_2F$ motif are also required for decapping activation although they are poorly conserved (**Supplementary Fig. 1c**). Importantly, the β-strand not only tethers to the Nudix domain but also locks the domain orientation with respect to the NRD–Dcp1 module through a network of interactions (**Fig. 3b–d**).

The extensively characterized and invariant Dcp2 residue W43 on helix α3 of the NRD[10,11,17,22] acts as an anchoring knob that intercalates

between I156 and Y158 of the Edc1 β-strand (**Fig. 3c**). moreover, W43 links to a buried water molecule that also coordinates D163 and T165 from strand β4 of the Nudix domain (**Fig. 3c**,**d**), thus suggesting a mechanism by which the activated orientation of the Nudix domain can be stabilized even in the absence of Edc1. The position of W43 in the activated conformation does not support a function as part of a split catalytic site[10,11], because W43 is completely buried by Edc1. Instead, the role of W43 as a structural anchor for the rotated Nudix domain provides a much simpler rationale for why mutation of W43 blocks the domain closure of Dcp2 as well as the Edc1-mediated activation of decapping[10,11]. It also explains why W43 (as well as D47, which possibly stabilizes the W43 rotamer) is required for full Dcp2 activity *in vitro* and *in vivo* in all organisms that have been tested[7,10,11,22].

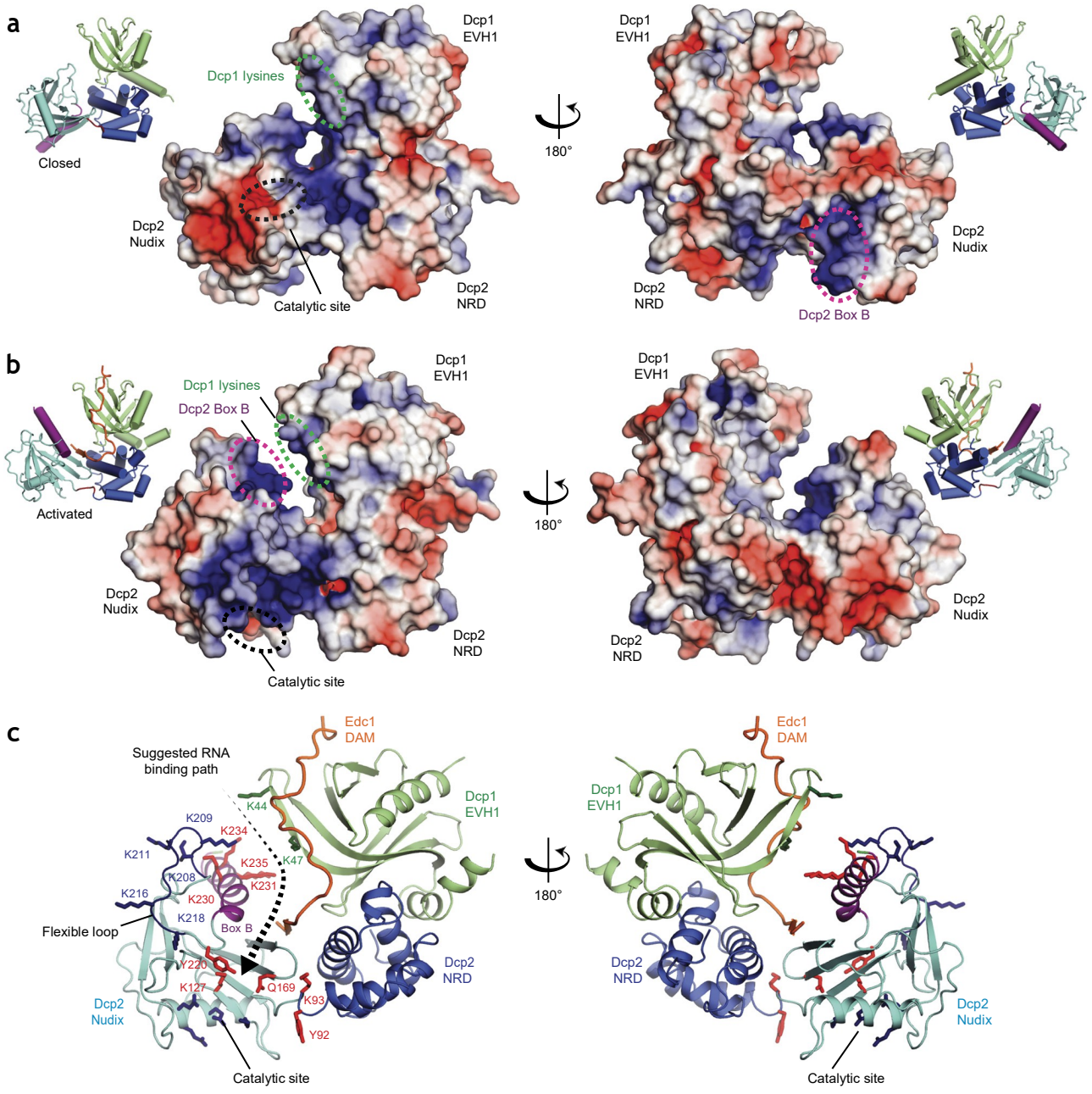


Figure 2 Reorganization of RNA-binding surfaces in the presence of Edc1. (**a**,**b**) Electrostatic surface potential (red to blue, −5 *kT*/*e* to +5 *kT*/*e*) plotted onto the molecular surface of the binary Dcp2–Dcp1 (closed) complex (**a**) or the ternary (activated) Dcp2–Dcp1–Edc1 complex (**b**). Domain orientations are indicated in insets. The rotation of the Nudix domain in the presence of Edc1 forms a deep channel (marked in **c**), lined by positively charged residues on opposing surfaces, as indicated by dotted green (Dcp1) and purple (Box-B helix) circles in **a** and **b**. The surface potentials were calculated as described in Online Methods. (**c**) Model of the activated mRNA-decapping complex explaining improved RNA binding (cartoon, oriented as in **b**). The proposed RNA binding path is highlighted by an arrow pointing toward the catalytic site. Catalytic site residues (dark blue) and previously mutated RNA-binding residues (red) are shown as sticks (**Supplementary Table 2**). RNA-binding residues on Dcp1 identified in this study are show in green. The flexible Nudix loop (F207–K218, dark blue; **Supplementary Note**) that is poised to close upon the RNA substrate is modeled as found in the binary complex with five lysines as sticks.

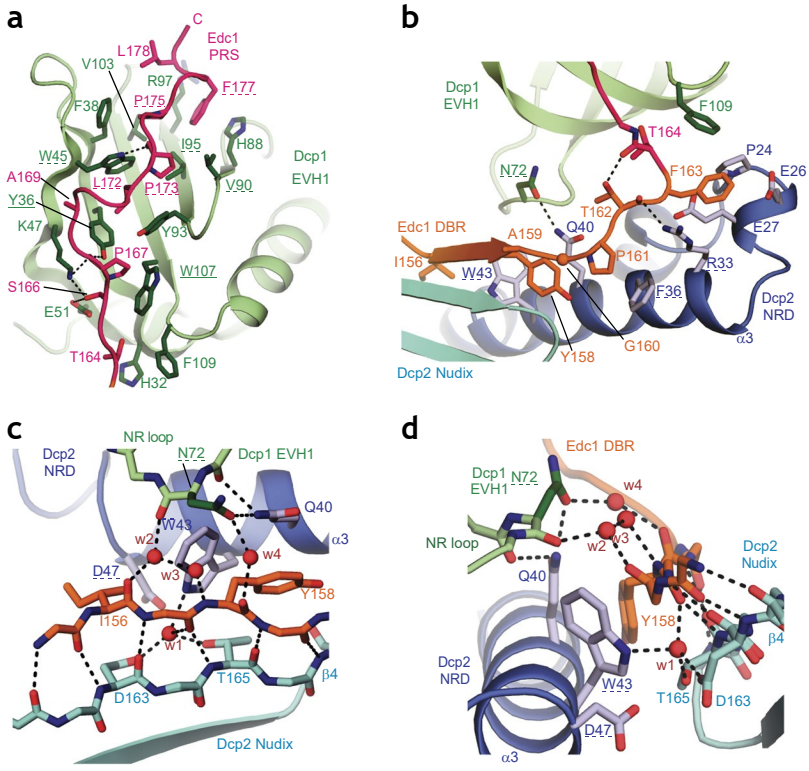


Figure 3 Interactions of the Edc1 DAM peptide fix the activated conformation of the decapping complex. (**a**) Close-up view of the Dcp1-Edc1 PRS interface. (**b**) Close-up view of the Edc1 interaction with the Dcp2 NRD and Nudix domain. (**c**,**d**) Close-up views of the Edc1 interaction with the Dcp2 Nudix domain, illustrating how the Edc1 DBR acts as a molecular bridge between the Nudix domain and the Dcp1–NRD module including a network of water-mediated contacts. Selected residues are shown as sticks with nitrogens in blue and oxygens in red. Selected hydrogen bonds are shown as dotted lines, and selected water molecules are shown as red spheres. Lowercase 'w' denotes waters. Residues mutated in this study and in the literature are highlighted with solid and dashed underlining, respectively.

In solution, small angle X-ray scattering (SAXS) of the Dcp2–Dcp1 complex indicated a mixture between open and closed conformations, although SAXS cannot distinguish between rotations of the Nudix domain and hence cannot determine the fraction of the activated conformation (**Supplementary Fig. 5** and **Supplementary Table 3**). It is clear, however, that activation by Edc1 correlates with increased compactness in SAXS after Edc1 addition, thus supporting that the conformation observed in the crystal represents the activated state.

Improved RNA binding of the activated conformation

The Dcp1 EVH1 domain also contributes to the stabilization of the Nudix-domain orientation in the presence of Edc1. This stabilization occurs via the conserved and functionally important Dcp1 NR-loop[22], in which N72 is fixed by Q40 from the Dcp2 NRD and makes main chain and side chain contacts to three defined water molecules that directly connect to the open edge of the Edc1 β-strand (**Fig. 3b–d**). moreover, the EVH1 domain defines and constrains the proposed RNA-binding channel, which emerges after rotation of the Nudix domain, and it is likely to contribute additional residues (K44 and K47) that cooperate with the previously characterized RNA-binding residues (K230, K231, K234 and K235) on the Nudix Box B helix[6] that at this point lines the other wall of the RNA binding channel (**Fig. 2c** and **Supplementary Table 2**). In agreement with such conformational activation, the RNA-binding activity of the Dcp2–Dcp1 complex increases after addition of Edc1 DAm peptide[17] (**Supplementary Fig. 4e**,**f**), and we found that alanine substitutions of Dcp1 residues K44 and K47 decreased the RNA-binding and decapping activity of the dimeric Dcp2–Dcp1 complex (**Supplementary Fig. 4g–i**).

DISCUSSION

We trapped the Dcp2–Dcp1 mRNA-decapping complex in a conformation that explains the decapping enhancement by Dcp1 and Edc1 primarily via improved binding of the capped mRNA substrate. Furthermore, given the widespread occurrence of the Edc1 and PNRC protein families, the presently described activation mechanism via an intrinsically disordered DAm peptide is probably conserved in all metazoans (**Supplementary Fig. 1c**) and adds an important facet to the functional repertoire of intrinsically disordered proteins in regulating essential enzymatic functions[23].

Our structures, together with previous data reported in the literature[6–11], suggest that in the absence of decapping activators, Dcp2 occurs in an ensemble of conformations, of which the activated conformation represents only a minor fraction. The inherent flexibility of Dcp2 prevents the enzyme from reaching full activity until it is assembled into the complete decapping complex together with Dcp1 and Edc1 (and potentially other decapping enhancers). This control of decapping activity through a cascade of interactions with multiple distinct decapping enhancers, which culminates in the conformational trapping of an activated state, provides multiple opportunities to regulate decapping to ensure that it does not occur inadvertently on functional mRNAs while also ensuring that mRNAs destined for degradation are decapped[2]. Furthermore, regulatory post-translational modifications on Dcp1, Dcp2, Edc1 or PNRC proteins might prevent decapping-complex assembly and hence full decapping activity of Dcp2. The unstructured nature of Edc1 or PNRC proteins makes them especially amenable to modifications, because they would be fully accessible to post-translational modifying enzymes in isolation.

Another interesting implication of our structures is that PRS-containing proteins that lack the Dcp2-binding region would still be able to bind the Dcp1 EVH1 domain but would not be able to facilitate the conformational change leading to decapping activation. These proteins may compete with Edc1 and PNRC proteins for Dcp1 binding and might effectively act as decapping inhibitors. For example, *Drosophila melanogaster* (*Dm*) XRN1 contains a PRS sequence that binds the DCP1 EVH1 domain[20], through a mode similar to those of Edc1 binding (this study) and human PNRC2 binding[18]. However, *Dm* XRN1 lacks a YAGX$_2$F motif. Thus, binding of *Dm* XRN1 to DCP1 may inhibit decapping or alternatively may activate decapping through a different mechanism. It remains to be seen whether PRS-containing decapping inhibitors exist and how decapping activators that lack a PRS stabilize the activated conformation of the Dcp2–Dcp1 complex. These decapping activators may act indirectly in destabilizing unfavorable conformations, such as the open or closed

conformations observed in previously available structures[6] or, alternatively, they may directly participate in cap hydrolysis. For any of these cases, the structural foundations have now been laid.

METHODS

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

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes PDB 5J3Q (*Sp* Dcp1–Edc1 complex), PDB 5J3Y (*Sp* Dcp2–Dcp1 complex) and PDB 5J3T (*Sp* Dcp2–Dcp1–Edc1 complex).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACknOWlEdgMEnTS

This work was supported by the max Planck Society. We thank R. Büttner for excellent technical assistance with crystallization experiments, D. Peter for help with the ITC experiments and J.-m. Schmidt and H. Bhatt for experiments at the initial stage of the project. We thank the staff at the PX beamlines of the Swiss Light Source, Villigen, for assistance with data collection. We thank R. Rambo and N. Cowieson (Diamond Light Source, UK) for assistance with SAXS data collection at beamline B21 of the Diamond Light Source, UK. The SAXS data collection was supported by funding (to E.V. and O.W.) from the European Community's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement no. 283570).

AUTHOR COnTRIBUTIOnS

S.J. and E.I. conceived and designed the project. S.m. carried out protein purifications, pulldown assays and crystallographic work under the supervision of S.J. and E.V. E.V. solved the structures. S.J. performed initial experiments relating to the Dcp2–Dcp1 structure. E.V. and S.m. collected and analyzed the SAXS data. C.-T.C. performed decapping and RNA binding assays and analyzed data. S.m., E.V., S.J. and O.W. analyzed the crystallographic data. E.V., S.J., O.W. and E.I. wrote the manuscript with contributions from S.m. and C.-T.C. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Arribas-Layton, M., Wu, D., Lykke-Andersen, J. & Song, H. Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim. Biophys. Acta* **1829**, 580–589 (2013).
2. Jonas, S. & Izaurralde, E. The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev.* **27**, 2628–2641 (2013).
3. Lykke-Andersen, J. Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol. Cell. Biol.* **22**, 8114–8121 (2002).
4. van Dijk, E. *et al.* Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.* **21**, 6915–6924 (2002).
5. Wang, Z., Jiao, X., Carr-Schmid, A. & Kiledjian, M. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc. Natl. Acad. Sci. USA* **99**, 12663–12668 (2002).
6. She, M. *et al.* Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell* **29**, 337–349 (2008).
7. She, M. *et al.* Crystal structure and functional analysis of Dcp2p from *Schizosaccharomyces pombe*. *Nat. Struct. Mol. Biol.* **13**, 63–70 (2006).
8. Deshmukh, M.V. *et al.* mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol. Cell* **29**, 324–336 (2008).
9. Floor, S.N., Jones, B.N. & Gross, J.D. Control of mRNA decapping by Dcp2: an open and shut case? *RNA Biol.* **5**, 189–192 (2008).
10. Floor, S.N., Jones, B.N., Hernandez, G.A. & Gross, J.D. A split active site couples cap recognition by Dcp2 to activation. *Nat. Struct. Mol. Biol.* **17**, 1096–1101 (2010).
11. Floor, S.N., Borja, M.S. & Gross, J.D. Interdomain dynamics and coactivation of the mRNA decapping enzyme Dcp2 are mediated by a gatekeeper tryptophan. *Proc. Natl. Acad. Sci. USA* **109**, 2872–2877 (2012).
12. Aglietti, R.A., Floor, S.N., McClendon, C.L., Jacobson, M.P. & Gross, J.D. Active site conformational dynamics are coupled to catalysis in the mRNA decapping enzyme Dcp2. *Structure* **21**, 1571–1580 (2013).
13. Steiger, M., Carr-Schmid, A., Schwartz, D.C., Kiledjian, M. & Parker, R. Analysis of recombinant yeast decapping enzyme. *RNA* **9**, 231–238 (2003).
14. She, M. *et al.* Crystal structure of Dcp1p and its functional implications in mRNA decapping. *Nat. Struct. Mol. Biol.* **11**, 249–256 (2004).
15. Dunckley, T., Tucker, M. & Parker, R. Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in *Saccharomyces cerevisiae*. *Genetics* **157**, 27–37 (2001).
16. Schwartz, D., Decker, C.J. & Parker, R. The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. *RNA* **9**, 239–251 (2003).
17. Borja, M.S., Piotukh, K., Freund, C. & Gross, J.D. Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. *RNA* **17**, 278–290 (2011).
18. Lai, T. *et al.* Structural basis of the PNRC2-mediated link between mRNA surveillance and decapping. *Structure* **20**, 2025–2037 (2012).
19. Fromm, S.A. *et al.* The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J.* **31**, 279–290 (2012).
20. Braun, J.E. *et al.* A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. *Nat. Struct. Mol. Biol.* **19**, 1324–1331 (2012).
21. Tharun, S. & Parker, R. Analysis of mutations in the yeast mRNA decapping enzyme. *Genetics* **151**, 1273–1285 (1999).
22. Chang, C.T., Bercovich, N., Loh, B., Jonas, S. & Izaurralde, E. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res.* **42**, 5217–5233 (2014).
23. Oldfield, C.J. & Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).

ONLINE METHODS

DNA constructs. cDNAs encoding *S. pombe* Dcp1, Dcp2 and Edc1 (SPAC18G6.09c) codon-optimized for expression in *Escherichia coli,* were obtained from Life Technologies. The regions of interest were inserted between NdeI and BamHI restriction sites in vectors pnEApG for Dcp1, pnYC and pnYCpm for Dcp2 and pnEAnvmBG[24] for the Edc1 peptide (**Supplementary Table 1**).

Protein expression and purification. All proteins were expressed in *E. coli* BL21(DE3) Star cells (Life Technologies) in LB medium, and their expression was induced with IPTG. For purification of the Dcp2–Dcp1 complex, N-terminally GST-tagged Dcp1 (residues 1–127) was coexpressed with untagged Dcp2 (residues 1–242), and expression was induced with IPTG overnight at 20 °C. The cells were lysed in lysis buffer (50 mm HEPES-NaOH, pH 7.0, 200 mm NaCl and 2 mm DTT supplemented with EDTA-free protease inhibitors (Roche), 1 mg/ml lysozyme and 5 μg/ml DNase I) with an EmulsiFlex-C3 homogenizer (Avestin). After initial purification with glutathione–Sepharose (macherey Nagel), the GST tag was cleaved from the fusion partner overnight with HRV 3C protease. Proteins were then purified on a HiTrap Heparin FF column (GE Healthcare) with a linear gradient to 800 mm NaCl on an ÄKTA Purifier system (GE Healthcare). The final purification step was size-exclusion chromatography on a 16/60 HiLoad Superdex 75 column (GE Healthcare) equilibrated in storage buffer containing 20 mm HEPES-NaOH, pH 7.0, 200 mm NaCl and 2 mm Tris(2-carboxyethyl)phosphine (TCEP).

N-terminally GST-tagged Dcp1 (residues 1–127) was also expressed and purified separately with lysis and storage buffer supplemented with 300 mm NaCl. After initial purification with glutathione–Sepharose, the GST tag was proteolytically cleaved with HRV 3C protease. The protein was then purified by size-exclusion chromatography on a 16/60 HiLoad Superdex 75 column, on which the GST tag was separated from the Dcp1 EVH1 domain.

The Edc1 construct used for pulldown assays carried an N-terminal mBP tag and a C-terminal GB1 tag[25] and was induced with IPTG at 30 °C for 5 h. The cells were lysed in lysis buffer containing 200 mm NaCl. mBP-Edc1-GB1 was purified with amylose resin (New England BioLabs) and then by anion-exchange chromatography with a HiTrap Q FF column (GE Healthcare) and elution with a linear gradient to 1 m NaCl. The final purification step was size-exclusion chromatography with a 16/60 HiLoad Superdex 75 column (GE Healthcare) equilibrated in storage buffer.

The Dcp2 construct used in pulldown assays was also mBP-tagged and expressed and purified with the same procedure described for Edc1. mutants of Dcp2, Dcp1 and Edc1 were purified with the same procedure described for the wild-type proteins. After purification, proteins were concentrated with Amicon centrifugal filter units (millipore) and used immediately for experiments or flash frozen in liquid nitrogen and stored at −80 °C.

Crystallization. For crystallization experiments, the *S. pombe* Edc1 peptide (residues S155–S180) was chemically synthesized by EmC microcollections and dissolved in 20 mm HEPES-NaOH, pH 7.0, and 200 mm NaCl to a 4 mm final concentration.

For crystallization of the ternary Dcp2–Dcp1–Edc1 complex, the purified $Dcp2_{1–242}$–$Dcp1_{1–127}$ complex (200 μm) was mixed with Edc1 peptide (500 μm) at a 1:2.5 (Dcp2–Dcp1–Edc1) molar ratio and incubated for 30 min on ice before crystallization. Crystals of the Dcp2–Dcp1–Edc1 ternary complex grew within 24 h by hanging-drop vapor diffusion at 22 °C. Drops contained 1 μl protein complex and 1 μl reservoir solution consisting of 1.55 m sodium formate and 0.1 m sodium acetate, pH 4.8. Before crystallization, 6% (v/v) glycerol was added to the protein complex. Crystals appeared after 1 d and grew to maximum size after 3 d.

For crystallization of the binary Dcp1–Edc1 complex, 400 μm Dcp1 (in storage buffer containing 300 mm NaCl) was mixed with 1 mm Edc1 peptide at a 1:2.5 (Dcp1–Edc1) molar ratio and incubated for 30 min on ice. After mixture of 200 nl complex with 200 nl reservoir solution containing (0.1 m PCB buffer sodium propionate, sodium cacodylate and bis-Tris propane in a molar ratio of 2:1:2, pH 6.0, and 25% (w/v) PEG 1500), crystals grew within 12–24 h by sitting-drop vapor diffusion.

Crystals of the Dcp2–Dcp1 complex were obtained by hanging-drop vapor diffusion at 22 °C. Crystals appeared 3 d after mixture of 1 μl of protein complex (10 mg/ml, 0.23 mm) and 1 μl of reservoir solution (0.1 m mES-NaOH, pH 6.5, 0.2 m ammonium sulfate and 30% (w/v) PEG 5000 mmE). For cocrystallization with cap analog or guanosine monophosphate, the Dcp2–Dcp1 protein complex was preincubated for 30 min on ice in a mixture containing 0.23 mm protein complex, 2 mm $m^7G(5')ppp(5')G$ cap analog (New England BioLabs) and 2 mm $mgCl_2$, or 2 mm guanosine monophosphate and 2 mm $mgCl_2$, in 0.8× storage buffer (16 mm HEPES-NaOH, pH 7.0, 160 mm NaCl and 1.6 mm TCEP).

For cryoprotection, Dcp1–Dcp2 crystals were briefly transferred to 0.1 m mES-NaOH, pH 6.5, 0.2 m ammonium sulfate and 30% (w/v) PEG 5000 mmE supplemented with 20% (v/v) glycerol and 2 mm cap analog before being flash frozen in liquid nitrogen. Dcp1–Edc1 crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol. Dcp2–Dcp1–Edc1 crystals were cryoprotected in 4 m sodium formate.

Data collection and structure determination. All diffraction data were recorded from a single crystal per complex at a wavelength of 0.9999 Å on a PILATUS 6m detector (Dectris) at the PXII beamline of the Swiss Light Source (Villigen, Switzerland). All data were processed with XDS[26] and scaled with AImLESS[27,28] from the CCP4 suite[29].

The Dcp2–Dcp1 data extended to a resolution of 3.3 Å and were processed in space group $P6_522$, thus indicating a new crystal form compared with those from previously published data[6]. The structure was solved by molecular replacement with PHASER[30] from the CCP4 suite and with the closed *Sp* Dcp2–Dcp1 complex (PDB 2QKm, chains A and B with protein atoms only)[6] as a search model.

molecular replacement resulted in two Dcp2–Dcp1 complexes per asymmetric unit with similar conformations. This was followed by an initial rigid-body (one body per domain) and full-atom refinement in REFmAC5 (ref. 31) from the CCP4 suite, which resulted in sufficiently well-defined difference density for us to start adjusting and rebuilding the structure. The model was improved by iterative cycles of building and refinement with COOT[32] and REFmAC5, respectively. The following residues are missing from the model: Dcp1, L125–R127 in chains A and C; and Dcp2, L75–W78 and T242 in chain D. Final refinement was carried out with PHENIX.REFINE[33] and with noncrystallographic symmetry torsional (NCS) restraints as well as translation/libration/screw (TLS) parameterization[34]. The final model was refined to R_{work}/R_{free} values of 0.224/0.252. The stereochemistry of the structure was validated with mOLPROBITY[35], and all residues were in the favored/allowed region of the Ramachandran plot. Despite the presence of cap analog in the crystallization condition, there was only very weak electron density in the former ATP-binding site of the Nudix domain[6] near Y220 of one of the Dcp2 copies in the complex (chain C). This result suggests that the nucleotide occupancy at this site is extremely low at best, and modeling of the ligand was not attempted.

The Dcp1–Edc1 data was processed in space group $P2_12_12_1$ to a resolution of 1.9 Å. The structure was solved with PHASER as implemented in PHENIX[36], despite the presence of strong translational pseudosymmetry, as indicated by PHENIX.XTRIAGE[37]. As a search model, we used the *Sp* Dcp1 structure, as determined from the binary *Sp* Dcp1–Dcp2 complex, identifying two translationally related copies per asymmetric unit. Continuous $F_o - F_c$ difference electron density indicated the presence of the Edc1 DAm peptide. We then used ARP/wARP[38], as implemented in the CCP4 suite to autobuild the two Dcp1–Edc1 complexes in an unbiased manner. The model was completed by several iterative rounds of manual adjustment in COOT and refined with BUSTER version 2.10.2 (https://www.globalphasing.com/buster/) with automated NCS restraints and TLS refinement[39]. The final model comprises two Dcp1–Edc1 complexes in the asymmetric unit and 142 water molecules. However, only the C-terminal PRS of the Edc1 DAm peptide was ordered. The following residues are missing from the model: Dcp1, m1 and E2 in chains A and C; Edc1, S155–T164 and H179–S180 in chain B; and S155–P161 and F177–S180 in chain D. The final model has R_{work}/R_{free} values of 0.207/0.242, and the stereochemistry was validated with mOLPROBITY. One residue, S42 of Dcp2 chain C, was in the disallowed region of the Ramachandran plot, and the rest were in the favored/allowed region.

The Dcp2–Dcp1–Edc1 data were processed in space group *C*2 to a resolution of 1.6 Å. The structure was solved by molecular replacement with PHASER from the CCP4 suite. We performed consecutive searches, first placing the Dcp1 EVH1 domain, then the Dcp2 NRD and then the Dcp2 Nudix domain (one copy each, as obtained from the binary Dcp1–Dcp2 complex). We then performed manual adjustment of the model in COOT and refinement in PHENIX.REFINE, which resulted in continuous difference electron density for the Edc1 DAm peptide in

doi:10.1038/nsmb.3232

the $F_o - F_c$ map. We then used ARP/wARP, as implemented in the CCP4 suite to autorebuild the entire complex including the Edc1 DAm peptide (residues S155–H179) in an unbiased manner. The model was completed by several iterative rounds of manual adjustment in COOT and refinement with BUSTER with TLS parameterization. The final model comprises one Dcp2–Dcp1–Edc1 complex in the asymmetric unit and 215 water molecules, five molecules of formic acid and one magnesium ion. The following residues are missing from the model: Dcp1, m1 in chain A; Dcp2, m1, F207–K216 and T242 in chain B; and Edc1, S180 in chain C. The final model has R_{work}/R_{free} values of 0.186/0.219, and the stereochemistry of the structure was validated with mOLPROBITY. One residue, E99 of Dcp1 chain A, was in the disallowed region of the Ramachandran plot, and the rest were in the favored/allowed region.

All structural figures were generated with PymOL version 1.8 (http://www.pymol.org/). For calculating the electrostatic surface potentials in **Figure 2**, loop residues F207–K216, which are disordered in the Dcp2–Dcp1–Edc1 structure, were also deleted in the Dcp2–Dcp1 structure for consistent visual comparison. The N- and C-terminal residues that were missing in either of the two structures were deleted in the other, such that the termini of both structures had identical visible residues in Dcp2 and Dcp1 chains. Calculation of the Poisson–Boltzmann equation at 310 K with the dielectric constants of 2.0 and 78.0 for protein and solvent, respectively, was performed with the Adaptive Poisson-Boltzmann Solver (APBS Tools version 2.1 in PymOL)[40]. The X-ray data collection and refinement statistics are presented in **Table 1**.

Pulldown assays. For the pulldown assays, equimolar amounts (40–50 µg, 2–5 µl) of mBP, mBP-Edc1-GB1 or mBP-Dcp2 (in storage buffer) were mixed with purified, untagged Dcp1 wild type or Y36A or W107A mutants (40–50 µg, 20–40 µl) (in storage buffer containing 300 mm NaCl). The protein mixtures were supplemented with 50 µl (50% slurry) of amylose resin (New England BioLabs). The volume was adjusted up to 1 ml with pulldown buffer (50 mm HEPES/NaOH, pH 7.0, 200 mm NaCl and 2 mm DTT), and samples were gently rotated for 1 h at 4 °C. The beads were washed three times for 1 min with 1 ml pulldown buffer. The proteins were eluted from the beads with pulldown buffer supplemented with 25 mm maltose. They were then precipitated with ice-cold 20% (v/v) trichloroacetic acid, resuspended in 40 µl denaturing SDS–PAGE sample buffer and analyzed by SDS–PAGE.

Decapping assays. The decapping assays were performed with 5 nm purified proteins and 0.5 µm *in vitro*–synthesized RNA (127 nucleotides) trace-labeled with [γ-^{32}P]GTP with a ScriptCap m^{7}G Capping System and ScriptCap 2′-O-methyltransferase kit (Epicentre Biotechnologies). Decapping reactions were carried out at 30 °C for the indicated times in a total volume of 10 µl of decapping buffer (50 mm Tris-HCl, pH 7.5, 50 mm ammonium sulfate, 0.1% BSA and 5 mm $mgCl_2$). Proteins were diluted to their working concentrations with the decapping buffer. Reactions were stopped by addition of up to 50 mm EDTA (final concentration), and 1 µl of each sample was spotted on polyethylenimine (PEI) cellulose thin-layer chromatography plates (merck) and developed in 0.75 m LiCl (**Supplementary Fig. 4a**,**h**). The released m^{7}GDP signal was quantified with ImageQuant TL software (GE Healthcare) and normalized to that of the total signal (capped RNA+ m^{7}GDP signal) for each condition. These normalized values were plotted against time (**Fig. 1d**). Error bars represent s.d. from three independent experiments.

Electrophoretic mobility shift assays (EMSA). The plasmid pcDNA3.1 (Life Technologies) was linearized with KpnI and used for *in vitro* transcription by T7 RNA polymerase. The 5′ end of the 38-nucleotide transcript was labeled with T4 Polynucleotide Kinase (Thermo) with [γ-^{32}P]ATP. Binding reactions containing 10 µm purified proteins and 5 nm [^{32}P]RNA were incubated on ice for 15 min in a total volume of 10 µl of 10 mm Tris-HCl, pH 7.4, 5 mm NaCl, 2.5 mm $mgCl_2$, 0.05% (w/v) BSA and 5% (v/v) glycerol. The RNA–protein complexes were analyzed by electrophoresis on 10% nondenaturing polyacrylamide gels in TBE buffer, pH 8.3, at 10 V cm^{-1}.

Isothermal calorimetry (ITC) analysis. The ITC experiments were performed with a VP-ITC microcalorimeter (microCal) at 20 °C. The solution of *Sp* Dcp2–Dcp1 at a concentration of 10 µm in the calorimetric cell (cell volume 1.4 ml) was titrated with a 20-fold-concentrated solution of synthetic Edc1 peptide (200 µm) that was dissolved in the same buffer (50 mm HEPES-NaOH, pH 7.0, 200 mm NaCl and 2 mm TCEP). The titration experiments consisted of an initial injection of 2 µl followed by 28 injections of 10 µl at an interval of 240 s and a stirring rate of 300 r.p.m. Each binding experiment was repeated three times. The thermodynamic parameters were calculated with a one-site binding model (Origin version 7.0), whereby the data point of the first injection was removed for the analysis[41].

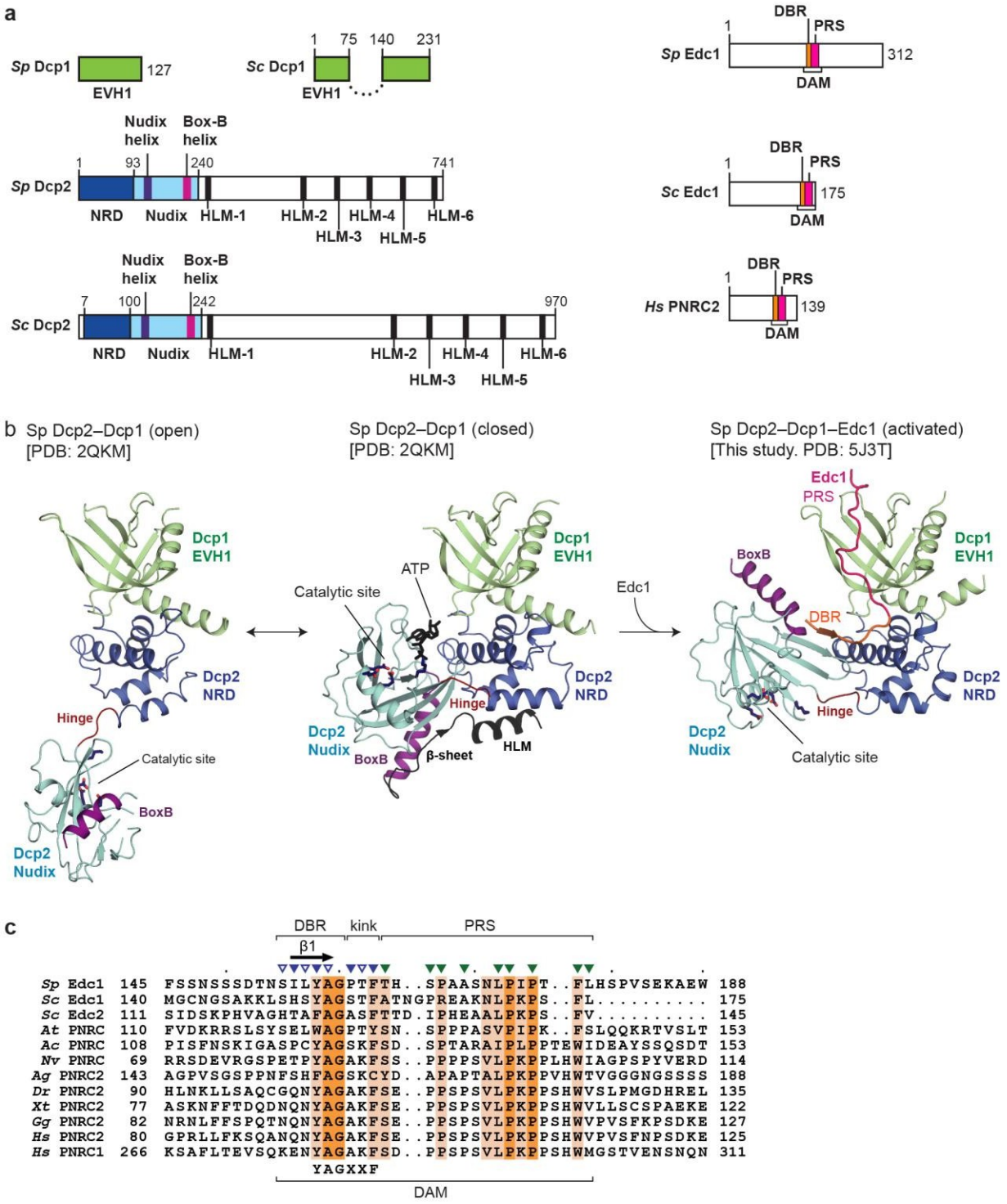
SAXS experiments. SAXS measurements were performed on beamline B21 at Diamond Light Source, Didcot, UK Samples were prepared on site with a Shodex Kw-403 size-exclusion column and Agilent HPLC (Agilent Technologies). Approximately 40–60 µl of sample was injected for SAXS at 20 °C with a sample-to-detector distance of 3.9 m and X-ray wavelength of 1 Å. Continuously eluting samples were exposed for 300 s in 10-s acquisition blocks. A low concentration of sucrose (1% v/v) was used as a free-radical scavenger, in buffer containing 20 mm HEPES-NaOH, pH 7.0, 200 mm NaCl and 2 mm TCEP, to minimize the radiation damage of the sample. Data were first normalized to the intensity of the transmitted beam. Then the scattering data from the buffer immediately before and after the protein peak were averaged and used to subtract the background. Images were corrected for variations in beam current, normalized for exposure time and processed into 1D scattering curves with in-house software. Buffer subtractions and all other subsequent analyses were performed with ScÅtter version 2.2b (http://www.bioisis.net/scatter/) and with programs within the ATSAS suite[42]. The forward scattering $I(0)$ and the radius of gyration R_g were computed with the Guinier approximation for $qR_g < 1.3$. The excluded volume of the particle, V, was computed with the Porod equation. The distance distribution function $P(r)$ was calculated with the indirect Fourier transform method implemented in GNOm[42]. This was limited to $q \leq 2.5$ nm^{-1} to minimize the error contribution from larger q angles.

SAXS ensemble and flexibility analysis. We also examined flexibility of the Dcp2–Dcp1 heterodimer in the presence and absence of Edc1 with the ensemble optimization method (EOm)[43]. The structures were treated as two rigid bodies, and the first body, comprising Dcp1–Dcp2$_{1–92}$ (EVH1 NRD domains) was fixed in space, whereas the second rigid body, corresponding to the catalytic Nudix domain of Dcp2 (residues P97–T242), was permitted to sample the full available range of conformational space, provided that it did not clash with the fixed body. The two rigid bodies were connected by a flexible linker of four dummy residues (corresponding to residues K93–I96 of Dcp2; **Supplementary Fig. 5a**,**b**), and a pool of 10,000 individual models was generated containing a random sampling of linker conformations and subunit positions. A genetic algorithm was then used to select an optimized ensemble of models whose combined theoretical scattering curve best represents the experimental SAXS profile. The EOm analysis of Dcp2–Dcp1 SAXS data yielded an optimized ensemble (containing two models) that fits the experimental scattering curve with a χ^2 value of 0.89 (**Supplementary Fig. 5a**). In the ensemble, the extended conformation model accounted for a fraction of 33% of the populated conformer species, whereas the compacted conformer accounted for a fraction of 67%. The EOm analysis of the Dcp2–Dcp1–Edc1 ternary complex SAXS data yielded an optimized ensemble (containing three models) that fit the experimental scattering curve with χ^2 value of 0.87 (**Supplementary Fig. 5b**). In the ensemble of models, the fraction of the populated conformers corresponding to the extended conformation decreased to just 10%, whereas the two compacted models, which accounted for 90% of the scattering, were both structurally highly similar to the closed conformation of the Dcp2–Dcp1 heterodimer. Analysis of the flexibility of the selected ensembles versus the initial random pool suggested an increase in the overall compaction, as quantified by a decrease in the R_{flex} metric[43] for the Dcp2–Dcp1–Edc1 complex and the progression toward a uniform, less flexible system than that of the Dcp2–Dcp1 complex in absence of Edc1 (**Supplementary Fig. 5c**).

24. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J. Struct. Biol.* **175**, 178–188 (2011).
25. Cheng, Y. & Patel, D.J. An efficient system for small protein expression and refolding. *Biochem. Biophys. Res. Commun.* **317**, 401–405 (2004).

26. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
27. Evans, P.R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 (2011).
28. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
29. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
30. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
31. Murshudov, G.N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
32. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
33. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
34. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).
35. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
36. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
37. Zwart, P.H., Grosse-Kunstleve, R.W., Lebedev, A.A., Murshudov, G.N. & Adams, P.D. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 99–107 (2008).
38. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
39. Smart, O.S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380 (2012).
40. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).
41. Mizoue, L.S. & Tellinghuisen, J. The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal. Biochem.* **326**, 125–127 (2004).
42. Petoukhov, M.V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
43. Tria, G., Mertens, H.D., Kachala, M. & Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).

doi:10.1038/nsmb.3232

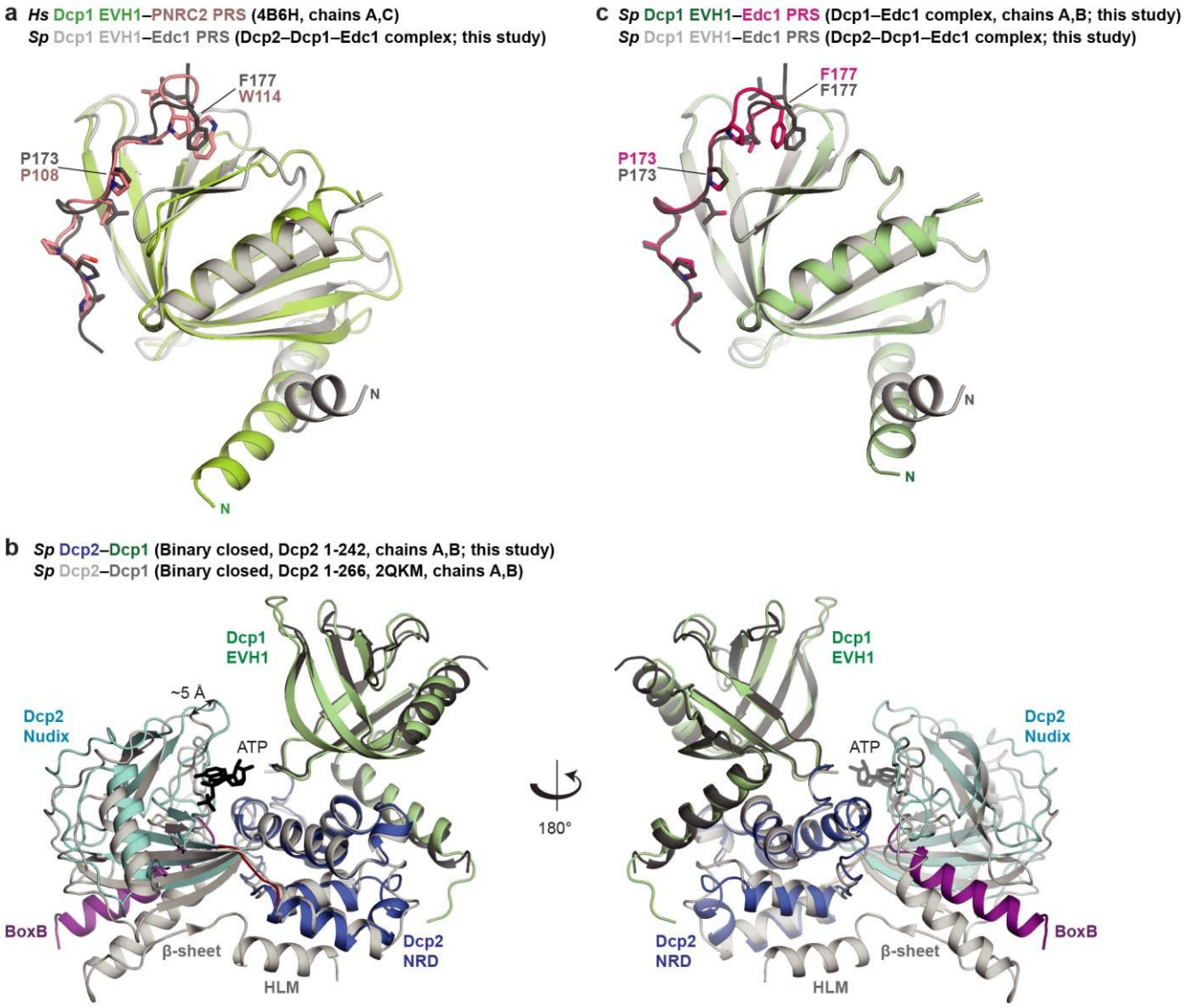


Supplementary Figure 1

Domain architecture and conformational states of the decapping complex, as revealed by structural studies.

(**a**) Domain organization of *Schizosaccharomyces pombe* (*Sp*) and *Saccharomyces cerevisiae* (*Sc*) Dcp1, Dcp2 and Edc1 proteins. The numbers above the schematic represent amino acid positions at domain/motif boundaries. Dcp1 contains an Ena/Vasp homology domain1 (EVH1). The *Sc* Dcp1 EVH1 domain has an unstructured insertion (dotted line). The Dcp2 proteins have an N-terminal regulatory domain (NRD) and a catalytic Nudix domain. The Dcp2 C-terminus is highly divergent and interspersed with six helical

leucine rich motifs (HLMs) in both *Sc* and *Sp*. The Edc1 proteins feature a conserved decapping activator motif (DAM) consisting of Dcp2-binding residues (DBR, comprising a $YAGX_2F$ motif) followed by a C-terminal proline-rich sequence (PRS) that binds to the Dcp1 EVH1 domain. (**b**) Cartoon representations of binary and ternary decapping complexes. The Dcp1 (green), Dcp2 N-terminal regulatory domain (NRD) (blue) and Dcp2 catalytic Nudix domain (cyan) adopt different relative orientations. The hinge region between the NRD and Nudix domains is shown in red. The βstrand extension and the HLM1 helix deleted in the Dcp2 construct used in this study are colored in black. The Box B helix of the Nudix domain is shown in purple. Conserved catalytic site residues in the Nudix domain are shown as purple sticks. The ATP co-crystallized with the closed conformation of the Dcp2–Dcp1 complex is shown as black sticks. The Edc1 DAM peptide is subdivided into a Dcp1-binding proline-rich sequence (PRS; magenta) and a Dcp2-binding residues (DBR; orange). (**c**) Sequence alignment of the DAM region of Edc1, Edc2 and PNRC proteins. Fully conserved residues are highlighted in dark orange, and those with >70% similarity are shown in light orange. Residues involved in Dcp1 and Dcp2 binding are indicated by green and blue triangles, respectively. Filled triangles mark residues contributing with their side-chain to the interface, and empty triangles signify pure backbone interactions. Species abbreviations are as follows: *Schizosaccharomyces pombe* (*Sp*), *Saccharomyces cerevisiae* (*Sc*), *Arabidopsis thaliana* (*At*), *Ancylostoma ceylanicum* (*Ac*), *Nematostella vectensis* (*Nv*), *Anopheles gambiae* (*Ag*), *Danio rerio* (*Dr*), *Xenopus tropicalis* (*Xt*), *Gallus gallus* (*Gg*), *Homo sapiens* (*Hs*).

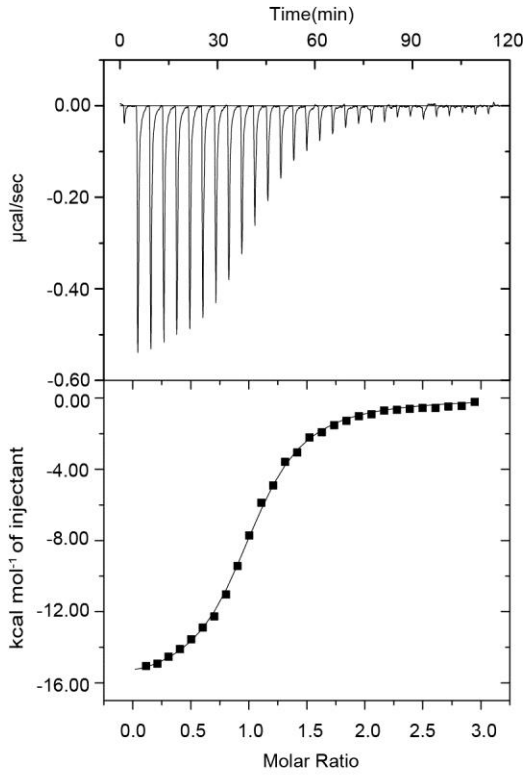


Supplementary Figure 2

Structural superpositions.

(a) Structural superposition of the Dcp1–Edc1 complex (colored in gray; this study) onto the structure of the human DCP1 EVH1 domain bound to the PRS peptide from PNRC2 (green, salmon; PDB code 4B6H)[18]. The RMSD is 1.57 Å over 124 equivalent residues. (b) Structural superposition of the Dcp2–Dcp1 heterodimer from this study (blue, green) onto the closed conformation of the Dcp2–Dcp1 binary complex bound to ATP (gray; PDB code 2QKM)[6]. The alignment is based on the Dcp1–NRD module of the complex (RMSD 1.5 Å over 208 equivalent residues) and shows that the Dcp2 Nudix domain is moved slightly closer (by ~5 Å) towards Dcp1 compared to the previous closed conformation (PDB code 2QKM)[6]. The ATP molecule is shown as black sticks. (**c**) Structural superposition of the Dcp1–Edc1 complex (colored in green and magenta) onto the structure of the Dcp2–Dcp1–Edc1 ternary complex (gray). The only structural difference in Dcp1 is a kink in the N-terminal helix in the ternary complex structure induced by crystal contacts. The RMSD is 0.72 Å over 129 equivalent residues.

a



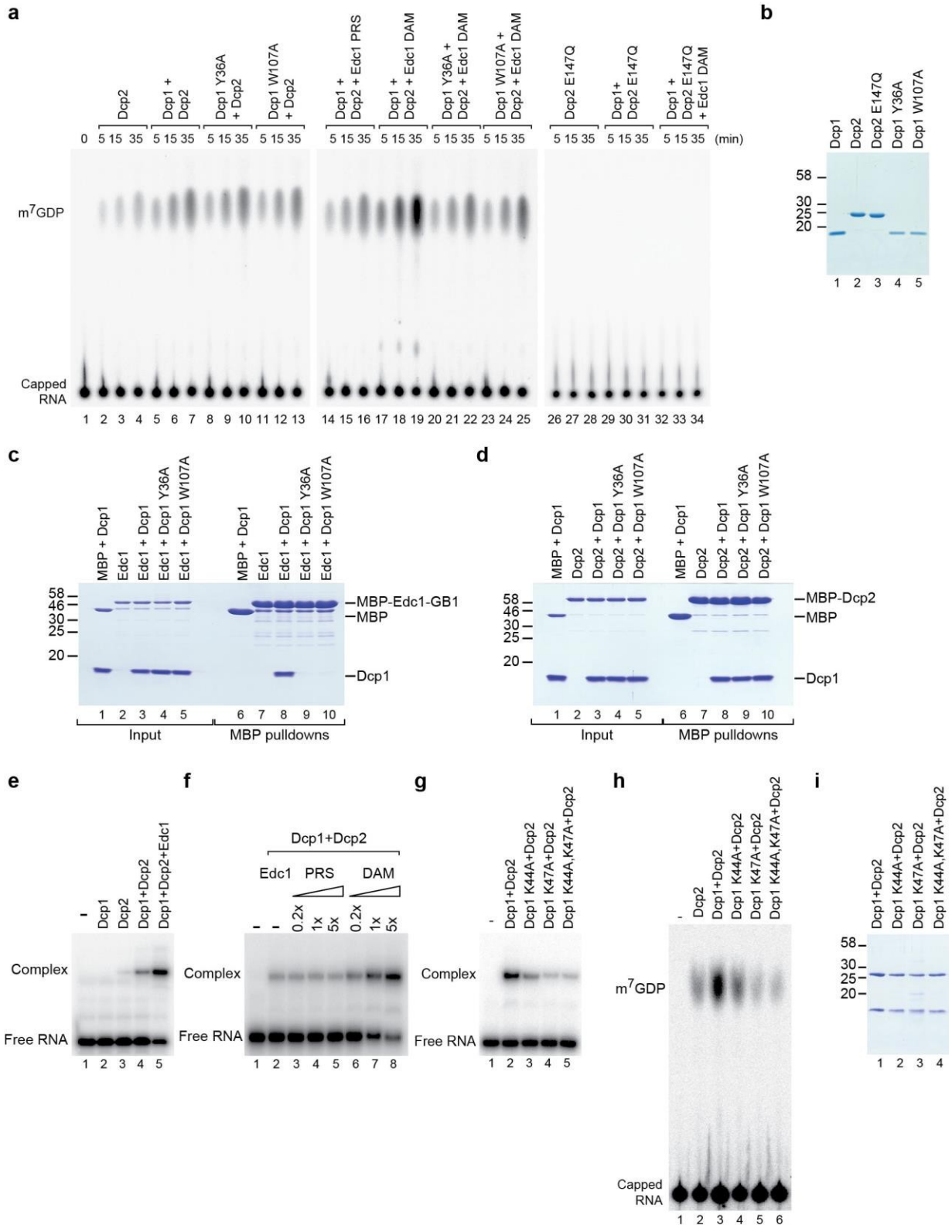
b Thermodynamic parameters for the interaction of Dcp2–Dcp1 with Edc1

K_D (M)	ΔH (kcal mol^{-1})	-TΔS (kcal mol^{-1})	ΔG (kcal mol^{-1})	Molar ratio
$6.6 \pm 1.1 \cdot 10^{-7}$	-15.86 ± 0.3	7.56 ± 0.4	-8.3	1.04 ± 0.03

Supplementary Figure 3

Thermodynamic parameters for the interaction of the Dcp2–Dcp1 complex with the Edc1 DAM peptide.

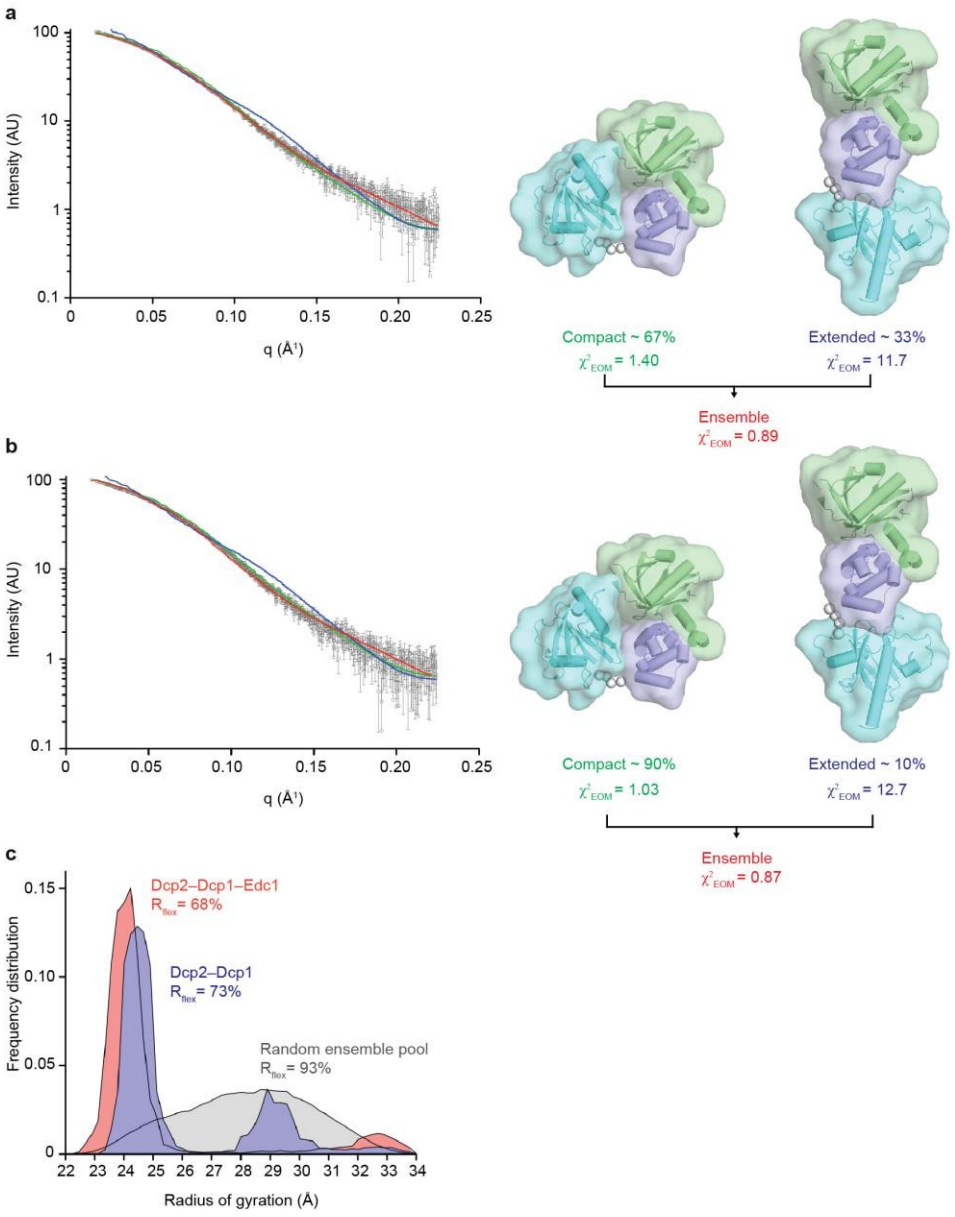
(**a**) Isothermal titration calorimetry (ITC) profile showing the interaction of the Dcp2**–**Dcp1 complex with the Edc1 DAM peptide. The top panel represents the raw data (µcal sec^{-1}) whereas the bottom panel shows the integrated data (kcal mol^{-1} of injectant) of the heat changes that were fitted using a one-site binding model. (**b**) Thermodynamic parameters for the interaction of the Dcp2**–**Dcp1 complex with the Edc1 DAM peptide. The ITC experiments were performed in triplicate, and mean values ± standard deviations are indicated.



Supplementary Figure 4

Contribution of the Edc1 DAM peptide and the Dcp1 EVH1 domain toward decapping activity.

(**a**) *In vitro* decapping assay using the indicated proteins and Edc1 DAM peptide (S155–S180) in comparison with the Edc1 PRS peptide (S166–S180) lacking the DBR portion. Decapping assays were performed over a time course of 35 min. The m^7GDP signals were quantified and normalized to that of the total signal (capped RNA+ m^7GDP signal) for each condition. These normalized values plotted against time are shown in **Figure 1d**. Error bars in **Figure 1d** represent standard deviations from three independent experiments. The samples contained in addition, Dcp2 alone or Dcp2 and Dcp1 (wild-type or the indicated mutants). A Dcp2 E147Q catalytic mutant served as negative control. (**b**) SDS-PAGE analysis of the proteins used in the decapping assay shown in **a**. (**c**) MBP pulldown assay showing that Dcp1 Y36A and W107A mutants do not interact with MBP-Edc1-GB1. (**d**) *In vitro* MBP pulldown assay showing that the Dcp1 Y36A and W107A mutants still interact with MBP-Dcp2. (**e, f**) RNA gel shift assays with the indicated proteins. The RNA-binding activity of Dcp2 is enhanced by Dcp1 and further enhanced by the Edc1 DAM peptide (**e**), in a concentration dependent manner (**f**). The PRS peptide is not sufficient to increase RNA-binding (**f**). (**g**) RNA gel shift assays using Dcp2–Dcp1 complexes containing Dcp1 wild-type or the indicated mutants. (**h**) A decapping assay was performed as described in panel **a** with Dcp2–Dcp1 complexes containing Dcp1 wild-type or the indicated mutants. Note that the Dcp1 K44A and K47A mutants reduce Edc1 binding. Therefore, RNA binding and decapping assays shown in panels **g** and **h** were performed in the absence of Edc1. (**i**) Proteins used in panels **g** and **h**.



Supplementary Figure 5

Analysis of the conformational dynamics of the decapping complex by using small angle X-ray scattering (SAXS).

(a) The simulated curves of the selected EOM ensemble (red) and individual conformer species (compact in green and extended in blue) is compared with the experimental SAXS logarithmic curve with open circles and error bars. The principal conformers obtained by EOM together with their population fraction expressed as percentage are shown. The $\chi 2$ values represent the goodness-of-fit of simulated curves to the SAXS data for the Dcp2–Dcp1 complex. (**b**) EOM conformer modeling analysis with the Dcp2–Dcp1–Edc1 SAXS data. (**c**) EOM Rg distribution of the initial random pool (gray) and the selected ensembles for Dcp2–Dcp1 (blue) and Dcp2–Dcp1–Edc1 (red). The bimodal distribution for the decapping complex in the absence of Edc1 shows greater conformational heterogeneity with two peaks representing the compact and extended conformers at the peak R_g values of 24.5 Å and 29 Å, respectively. In presence of Edc1, the distribution becomes more homogeneous in favor of the compact conformer. Increase in conformer compaction is reflected in the decrease of the EOM-derived R_{flex} metric from 73% for Dcp2–Dcp1 to 68% Dcp2–Dcp1–Edc1 complex.

Supplementary Table 1. Constructs used in this study.

Name	**Dcp1**	**Vector***
Dcp1	*Sp* Dcp1 1−127	pnEA npG *Sp* Dcp1 1−127
Dcp1 Y36A	*Sp* Dcp1 1−127 Y36A	pnEA npG *Sp* Dcp1 1−127 Y36A
Dcp1 W107A	*Sp* Dcp1 1−127 W107A	pnEA npG *Sp* Dcp1 1−127 W107A
Dcp1 K44A	*Sp* Dcp1 1−127 K44A	pnEA npG *Sp* Dcp1 1−127 K44A
Dcp1 K47A	*Sp* Dcp1 1−127 K47A	pnEA npG *Sp* Dcp1 1−127 K47A
Dcp1 K44A,K47A	*Sp* Dcp1 1−127 K44A K47A	pnEA npG *Sp* Dcp1 1−127 K44A,K47A

Name	***Dcp2***	***Vector***
Dcp2	*Sp* Dcp2 1−242	pnYC *Sp* Dcp2 1−242
Dcp2 E147Q	*Sp* Dcp2 1−242 E147Q	pnYC npM *Sp* Dcp2 1−242 E147Q
MBP-Dcp2	MBP *Sp* Dcp2 1−242	pnYC npM *Sp* Dcp2 1−242

Name	***Edc1***	***Vector***
MBP-Edc1-GB1	*Sp* Edc1 S155−S180	pnEA nvMBG *Sp* Edc1 155−180
Edc1 peptide	*Sp* Edc1 S155−S180	Synthetic peptide
Edc1 short peptide	*Sp* Edc1 S166−S180	Synthetic peptide

* Vector and construct annotation:
pnEA − derived from the pET−MCN (Multi Cloning and expression) series using the pET15b backbone bearing ampicillin resistance;
pnYC − pACYC11b backbone bearing chloramphenicol resistance;
npG – N-terminal GST followed by 3C protease site;
npM – N-terminal MBP followed by 3C protease site;
nvMBG – N-terminal MBP followed by TEV protease site with C-terminal GB1

Supplementary Table 2. Dcp2, Dcp1 and Edc1 mutants reported in the literature.

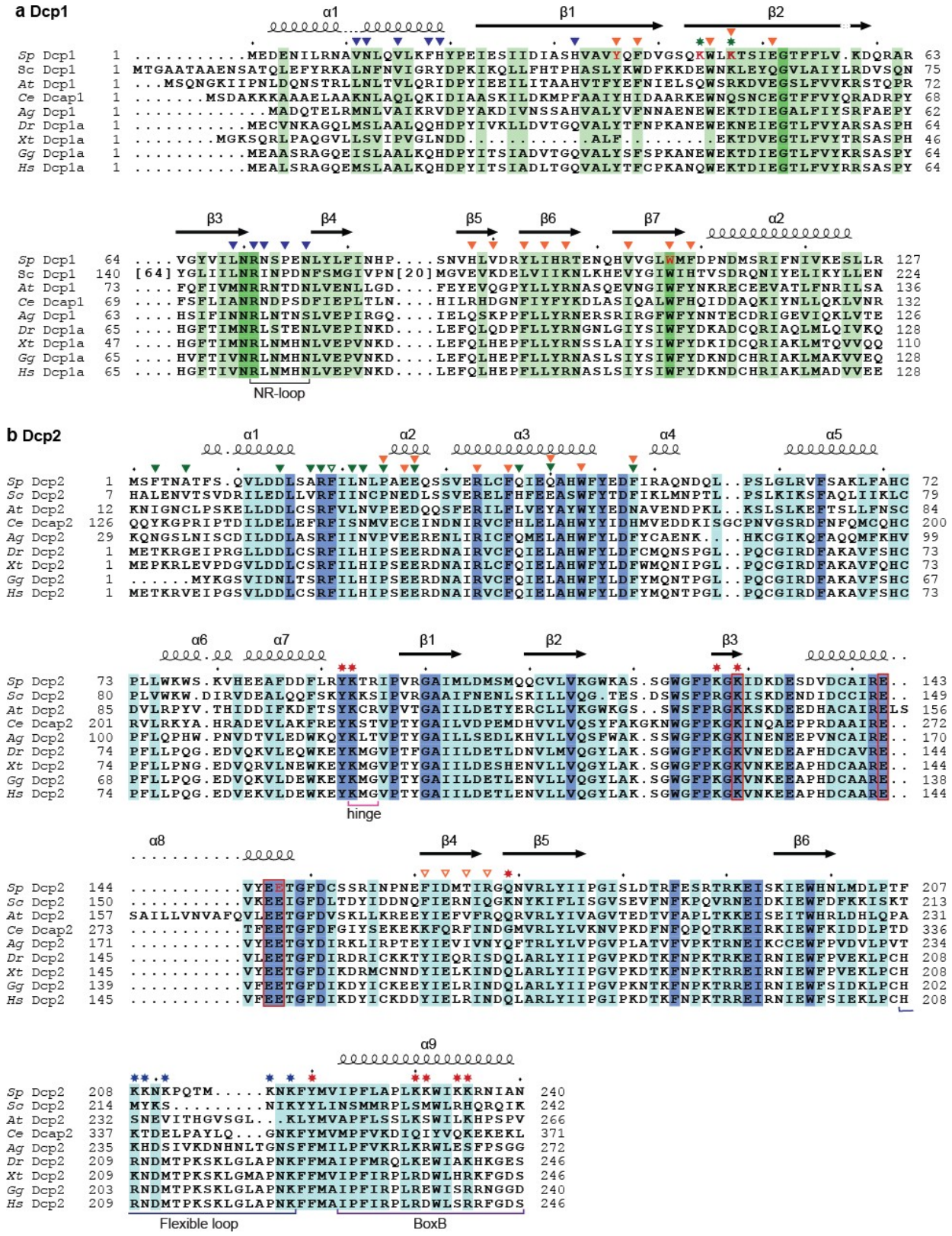
	Residue *Sp*	Orthologs tested	Effect of mutation	Ref.
Dcp1				
PRS binding cleft	**Y36**	*Dm* Y34, *Sc* Y47	reduced decapping *in vivo* & *in vitro*; reduced Edc1 binding	17, 20, 21
	W45	*Sc* W56	reduced decapping *in vivo*	21
	V90	*Sc* K187	reduced decapping *in vivo*	21
	D91	*Sc* D188	reduced decapping *in vivo*	21
	I95	*Hs* L96	abolishes PNRC2 binding	18
	W107	*Sc* W204	reduced decapping *in vivo* & *in vitro*; reduced Edc1 binding	17, 21
NR loop	**R71 - P74**	*Hs* R72-M75	reduced decapping *in vivo* & *in vitro*	22
Dcp2				
Edc1 binding	**R33**	*Sc* R40	reduced decapping *in vivo*	7
	F36	*Sc* F43	reduced decapping *in vivo*	7
NRD-Nudix interface	**W43**	*Hs* W44, *Sc* W50, *Sp*	NMR shift upon ATP addition; blocked conformational change; reduced activation by Dcp1/Edc1; reduced decapping *in vivo* & *in vitro*	10, 11, 22; 7
	D47	*Sc* D54, *Sp*	NMR shift upon ATP addition; reduced decapping *in vivo* & *in vitro*	7, 10
NRD-Nudix hinge	**R95**	*Hs* G96, *Sc* S102, *Sp*	blocked conformational change; reduced decapping *in vivo* & *in vitro*	6, 22
	I96	*Hs* V97, *Sp*	blocked conformational change; reduced decapping *in vivo* & *in vitro*	6, 22
substrate binding	**Y92**	*Sc* Y99	reduced decapping *in vivo*	7
	K93	*Sc* K100, *Sp*	reduced decapping *in vivo* & *in vitro*; reduced growth	7, 8
	K127	*Sc* R133	reduced decapping *in vitro*	8
	Q169	*Sc* K175, *Sp*	reduced decapping *in vitro*	6, 8
	Y220	*Sc* Y222, *Sp*	reduced decapping *in vitro*; reduced growth	6, 8
	K230	*Sp*	reduced decapping *in vitro*; reduced RNA binding	6
	K231	*Sp*	reduced decapping *in vitro*; reduced RNA binding	6
	K234	*Sp*	reduced decapping *in vitro*; reduced RNA binding	6
	K235	*Sp*	reduced decapping *in vitro*; reduced RNA binding	6
Edc1 DAM				
DBR	**I156-G160**	*Sc* H151-G155	reduced decapping *in vitro*	17
	I156-H165	*Sc* H151-T160	reduced decapping *in vitro*; reduced Dcp1:Dcp2 binding	16, 17
PRS	**L172**	*Sc* L169	reduced Dcp1 binding	17
	P173	*Hs* P108, *Sc* P170	reduced Dcp1 binding; reduced decapping *in vivo* & *in vitro*	17, 18
	P175	*Sc* P172	reduced Dcp1 binding; reduced decapping *in vitro*	17
	F177	*Hs* W114, *Sc* F174	reduced Dcp1 binding; reduced decapping *in vivo* & *in vitro*	17, 18
	S166-L178	*Sc* G162-L175	reduced decapping *in vitro*	16

Abbreviations: *Dm - Drosophila melanogaster*, *Hs - Homo sapiens*, *Sc - Saccharomyces cerevisiae*, *Sp - Schizosaccharomyces pombe*

Supplementary Table 3. SAXS data collection and processing statistics.

	Sp Dcp2:Dcp1	*Sp* Dcp2:Dcp1:Edc1
q-range (Å^{-1})	0.015–0.224	0.015–0.224
Guinier R_g (Å)	26.4	25.2
Real-space R_g (Å)	25.9	24.1
$I(0)$	106	89
D_{max} (Å)	83	78
Porod volume (Å^3)	81,800	75,300
Estimated molecular mass (Da)*	48,200	44,294
Theoretical molecular mass (Da)	43,445	46,185

* From Porod volume (Vp/1.7).



Supplementary Note. Sequence analysis of decapping factors. a, Sequence alignment of the EVH1 domain of Dcp1 proteins. Fully conserved residues are highlighted in dark green, and those with >70% similarity are shown in light green. Residues involved in Dcp2 and Edc1

binding are indicated by blue and orange triangles, respectively. Residues that are implicated in RNA binding (corresponding to residues shown as sticks in Fig. 2c) are marked with green asterisks. Residues mutated in this study are highlighted in red in the sequence of *Sp* Dcp1. Species abbreviations are as follows: *Schizosaccharomyces pombe* (*Sp*), *Saccharomyces cerevisiae* (*Sc*), *Arabidopsis thaliana* (*At*), *Caenorhabditis elegans* (*Ce*), *Ancylostoma ceylanicum* (*Ac*), *Nematostella vectensis* (*Nv*), *Anopheles gambiae* (*Ag*), *Danio rerio* (*Dr*), *Xenopus tropicalis* (*Xt*), *Gallus gallus* (*Gg*), *Homo sapiens* (*Hs*). **b,** Sequence alignment of the NRD and Nudix domains of Dcp2. Fully conserved residues are highlighted in dark blue, while those with >70% similarity are shown in light blue. Residues involved in Dcp1 and Edc1 binding are indicated by green and orange triangles, respectively, with filled triangles signifying side chain interactions and empty triangles indicating main chain binding. Residues that were implicated in substrate binding in the literature are marked by red asterisks, potential RNA binding lysine residues in the flexible loop are marked with blue asterisks (corresponding to residues shown as sticks in Fig. 2c). Catalytic site residues are framed in red. Species abbreviations are as in (b). Residues mutated in this study are highlighted in red in the sequence of *Sp* Dcp2

Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation

Tobias Raisch[1,2,3], Chung-Te Chang[1,3], Yevgen Levdansky[1,3], Sowndarya Muthukumar[1], Stefan Raunser[2], and Eugene Valkov[1,*]

[1]Department of Biochemistry, Max Planck Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany.

[2]Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany.

[3]These authors contributed equally.

*Corresponding author: eugene.valkov@tuebingen.mpg.de

Running title: Biochemical characterization of human CCR4-NOT

Keywords: Deadenylation; poly(A) tail; mRNA decay; gene expression

ABSTRACT

CCR4-NOT is a conserved multiprotein complex which regulates eukaryotic gene expression principally via shortening of poly(A) tails of messenger RNA or deadenylation. Here, we reconstitute a complete, recombinant human CCR4-NOT complex. Our reconstitution strategy permits strict compositional control to test mechanistic hypotheses with purified component variants. CCR4-NOT is more active and selective for poly(A) than the isolated exonucleases, CCR4a and CAF1, which have distinct deadenylation profiles in vitro. The exonucleases require at least two out of three conserved non-enzymatic modules (CAF40, NOT10:NOT11 or NOT) for full activity in CCR4-NOT. CAF40 and the NOT10:NOT11 module both bind RNA directly and stimulate deadenylation in a partially redundant manner. Linear motifs from different RNA-binding factors that recruit CCR4-NOT to specific mRNAs via protein-protein interactions with CAF40 can inhibit bulk deadenylation. We reveal an additional layer of regulatory complexity to the human deadenylation machinery, which may prime it either for general or target-specific degradation.

INTRODUCTION

The poly(A) tails at 3′ ends of eukaryotic mRNAs are crucial for their cytoplasmic stability and to enhance the initiation of translation. Newly synthesized metazoan mRNAs possess long poly(A) tails[1], and following export to the cytoplasm the tails are reported to be ~60-80 nucleotides on average at steady state[2]. Poly(A) tails are also important for translational efficiency at the embryonic stage[2] and the length of the poly(A) tail was reported to be correlated with translational efficiency[3]. The multisubunit CCR4-NOT complex is principally responsible for efficient processive shortening of poly(A) tails, or deadenylation, in addition to other functions[4–7]. In addition to its role in bulk mRNA decay, CCR4-NOT can also catalyze the deadenylation or promote translational repression of specific mRNA targets to which it is recruited by RNA binding proteins, such as Nanos, Roquin and Puf/Pumilio proteins[8–13]. In animals, CCR4-NOT functions in cytoplasmic microRNA-mediated gene silencing via a direct interaction with the GW182/TNRC6 proteins[8,14,15], as well as in the nuclear miRNA-mediated gene silencing, which is critical for stem cell differentiation[16].

The CCR4-NOT complex consists of two exonucleases, CCR4 and CAF1, as well as the non-enzymatic proteins NOT1, NOT2, NOT3, and CAF40 which are conserved in all eukaryotes (Fig. 1a). NOT1 is an essential subunit of CCR4-NOT and functions as a scaffold on which other subunits and modules dock[17–19]. The highly conserved CAF40 and NOT2/3 subunits act as protein-protein interaction platforms for sequence-specific RNA-binding proteins[10,11,13,20]. The exonucleases CCR4a (NOT6) and CCR4b (NOT6L) are vertebrate orthologs of the yeast Ccr4 whilst NOT7 and NOT8, which belong to a large family called Caf1, are vertebrate orthologs of the yeast Pop2, respectively[21–23]. Ccr4 is the key functional deadenylase in yeast[24] whereas CAF1 was shown to be crucial for deadenylation in nematodes[25] and in *Drosophila* S2 cells[26]. In human cells, CAF1 is important for microRNA- or small interfering RNA-mediated deadenylation[27]. Recently it was shown that CCR4 is the dominant deadenylase of the human and yeast CCR4-NOT complexes on tails coated with poly(A) binding protein PABPC1 whereas CAF1 is blocked[7,28]. Other species-specific compositional differences relate to the NOT4 subunit, which functions as an E2-dependent RING E3 ligase. It is stably incorporated within the yeast Ccr4-Not but not in *Drosophila* S2 and human cells[18,21–23,29,30]. The largely uncharacterized subunits NOT10 and NOT11 are widely conserved in eukaryotes, except for yeast, and interact stably with the NOT1 subunit (Fig. 1a)[19,23,31].

One long-standing question regarding the CCR4-NOT complex has been that of the role of the non-enzymatic subunits of CCR4-NOT in supporting deadenylation and/or substrate selectivity. Previous results showed that those subunits support target-specific

deadenylation by providing binding sites for factors which recruit CCR4-NOT to specific transcripts[10,11,13,20,32–34], but the question whether other, more direct mechanisms of regulating the nucleases might exist, remained unanswered. Recombinant fission yeast Ccr4-Not complex is strikingly more active and sequence-selective than the Ccr4/Caf1 exonucleases alone[35]. However, the species-specific differences in the subunit composition mean that these results are not directly transferable to the mammalian complex. Hence, in order to investigate the molecular mechanisms of mammalian CCR4-NOT, access to reproducible, active and compositionally-defined preparation is essential. However, despite intense efforts, the isolation of fully assembled and compositionally homogeneous mammalian CCR4-NOT complex has not yet been reported. This limits our mechanistic understanding to studies with isolated subunits and subcomplexes[7,32,33,36–41].

Here, we describe a procedure to reconstitute the human CCR4-NOT complex using a stepwise assembly of purified recombinant components. Our modular assembly approach permits a simple exchange of variants such as truncated constructs and mutants to rapidly exert full compositional control to test mechanistic hypotheses. We observe that the intact human CCR4-NOT has substantially increased deadenylation activity and sequence selectivity in vitro compared to the exonucleases alone. However, this is not strongly influenced by the sequence composition of the segment preceding the poly(A) tail. Several non-enzymatic modules directly stimulate deadenylation by supporting RNA binding in an apparently redundant manner. Binding of RNA-binding proteins to the CAF40 subunit, in turn, inhibits this intrinsic stimulation of deadenylation. Biochemical reconstitution of human CCR4-NOT presents a strikingly holistic view of this complex in which multiple subunits act in concert to regulate deadenylation.

RESULTS

Production of assembly intermediates of human CCR4-NOT

Functional native CCR4-NOT has been isolated from yeast, *Drosophila* and human cells[18,22,23,35]. However, isolation of native complexes is challenging, time-consuming and results in compositionally heterogeneous preparations, which are not manipulable or easily tractable for biochemical study. To solve this problem, we focused our approach on in vitro reconstitution with highly purified, recombinant human proteins. First, we generated a single recombinant baculovirus with genes encoding all eight subunits (CCR4a, CAF1, NOT1, NOT2, NOT3, CAF40, NOT10, and NOT11; Fig. 1a; Supplementary Table 1) using the MultiBac system[42,43]. However, production of the entire complex in insect cells was hampered by low yields and poor subunit stoichiometry. We then revised our approach to co-produce subcomplexes with subsequent reconstitution. We generated a baculovirus comprising the full-length NOT1 together with a minimal set of three full-length proteins (NOT2, NOT3, and CAF40; Supplementary Fig. 1a) rationalized by observations that these four subunits are mutually stabilizing[33,34,39,44]. A decahistidine tag fused to the N-terminus of the NOT1 scaffold subunit was used for metal affinity purification with typical yields of 4-5 mg per liter of insect cell culture (Supplementary Fig. 1b). Remarkably, we observed almost no proteolytic degradation of NOT1, NOT2, and NOT3 despite their size and presence of low-complexity regions (Fig. 1b, lane 3).

The heterodimers of NOT10:NOT11 and the CCR4a:CAF1 exonucleases were recombinantly produced in bacteria (Supplementary Fig. 1c-f). A heterodimer of CCR4a and CAF1 full-length exonucleases was purified by metal affinity capture followed by size exclusion chromatography and a final high-resolution anion exchange step to obtain a complex of 1:1 stoichiometry (Supplementary Fig. 1c,d). Previous work indicated that the C-terminal portion of NOT11 (residues 257-498) was sufficient to form a stable trimeric complex with an N-terminal region of NOT1 as well as NOT10[19,20]. This NOT11 construct including a hexahistidine tag was co-produced in bacteria with NOT10 (residues 25-707) fused to an N-terminal maltose-binding protein (MBP) tag for stability (Supplementary Fig. 1e). Metal affinity capture was followed by proteolytic removal of MBP and subsequent size exclusion chromatography (Supplementary Fig. 1f). The purity of the heterodimeric subcomplexes was confirmed by SDS-PAGE (Fig. 1b, lanes 1 and 2).

Modular reconstitution of the CCR4-NOT complex

We observed that an intact CCR4-NOT comprising eight subunits, which we termed CCR4-NOT_{FULL} (Fig. 1a,c), could be simply and efficiently assembled from three purified

subcomplexes by incubating the NOT1:NOT2:NOT3:CAF40 subcomplex produced in insect cells with a two-fold molar excess of CCR4a:CAF1 and NOT10:NOT11 heterodimers on ice for two hours followed by a separation of the mixture of complexes by size exclusion chromatography (Supplementary Fig. 2a,b). SDS-PAGE analysis with Coomassie staining confirmed that the reconstituted CCR4-NOT$_{FULL}$ comprised all eight protein subunits in an equimolar stoichiometric ratio, as expected (Fig. 1b, lane 5). This was further corroborated by multiangle light scattering coupled with a size exclusion column (SEC-MALS), which indicated that the estimated molecular mass of the peak corresponding to the eight-subunit complex (659 +/- 4 kDa [+/- value indicates technical measurement uncertainty]) was in close agreement with the calculated mass of 647 kDa and the complex was monodisperse in solution (Fig. 1d). Thus, we established a simple and reproducible modular strategy to reconstitute biochemically tractable preparations of human CCR4-NOT with a strictly defined subunit composition.

CCR4-NOT$_{FULL}$ is more active and specific than CCR4a:CAF1

To assess the deadenylation activity of the reconstituted human CCR4-NOT complex we utilized a synthetic RNA substrate of seven nucleotides (5′-UCUAAAU-3′) followed by a polyadenosine tail of 20 nucleotides (A_{20}) and labeled with fluorescein at the 5′ end for visualization (Fig. 2a)[37]. The deadenylation reactions then were analyzed on denaturing gels with single nucleotide resolution of products[45].

The CCR4a:CAF1 exonuclease heterodimer demonstrated very low deadenylation activity at an equimolar ratio of enzyme heterodimer to the RNA substrate in a time course experiment, consistent with previous studies[37,46] (Fig. 2b). Efficient deadenylation was observed after the ratio of the enzyme complex was increased to be in excess over the substrate (Supplementary Fig. 2c), up to maximal five-fold (Fig. 2c). The substrate was degraded first to an intermediate product containing, in addition to the 7-mer body, just one or two As (Fig. 2c, 16-24 min) and degradation terminated at the 5′-terminal 5′-UCU-3′ trinucleotide (Fig. 2c, 32-48 min). This is consistent with previous observations that the CCR4a:CAF1 exonuclease module has a preference rather than strict specificity for adenosine[36,46].

In contrast, CCR4-NOT$_{FULL}$ efficiently degraded the poly(A) tail at an equimolar ratio of complex to the substrate (Fig. 2d vs 2b). CCR4-NOT$_{FULL}$ was not only more efficient than the CCR4a:CAF1 heterodimer but also appeared to be more selective for poly(A), as evident from the stability of the A_1 and A_0 products even following extended incubation (Fig. 2c, 24-48 min vs Fig. 2d, 16-48 min). This suggests that as in yeast the human CCR4-NOT complex exhibits much greater selectivity for poly(A) compared to its constituent exonucleases.

Two non-enzymatic modules suffice to stimulate CCR4a:CAF1

CCR4-NOT$_{FULL}$ is composed of four distinct structural modules: the NOT10:NOT11, CCR4a:CAF1 exonuclease, CAF40 and NOT modules (Fig. 1a,c). We asked if we could delineate which modules or even individual subunits may contribute directly towards the increased activity and selectivity of deadenylation.

To this end, in addition to CCR4-NOT$_{FULL,}$ we generated a six-subunit complex variant, termed CCR4-NOT$_{\Delta 10:11}$, in which the NOT10:NOT11 heterodimer was omitted during the reconstitution procedure (Fig. 1b,c). The speed and selectivity of deadenylation by CCR4-NOT$_{\Delta 10:11}$ appeared very similar to CCR4-NOT$_{FULL}$ (Fig. 2e vs 2d). This suggests that under the conditions of the assay and with the same substrate the NOT10:NOT11 heterodimer does not stimulate deadenylation of the CCR4-NOT in presence of the other modules.

Next, we tested another six-subunit complex comprising, in addition to the CCR4a:CAF1 full-length exonucleases, only the C-terminal half of NOT1 (residues 1093-2376), CAF40 (residues 19-285), and the minimal constructs of NOT2 (residues 344-540) and NOT3 (residues 607-753). In these constructs, the extensive N-terminal low-complexity regions of NOT2 and NOT3, as well as the NOT3 coiled-coil region, were removed (i.e. the constructs comprise the functional and structural core of the NOT module; Fig. 1c and Supplementary Fig. 2d). Unlike the larger CCR4-NOT assemblies containing the full-length NOT1 scaffold, all subunits used to reconstitute this complex were produced in bacteria rather than insect cells. Surprisingly, this minimal complex still retained the deadenylation activity almost at the same level as CCR4-NOT$_{FULL}$ (Fig. 2f vs 2d) and we termed it CCR4-NOT$_{MINI}$ (Fig. 1c).

Finally, we reconstituted an assembly in which the CAF40 module comprising CAF40 and one domain from the NOT1 scaffold was the only other module present in addition to the exonuclease module comprising CCR4a, CAF1 and a MIF4G domain of NOT1 (the NOT1 construct consists of residues 1093-1607; Fig. 1c and Supplementary Fig. 2d). This four-subunit complex lacking the NOT module, which we termed CCR4-NOT$_{CORE}$, was strikingly less active than CCR4-NOT$_{MINI}$ (Fig. 2g vs 2f), but CCR4-NOT$_{CORE}$ was still considerably more active than the CCR4a:CAF1 exonuclease heterodimer (Fig. 2g vs 2b) consistent with previous observations[47].

In summary of the results presented so far, the systematic compositional dissection of human CCR4-NOT revealed that in addition to the CCR4a:CAF1 exonucleases, the non-enzymatic CAF40 and NOT modules are together necessary and sufficient to fully stimulate

deadenylation in vitro. In contrast, the NOT10:NOT11 module did not provide substantial additional stimulation. Importantly, in all the CCR4-NOT complexes we compared, the CCR4a:CAF1 exonucleases were purified and incorporated into larger assemblies in an identical manner. This strategy ensured that activity assays were consistently reproducible between different, independently purified batches of complexes.

At 1 µM concentration the NOT10:NOT11 heterodimer is the only subcomplex to bind the 7-mer-A_{20} RNA under the non-equilibrium conditions of electrophoretic mobility shift assays (EMSA) (Fig. 2h). The binding of the CAF40 module was detectable only at 25 µM (Supplementary Fig. 2e). Although we did not observe binding of other modules under these assay conditions, we cannot rule out that these interactions occur in the context of the intact complex. We then performed UV crosslinking experiments that provide further evidence of interactions of the labeled poly(U) RNA with the CCR4-NOT_{FULL}, the NOT10:NOT11 heterodimer and, to a lesser degree, also with MBP-tagged CAF40 and the nuclease module (Supplemental Fig. 2f). This indicates that the NOT10:NOT11 and CAF40 modules can directly interact with the RNA, which implies that they may contribute to improved substrate binding. Finally, we show that CCR4-NOT_{FULL} binds preferentially to poly(U) and poly(G) sequences compared to poly(A) and poly(C) under the EMSA conditions (Supplementary Fig. 2g).

Distinct activities of CCR4a and CAF1 in CCR4-NOT_{FULL}

The presence of two seemingly redundant nucleases in the CCR4-NOT complex raises the interesting question as to whether they are indeed redundant or have rather distinct functions. In *S. pombe*, single deactivating mutations of either Ccr4 or Caf1 exonuclease only mildly impaired the activity of the intact Ccr4-Not in vitro and complete deadenylation block was observed only when both catalytic mutants were combined[35]. However, the human CCR4a:CAF1 heterodimer stabilized by the accessory factor BTG2 was reported to be inactive in vitro when either of the exonucleases is mutated[48].

To examine the individual contributions of CCR4a and CAF1 we reconstituted heterodimers where one of the enzymes contained an inactivating catalytic mutation: CCR4a^{E240A}:CAF1 and CCR4a:CAF1^{D40A} [35,37,49]. We observed that, in contrast to the reported findings in the presence of BTG2[48], the catalytic inactivation of either CAF1 or CCR4a had little impact on the deadenylation pattern compared to the wildtype CCR4a:CAF1 (Fig. 3a,b vs 2c). CCR4a^{E240A}:CAF1 was slightly less efficient than CCR4a:CAF1^{D40A} (Fig. 3b vs 3a, 8-24 min). However, there was a clear difference in sequence selectivity in the mutant context with CCR4a:CAF1^{D40A} degrading the substrate to a stable A_1 product (5′-UCUAAAU<u>A</u>-3′) whereas CCR4a^{E240A}:CAF1 efficiently degraded to

the 5′-UCU-3′ trinucleotide (Fig. 3a, 32-48 min vs 3b, 40-48 min). Thus, CCR4 appears to be more specific for adenosine nucleotides compared to CAF1[37].

We then asked if CCR4a retains its higher activity and selectivity compared to CAF1 when incorporated into CCR4-NOT$_{FULL}$. Intriguingly, CCR4a:CAF1^{D40A} showed slightly reduced activity in the context of CCR4-NOT$_{FULL}$ compared with wildtype (Fig. 3c, 8-24 min vs 2d, 8-24 min) but higher sequence selectivity than wildtype CCR4-NOT$_{FULL}$ as the A_1 product was remarkably stable over the deadenylation time course (Fig. 2d, 24-48 min vs 3c, 32-48 min). Inactivating CCR4a in the CCR4-NOT$_{FULL}$, however, drastically reduced the deadenylation efficiency compared to wildtype (Fig. 3d vs 2d) or CCR4-NOT$_{FULL}$ reconstituted with CCR4a:CAF1^{D40A} (Fig. 3d vs 3c). Importantly, catalytic inactivation of CCR4a resulted in a much more severe deadenylation defect in the context of CCR4-NOT$_{FULL}$ than in the isolated exonuclease heterodimer (Fig. 3d vs 3b) suggesting that the non-enzymatic subunits modulate distinct exonuclease activities.

We have also reconstituted the CCR4-NOT$_{\Delta 10:11}$ and CCR4-NOT$_{MINI}$ complexes with the CCR4a^{E240A}:CAF1 and CCR4a:CAF1^{D40A} exonuclease mutants. The same overall pattern where CCR4a was more active and selective than CAF1 on synthetic substrates was also observed in these complexes (Supplementary Fig. 3a-e). Thus, the distinct properties of CCR4a and CAF1 appear not to alter in response to the non-enzymatic subunit composition. We have also tested a CCR4a^{E240A}:CAF1^{D40A} complex where both enzymes were inactivated, and incorporated these inactive exonucleases into CCR4-NOT$_{FULL}$ and CCR4-NOT$_{MINI}$. None of these complexes were active, which suggests our purification procedures were effective in removing any contaminant nuclease activity (Supplementary Fig. 3f-h).

Taken together, these results support a model in which CCR4a rather than CAF1 makes the dominant contribution towards the activity and specificity of deadenylation in the intact human CCR4-NOT complex in vitro and in absence of other factors.

Peptide motifs compete with RNA for binding to CAF40

The CAF40 subunit is positioned next to CCR4a:CAF1 on the NOT1 scaffold (Fig. 1a) where it serves as a binding platform not only for nucleic acids in vitro (Supplementary Fig. 2e-f)[41] but also for the α-helical CAF40-binding motifs (CBMs) of metazoan NOT4[50] as well as the *Drosophila melanogaster* (*Dm*) proteins Bag-of-marbles (Bam; Supplementary Fig. 4a)[20] and Roquin[13]. To confirm that the CAF40 subunit retains the same binding to the CBMs in the context of the CCR4-NOT$_{FULL}$ and CCR4-NOT$_{MINI}$ complexes, we carried out pulldown assays with MBP-tagged Bam CBM and the Roquin C-terminal region (Supplementary Fig. 4b,c).

Protein-protein and protein-RNA interactions on CAF40 are shared via a common surface[13,20,41,50], and consequently, they are likely to be mutually exclusive. To test this, we initially set out to test competition between RNA and protein binding by CAF40. We selected the CBM of Bam as a candidate deadenylation antagonist due to its nanomolar affinity for CAF40 and structurally characterized binding mode[20] (Supplementary Fig. 4d). Indeed, MBP-tagged Bam CBM^{WT} efficiently prevented CAF40 from RNA binding in an EMSA (Fig. 4a). In contrast, a Bam CBM containing amino acid substitutions L17E and M24E designed to disrupt binding to CAF40 (Supplementary Fig. 4d)[20], which we termed Bam CBM^{MUT}, had almost no effect on RNA binding by CAF40 (Fig. 4a).

Then, we devised an assay to investigate the impact of Bam CBM^{WT} on the CAF40-mediated stimulation of deadenylation. We chose to initially test it on CCR4-NOT$_{CORE}$ as this complex is more active than the exonucleases and the CAF40 module is the only non-enzymatic module present.

The deadenylation activity of CCR4-NOT$_{CORE}$ was strongly reduced when we titrated increasing amounts of MBP-tagged Bam CBM^{WT} (Fig. 4b, lanes 5-8) compared to the MBP control (Fig. 4b, lanes 1-4). Strong inhibition was also evident in a time course assay in the presence of an almost saturating concentration of the Bam CBM^{WT}, equivalent to 50-fold molar excess over CAF40 (Fig. 4c, left and middle panels). Indeed, the activity of CCR4-NOT$_{CORE}$ with this concentration of Bam CBM^{WT} was reduced almost to the level of the CCR4a:CAF1 heterodimer alone (Fig. 4c, middle panel vs Fig. 2b). The activity of the exonuclease heterodimer alone was not inhibited by Bam CBM^{WT} (Supplementary Fig. 4e). Importantly, the activity of CCR4-NOT$_{CORE}$ was not affected when we added an equivalent molar excess of Bam CBM^{MUT} or with the MBP control (Fig. 4c, left and right panels). This suggests that deadenylation inhibition is directly correlated to high-affinity binding of this linear peptide motif to CAF40.

The NOT10:NOT11 module compensates for CAF40 unavailability

To assess if the Bam CBM can inhibit if the other stimulatory sites of the CCR4-NOT complex are available, we tested its effect on the deadenylation ability of our larger assemblies. Initially, we observed that CCR4-NOT$_{MINI}$ was also progressively inhibited by increasing amounts of Bam CBM (Fig. 4d), but not as efficiently as CCR4-NOT$_{CORE}$ (Fig. 4b), suggesting that the NOT module can partially compensate for the loss of deadenylation stimulation by CAF40. CCR4-NOT$_{FULL}$ was even more resistant to inhibition by Bam CBM^{WT}. Unlike with CCR4-NOT$_{MINI}$, the activity of CCR4-NOT$_{FULL}$ was only slightly affected during titration with Bam CBM^{WT} (Fig. 4e). This difference between the two complexes suggests that the NOT10:NOT11 module is primarily responsible for the resistance of CCR4-NOT$_{FULL}$

against the inhibition by the Bam CBM. This was confirmed by the observation that in presence of Bam CBM^{WT}, the deadenylation activity of CCR4-$NOT_{\Delta 10:11}$ was significantly inhibited compared to the CCR4-NOT_{FULL} and rather resembled the situation in CCR4-NOT_{MINI} (Fig. 4f).

Given that CAF40 is conserved and in an assumed spatial proximity to the exonucleases, we suggest that CAF40 provides the principal stimulatory surface. If this is blocked by a linear motif from an interacting partner, then the NOT and NOT10:NOT11 modules can potentially compensate.

Multiple CCR4-NOT interactors can inhibit deadenylation

To determine if the deadenylation inhibition via the CBM is unique to Bam or whether CBMs from other RNA-binding recruitment factors can also impact on CCR4-NOT activity, we tested an extended C-terminal fragment of human Roquin1 (Roq-C), which binds directly to the CAF40 subunit and the NOT module (Supplementary Fig. 4f)[13,20], as well as a CBM from the C-terminus of NOT4 which binds to the same surface as the Bam CBM, but in a structurally unrelated manner (Supplementary Fig. 4g)[50]. In a series of deadenylation time courses, where we systematically tested the effect of all three MBP-tagged peptides (Bam CBM^{WT}, Roq-C, and NOT4 CBM) on CCR4-NOT_{MINI} (Fig. 5a), CCR4-NOT_{FULL} (Fig. 5b) and CCR4-$NOT_{\Delta 10:11}$ (Fig. 5c), we observed that peptide motifs from all three factors inhibit CCR4-NOT-mediated deadenylation in vitro. However, the extent of inhibition varied among the complexes. The complexes lacking the NOT10:NOT11 heterodimer, CCR4-$NOT_{\Delta 10:11}$ and CCR4-NOT_{MINI}, were strikingly more sensitive to the inhibition by the peptides compared to the CCR4-NOT_{FULL}. Taken together, these results show that multiple interaction partners of the CCR4-NOT complex can repress deadenylation via direct binding to CAF40 and the NOT10:NOT11 module compensates by restoring deadenylation stimulation.

3′-UTR length but not composition affects deadenylation

The interactions of several stimulatory modules with RNA suggested that longer substrates may engage several stimulatory surfaces at the same time. This would stabilize the enzyme-substrate complex and lead to more efficient deadenylation. To test this, we performed deadenylation assays using a substrate with an extra 13 randomly chosen nucleotides, which are predicted to be unstructured, 5′ of the 7-mer-A_{20} (Supplementary Table 4). This RNA substrate interacts stably with the CCR4-NOT_{FULL} complex as well as the NOT10:NOT11 heterodimer under the EMSA conditions (Supplementary Fig. 5a), and at higher concentrations with the CAF40 module (Supplementary Fig. 5b), similar to the shorter 7-mer-A_{20} substrate (Fig. 2h and Supplementary Fig. 2e). We observed an enhanced

deadenylation activity of CCR4-NOT$_{FULL}$ on this 13+7-mer-A$_{20}$ substrate compared to 7-mer-A$_{20}$ (Fig. 6a, 16 min vs 2d, 48 min, and Fig. 6d). This result is in contrast to the reported situation in fission yeast where the length upstream of the poly(A) did not significantly influence activity[35]. At the same time, the CCR4-NOT complex did not degrade this substrate beyond the first two consecutive non-A nucleotides, consistent with the notion that the high level of selectivity for adenosine is not influenced significantly by the overall length of the substrate. Strikingly, we did not observe a similar enhancement with the longer substrate on the activity of the CCR4a:CAF1 heterodimer (Fig. 6b vs 2c). Furthermore, the absence of the NOT10:NOT11 module reduced the deadenylation activity on the longer substrate only slightly (Fig. 6c vs 6a, and Supplementary Fig. 5c).

To assess whether the sequence of the region preceding the poly(A) influences deadenylation by CCR4-NOT$_{FULL}$, we performed deadenylation assays using four different substrates, each with 20 nucleotides upstream of A$_{20}$ (Supplementary Table 4). Surprisingly, the differences between three substrates with unstructured 5′ regions were not very pronounced (Fig. 6d,e, Supplementary Fig. 5d), suggesting that the sequence composition of 3′-UTR does not significantly influence deadenylation. Introduction of a short stem-loop structure led to an initial decrease in the rate of deadenylation but the reaction still proceeded to completion (Fig. 6f vs 6d,e). This suggests that the presence of secondary structure in the 3′-UTR is not a critical determinant of deadenylation efficiency.

The CCR4-NOT complex is a conformationally flexible assembly

We used negative stain electron microscopy (EM) to characterize the particle shape and dimensions of CCR4-NOT$_{FULL}$. This revealed many particles of different size and shape (Supplementary Fig. 6a). Their size heterogeneity and globular appearance suggested they were isolated subunits or subcomplexes. Localized particle clustering further indicated that CCR4-NOT$_{FULL}$ disassembles when stained with acidic uranyl formate. To stabilize complexes for EM, we employed mild crosslinking using glutaraldehyde during sucrose density gradient centrifugation[51]. This markedly reduced nonspecific aggregation and complex disassembly resulting in a more homogeneous particle size distribution (Supplementary Fig. 6b). The average length of a particle was approximately 22-24 nm in the longest direction, consistent with the reported dimensions observed by negative stain EM for the yeast Ccr4-Not complexes[52,53]. However, particle shape and appearance indicates that CCR4-NOT$_{FULL}$ adopts not one single conformation but rather a continuum of many different conformations, which means that many of the subunits and modules may be able to locate in close spatial proximity.

373

DISCUSSION

The human CCR4-NOT complex was recently shown to be the principal mediator of deadenylation[7]. Here, we describe the production of a fully recombinant human CCR4-NOT complex comprising all eight subunits, with only the NOT11 subunit containing a significant truncation of the unstructured N-terminal region. Through a modular reconstitution approach, we systematically characterized the contribution of non-enzymatic modules towards deadenylation.

We observed that the intact complex is strikingly more active compared to the isolated CCR4a:CAF1 heterodimer, consistent with observations in fission yeast[35]. The CCR4-NOT complex is also much more sequence-specific as evident by stalling following efficient shortening of the substrate poly(A) tail. Previously it was shown that recombinant CCR4b is significantly more selective towards polyadenosine in synthetic substrates compared to CAF1[36,37,54]. Our studies with mutant nucleases incorporated into the CCR4-NOT indicated that CCR4a is principally responsible for sequence selectivity and activity whereas CAF1 was rather inactive in three distinct compositional variants of the CCR4-NOT under the assay conditions. This contrasts markedly with the fission yeast complex where Caf1 and Ccr4 catalytic mutants had almost identical deadenylation profiles[35]. CCR4a and CAF1 were reported to be functionally distinct depending on whether the poly(A) tail is coated with the poly(A)-binding protein (PABPC1)[7,35]. Our results indicate that the functional distinction between CCR4a and CAF1 exists also in the absence of PABPC1.

The activity of the fission yeast Ccr4-Not was shown to depend on the sequence and the secondary structure content of the 3′-UTR but not its length. It was thus proposed that the Ccr4-Not has an intrinsic ability to recognize and respond to the sequence and/or structure context upstream of the poly(A) tail[35]. Human CCR4-NOT, in contrast, showed a strong dependence on the length of the 3′-UTR segment. Since the exonuclease module does not bind the substrate well, the purpose of multiple independent RNA-binding sites on spatially distinct modules may be to strengthen the interaction of the complex with its substrate via avidity effects and thereby improve deadenylation efficiency. Furthermore, our data indicate that the human CCR4-NOT is not selective at the sequence level for 3′-UTR and deadenylation activity is not significantly impacted by the secondary structure. This is consistent with a view that the CCR4-NOT may be more reliant on extrinsic factors for 3′-UTR discrimination than the yeast complex. This may increase the potential for a very fine, transcript-specific regulation of gene expression necessary in multicellular organisms.

We observed that the CCR4-NOT-mediated deadenylation is only efficiently terminated once a stretch of more than two non-adenosine residues is encountered by the complex. The CCR4-NOT thus has an intrinsic preference for poly(A), which is enhanced by the non-enzymatic subunits and is independent of the sequence context of the 3′-UTR. Collectively, our findings support a view of the human CCR4-NOT where non-enzymatic subunits serve not only to enhance the exonuclease activity but to exquisitely tune it for the purpose of poly(A) shortening and to minimize nonspecific RNA degradation.

We observed that at least three independent sites on the human CCR4-NOT act in concert to stimulate deadenylation. The CAF40 subunit binds RNA directly via a conserved surface and is sufficient to strongly stimulate activity. CAF40 is proximal to the exonuclease module, which does not bind substrates stably, and thus CAF40 may stimulate deadenylation through improved substrate binding or through optimized orientation of the RNA for degradation.

In addition to the CAF40 module, the structural core of the NOT module, consisting of the NOT2 and NOT3 subunits as well as the C-terminal domain of NOT1, is necessary and sufficient to achieve full stimulation in vitro. In our binding assays, the human NOT module does not bind RNA directly unlike the yeast NOT module, which was shown to bind nucleic acid in vitro[44]. The NOT module is critical for the stability of CCR4-NOT in *Drosophila* S2 cells[39] and we suggest this role is conserved in the human complex thus aiding its activity but the precise mechanistic contribution remains unclear at present.

The NOT10:NOT11 module provides the third potent stimulatory site. The module is conserved in many eukaryotes, with the exception of unicellular yeasts, but its function remains unknown although it appears to serve an important function in mammals as *Cnot10* deficiency in mice causes embryonic lethality[55]. It binds RNA directly and with the highest relative affinity of all modules tested in this study (Fig. 2h). Thus, we suggest that NOT10:NOT11 stimulates deadenylation through direct stabilization of the substrate RNA but only under distinct conditions where the principal stimulatory CAF40 surface may not be available for RNA binding.

Taken together, we propose a mechanistic model in which RNA-binding surfaces on distinct modules cooperate to enhance interaction with the substrate to elicit efficient deadenylation. Such interactions may also exploit the inherent structural flexibility of CCR4-NOT to position the mRNA substrate in an optimal conformation (Fig. 7a).

Linear motifs from several mRNA-binding proteins can bind to the highly conserved nucleic acid binding site on CAF40 to compete with RNA binding and interfere with

deadenylation stimulation. The long-standing notion of targeted deadenylation by the CCR4-NOT is a unidirectional one in which CCR4-NOT acts on bulk mRNA and is repurposed toward specific mRNA targets by protein factors. Our biochemical data suggests that blocking an RNA-binding surface on CAF40 by recruitment factors may reduce nonspecific substrate binding and bulk deadenylation but this model remains to be tested in vivo (Fig. 7b). Our findings indicate that there is some RNA-binding redundancy in the CCR4-NOT modules. Recruitment factors such as GW182[33], TTP[56], and Roquin[13] bind to several CCR4-NOT subunits through extended motifs. Multiple recruitment and regulatory events could perhaps occur simultaneously on the multisubunit CCR4-NOT and access to specific stimulatory sites may be regulated in specific contexts. An important future goal is to understand how the conformational dynamics of the CCR4-NOT complex are influenced through combinatorial control from multiple regulatory inputs in response to stimuli to ultimately redirect the gene expression program of specific targets.

In conclusion, our compositional dissection of the human CCR4-NOT reveals that many parts of this multisubunit complex act together and intrinsically modulate each other to coordinate the shortening of poly(A) tails. We propose a mechanistic model of allosteric and cooperative stimulation of the exonucleases where several non-enzymatic modules mediate interactions with the RNA substrate. Means of producing the recombinant and biochemically tractable intact human CCR4-NOT will facilitate the future studies of targeted deadenylation, translational repression, and co-translational decay.

METHODS

DNA constructs

For expression of the NOT1:NOT2:NOT3:CAF40 complex (Supplementary Fig. 1a and 7a), full-length NOT1 was amplified from a human cDNA library, and sequences encoding an N-terminal His_{10} (fused to NOT1 via a TEV protease site) and two consecutive C-terminal StrepII tags were added during PCR. The DNA sequence encoding full-length CAF40 was amplified from the same cDNA library. Then, the NOT1 and CAF40 PCR products were fused together, each with a downstream SV40 poly(A) signal and an upstream polyhedrin promoter, into a single cassette by overlap extension PCR, and this cassette was inserted between the SalI and ScaI restriction sites into the pACEBac1 vector backbone (Supplementary Fig. 7b). The oligonucleotide primers used in this study are listed in Supplementary Table 5.

cDNA encoding full-length NOT2 was also amplified from the cDNA library, and the cleaved PCR product inserted between the XhoI and NcoI restriction sites of the pIDK plasmid. Amplified cDNA encoding full-length NOT3 was inserted in the KpnI restriction site of the pIDK plasmid. Following this, the NOT2 cassette including the upstream polyhedrin promoter and the downstream SV40 poly(A) signal was amplified from the pIDK-NOT2 plasmid and inserted into the ClaI restriction site of pIDK-NOT3 to create the pIDK-NOT2-NOT3 plasmid (Supplementary Fig. 7b).

Finally, the pACEBAC-NOT1-CAF40-NOT2-NOT3 bacmid was created in vitro by Cre-Lox recombination using Cre recombinase (New England Biolabs) with 200 fmol of pACEBac-NOT1-NOT9 and 300 fmol pIDK-NOT2-NOT3 in the reaction.

cDNA encoding full-length CCR4a was inserted into the pMCSG19c plasmid vector[57] between KpnI and BamHI restriction sites (Supplementary Fig. 7c). In this vector an insert is fused in-frame with a gene encoding maltose-binding protein (MBP) with an adjacent TVMV protease cleavage site. Low-level production of encoded TVMV protease from the same plasmid cleaves the fused MBP in vivo resulting in a CCR4a protein construct with an N-terminal His_6-tag followed by the TEV protease recognition site.

CAF1 cDNA fused to an N-terminal TEV site was inserted between the BamHI and HindIII restriction sites of the pET28b plasmid (Merck). To improve the stability of CAF1, a SUMO3 tag with an N-terminal His_6 tag was inserted between NdeI and BamHI restriction sites, resulting in a SUMO-His_6-CAF1 fusion construct cleavable by TEV protease (Supplementary Fig. 7c).

The NOT10:NOT11 heterodimer for reconstitution was co-expressed from two plasmids: cDNA encoding NOT10 (residues 25-707) was inserted between the XhoI and BamHI restriction sites of the pnYC-pM plasmid[58], resulting in an MBP-tagged fusion construct cleavable by HRV-3C protease (Supplementary Fig. 7d). The NOT11 C-terminal construct (residues 257-498) was inserted between the same restriction sites of the pnEA/vH plasmid[58], and the resulting C-terminally His_6-tagged construct was cleavable by TEV protease (Supplementary Fig. 7d).

For expression of the tetramer which serves as a scaffold of the CCR4-NOT_{MINI} complex (Supplementary Fig. 8a), two plasmids were used. MBP-tagged NOT1 (residues 1093-2376) encoded on a plasmid of the pnYC backbone which has been described before[13] (Supplementary Fig. 8b). A tricistronic plasmid encoding for MBP-tagged NOT2 (residues 344-540), His_6-tagged NOT3 (residues 607-753) and His_6-tagged CAF40 (residues 19-285) was constructed based on the pnEA architecture[58]. All tags were cleavable by HRV-3C protease apart from the His_6 tag fused to NOT3 which was non-cleavable (Supplementary Fig. 8b).

Two plasmids were used for the expression and purification of the NOT1:CAF40 complex serving as the base for reconstitution of the CCR4-NOT_{CORE} complex (Supplementary Fig. 8c). cDNA encoding for NOT1 (residues 1093-1607) was inserted between the XhoI and BamHI restriction sites of the pnYC-pM plasmid backbone. CAF40-encoding cDNA (residues 19-285) was inserted between the XhoI and BamHI restriction sites of a pnEA plasmid which does not encode any affinity tags (Supplementary Fig. 8d).

For expression of MBP-tagged CAF40, cDNA encoding CAF40 residues 18-293 was inserted between the XhoI and BamHI restriction sites of the pnEA-pM plasmid[58].

The DNA constructs for the production of the individual CCR4-NOT modules (the CAF40 module comprising NOT1 residues 1351-1588 and CAF40 residues 19-285, and the NOT module core comprised of NOT1 residues 1833-2361, NOT2 residues 350-540 and NOT3 residues 607-748), MBP-tagged *Drosophila* Bam CBM and MBP-tagged *Hs* Roquin1 CBM have been described before, as well as the detailed description of expression and purification of these constructs[11,13,20,33].

Site-directed mutagenesis was carried out according to modified QuickChange™ protocol[59]. All the plasmid constructs and mutants used in this study were confirmed by DNA sequencing and are listed in Supplementary Table 2.

Baculovirus production

Recombined multi-plasmid vectors were transformed into chemically competent DH10EmbacY cells (a kind gift from Imre Berger). Positive baculovirus genome integrands were selected using blue/white screening on LB-agar plates containing IPTG and Bluo-Gal (Gold Biotechnology) with 25 µg/ml kanamycin, 10 µg/ml tetracycline, 34 µg/ml chloramphenicol and 10 µg/ml gentamicin. Two white colonies were used to inoculate 5 ml LB cultures with all four antibiotics at the same concentrations and cultured at 37°C overnight. Bacmid DNA was purified by alkaline lysis with buffers P1, P2, and N3 (Qiagen). Following isopropanol precipitation, pellets were washed with ice-cold 70% (v/v) ethanol and centrifuged at 4°C before being resuspended in sterile, milliQ-grade water.

Sterile 1–5 µg bacmid DNA was transfected into a standard sterile 6-well plate (Greiner) with technical replicates of 1.0 x 10^6 adherent *Sf*21 cells (a kind gift from Imre Berger) using Fugene HD reagent (Promega). The supernatant containing the initial V_0 low-titer baculovirus was collected once at least 50% of cells were fluorescent due to the expression of the YFP marker. Suspension cultures at 1.0 x 10^6 cells/ml were infected with approximately 10% (v/v) V_0 virus stock. A working V_1 stock of baculovirus was collected 72 hours after infection. Cells were cultured in Sf900II serum-free medium (Thermo Fisher Scientific) without any supplements.

Purification of the CCR4a:CAF1 exonuclease heterodimer

Full-length CCR4a and CAF1 were co-expressed in *E. coli* BL21(DE3) Star cells (Thermo Fisher Scientific) in LB medium at 20 °C as fusion proteins carrying N-terminal His_6 and His_6-SUMO tags, respectively. Cells were lysed using an Emulsiflex-C3 homogenizer (Avestin) in a buffer containing 50 mM potassium phosphate pH 7.5, 300 mM NaCl and 25 mM imidazole supplemented with complete EDTA-free protease inhibitors, 5 µg/ml DNaseI and 1 mg/ml lysozyme. The proteins were isolated from the cleared lysate by binding to a Nickel-charged HiTrap IMAC column (GE Healthcare) and eluted from the column by a linear gradient to the same buffer supplemented with 500 mM imidazole. Both tags were then cleaved off overnight by incubation with recombinant TEV protease while dialyzing against a buffer containing 30 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 2 mM DTT. Then, the CCR4a:CAF1 heterodimer was eluted on a Superdex 200 26/600 size exclusion chromatography column (GE Healthcare) equilibrated in a buffer containing 10 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 2 mM DTT. Finally, the NaCl concentration was diluted to 75 mM, and the complex bound to a Source 15Q column (GE Healthcare) was subsequently eluted by a linear gradient to buffer containing 10 mM HEPES/NaOH pH 7.5, 1000 mM NaCl and 2 mM DTT. The eluted CCR4a:CAF1 complex was concentrated to 6-7

mg/ml, flash-frozen and stored at -80 °C. Mutant variants of CCR4a:CAF1 were purified in the same way as the wild-type heterodimer. The CCR4a:CAF1 dimer purified following this scheme was used for reconstitution of all larger complexes (CCR4-NOT_{FULL}, CCR4-$NOT_{\Delta 10:11}$, CCR4-NOT_{MINI} and CCR4-NOT_{CORE}).

Purification of the NOT10:NOT11 heterodimer

Both proteins were co-expressed in *E. coli* BL21(DE3) Star cells in LB medium at 20 °C as fusion proteins carrying C-terminal His_6 (NOT11, residues 257-498) and N-terminal MBP (NOT10, residues 25-707) tags, respectively. Cells were lysed using an Emulsiflex-C3 homogenizer (Avestin) in a buffer containing 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 25 mM imidazole supplemented with complete EDTA-free protease inhibitors, 5 µg/ml DNaseI and 1 mg/ml lysozyme. The proteins were isolated from the cleared lysate by using a nickel-charged HiTrap IMAC column (GE Healthcare) and eluted by a linear gradient to a buffer containing 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 500 mM imidazole. Subsequently, the tags were cleaved off overnight by incubation with recombinant HRV-3C and TEV proteases, while dialyzing against a buffer containing 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10% (v/v) glycerol and 2 mM DTT. Then, the NOT10:NOT11 heterodimer was eluted on a Superdex 200 26/600 size exclusion chromatography column (GE Healthcare) equilibrated in a buffer containing 10 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10% (v/v) glycerol and 2 mM DTT. The NOT10:NOT11 heterodimer was concentrated to 6-7 mg/ml, flash-frozen and stored at -80 °C.

MBP-CAF40 purification

CAF40 (residues 18-293) was expressed in *E. coli* BL21(DE3) Star cells in LB medium at 20 °C with an N-terminal MBP tag. Cells were lysed using an Emulsiflex-C3 homogenizer (Avestin) in a buffer containing 50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, and 2 mM DTT. The complex was isolated from the crude lysate using amylose resin (New England Biolabs) and eluted with lysis buffer supplemented with 30 mM D-(+)-maltose. The eluted protein was diluted to reach a final concentration of 100 mM NaCl. This was then applied on a 5 ml heparin column (GE Healthcare) and eluted with a linear gradient to 1 M NaCl. This was followed by size exclusion chromatography on a Superdex 200 26/60 column (GE Healthcare) equilibrated in a buffer containing 10 mM HEPES/NaOH pH 7.5, 200 mM NaCl, 10% (v/v) glycerol, and 2 mM DTT. The protein was concentrated to approximately 25 mg/ml.

Reconstitution of CCR4-NOT$_{FULL}$ and CCR4-NOT$_{\Delta 10:11}$ complexes

*Sf*21 insect cells were grown at 27 °C to a density of 2.0 x 10^6 cells/ml in Sf900II medium (Thermo Fisher Scientific) and infected with recombinant NOT1:NOT2:NOT3:CAF40 baculovirus stock (1:100 v/v). The cells were harvested by centrifugation 48 h after they stopped dividing and pellets stored at -80 °C.

Pellets of insect cells expressing the NOT1:NOT2:NOT3:CAF40 tetramer were thawed, resuspended in lysis buffer (50 mM HEPES/NaOH pH 7.5, 500 mM NaCl, 10 mM potassium phosphate pH 7.5, 50 mM imidazole, 5 µg/ml DNAseI) and lysed by sonication (Branson Sonifier; level 2; 20% output; 1 min). The lysate was cleared by centrifugation (15 minutes at 12,500 x *g*) and filtered through 0.45 µm and 0.22 µm syringe-driven filters (Millipore). The cleared and filtered lysate was applied to a nickel-charged 1 ml HiTrap IMAC column (GE Healthcare) and the bound complex eluted by a linear gradient to a buffer containing 50 mM HEPES/NaOH pH 7.5, 300 mM NaCl, 10 mM potassium phosphate pH 7.5 and 500 mM imidazole.

For reconstitution of CCR4-NOT$_{FULL}$, the NOT1:NOT2:NOT3:CAF40 subcomplex was mixed in a 1:2:2 molar ratio with CCR4a:CAF1 (wildtype or mutants) and NOT10:NOT11 complexes. The mixture was then incubated on ice for 2 hours to assure the association of the subcomplexes with each other, and then the assembled complex was separated from the excess of CCR4a:CAF1 and NOT10:NOT11 heterodimers by size exclusion chromatography on a Sephacryl HR-300 26/60 column (GE Healthcare) equilibrated in a buffer containing 10 mM HEPES/NaOH pH 7.5, 200 mM NaCl and 2 mM DTT. The complex was concentrated to approximately 1 mg/ml, and for deadenylation assays, it was subsequently aliquoted, flash-frozen and stored at -80 °C.

Procedures were identical for the reconstitution of the CCR4-NOT$_{\Delta 10:11}$ complex with the difference that the NOT10:NOT11 heterodimer was omitted from the mixture of subcomplexes.

Reconstitution of the CCR4-NOT$_{MINI}$ complex

MBP-tagged NOT1 (residues 1093-2371), MBP-tagged NOT2 (residues 344-540), His_6-tagged NOT3 (residues 607-753) and His_6-tagged CAF40 (residues 19-285) were co-expressed in *E. coli* BL21(DE3)Star cells in LB medium at 20 °C. Cells were lysed using an Emulsiflex-C3 homogenizer (Avestin) in a buffer containing 50 mM potassium phosphate pH 7.5 and 300 mM NaCl supplemented with complete EDTA-free protease inhibitors, 5 µg/ml DNaseI and 1 mg/ml lysozyme. The complex was isolated from the crude lysate using

amylose resin (New England Biolabs) and eluted with lysis buffer supplemented with 25 mM D-(+)-maltose. The tags were removed by cleavage with HRV-3C protease.

Subsequently, a two-fold molar excess of purified CCR4a:CAF1 heterodimer (wildtype or mutants) was added and the mixture incubated for two hours. Finally, the assembled complex was separated from the affinity tags and the excess of CCR4a:CAF1 by size exclusion chromatography on a Superdex 200 26/60 column (GE Healthcare) equilibrated in a buffer containing 20 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 2 mM DTT. The complex was concentrated to approximately 1 mg/ml.

Reconstitution of the CCR4-NOT$_{CORE}$ complex

The heterodimer of NOT1 (residues 1093-1607) and CAF40 (residues 19-285) was co-expressed in *E. coli* BL21(DE3) Star cells in LB medium at 20 °C with N-terminal MBP and His_6 tags, respectively. Cells were lysed using an Emulsiflex-C3 homogenizer (Avestin) in a buffer containing 50 mM potassium phosphate pH 7.5 and 300 mM NaCl supplemented with complete EDTA-free protease inhibitors, 5 µg/ml DNaseI and 1 mg/ml lysozyme. The complex was isolated from the crude lysate using amylose resin (New England Biolabs) and eluted with lysis buffer supplemented with 25 mM D-(+)-maltose. The tags were removed by cleavage with HRV-3C protease.

Then, a two-fold molar excess of purified CCR4a:CAF1 heterodimer was added and the mixture incubated for two hours. Finally, the assembled complex was separated from the affinity tags and the excess of CCR4a:CAF1 by size exclusion chromatography on a Superdex 200 26/60 column (GE Healthcare) equilibrated in a buffer containing 20 mM HEPES/NaOH pH 7.5, 300 mM NaCl and 2 mM DTT. The complex was concentrated to approximately 1 mg/ml.

The residue counts and molecular weights of all recombinant complexes and their subunits are listed in Supplementary Table 3.

Size exclusion chromatography with light scattering

The purified CCR4-NOT$_{FULL}$ complex (40 µl at 0.9 mg/ml) was applied to a Superose 6 5/150 column (GE Healthcare) equilibrated in a buffer containing 10 mM HEPES/NaOH pH 7.5 and 200 mM NaCl, which was connected to the miniDAWN TREOS and Optilab rEX instruments (Wyatt Technologies). Samples were analyzed by multiangle static light scattering, and the absolute molecular weight of each protein was calculated from the light scattering data with ASTRA (Wyatt Technologies).

Deadenylation assays

Deadenylation assays were carried out as described previously[35,45] and with minor modifications. Briefly, the deadenylase complex to be tested was mixed with a synthetic RNA substrate (sequences are listed in Supplementary Table 4) carrying a 6-FAM label at the 5′ end (Biomers). Reactions were performed at 37 ˚C in a buffer containing 20 mM PIPES pH 6.8, 10 mM KCl, 40 mM NaCl and 2 mM $Mg(OAc)_2$. The total reaction volume was 60 µl, the concentration of FAM-labeled RNA 50 nM and the concentration of deadenylase complex 50 nM (or 250 nM in the case of the CCR4a:CAF1 dimer). The reactions were stopped at the indicated time points by adding 180 µl of 2x RNA loading dye (95% (v/v) deionized formamide, 17.5 mM EDTA pH 8, 0.01% (w/v) bromophenol blue). In the case of competition assays with CBM peptides, the indicated final concentrations of the respective MBP-tagged peptide (or MBP as control) was incubated with the protein complex for 15 minutes before mixing with the RNA substrate, and the deadenylation reactions were stopped 32 minutes after adding the substrate.

The reaction products were separated on denaturing TBE-urea polyacrylamide gels [20% (w/v) 19:1 acrylamide-bis acrylamide, 7 M urea, 1 x TBE (Tris-Borate-EDTA) buffer] at 300 V for 2 h for the 7-mer-A_{20} substrate, and 3.5 h for the longer substrates, followed by analysis using a Typhoon RGB Biomolecular Imager (GE Healthcare).

All deadenylation assay experiments were performed in triplicates. All assays were also extensively validated with proteins and complexes independently purified and reconstituted in different batches.

Electrophoretic mobility shift assay (EMSA)

Binding reactions contained 100 nM of labeled RNA and 1 µM or 25 µM of the indicated proteins in a total reaction volume of 10 µl of binding buffer (20 mM Tris/HCl 7.5, 10 mM NaCl, 2 mM $MgCl_2$, 0.1% BSA, 0.1% Orange G, 3% Ficoll 400). For the competition assay, the reaction mixture contained 100 µM of the NOT1:CAF40 complex and 200 µM of the MBP-tagged Bam CBM (wildtype or mutant as indicated). The proteins were pre-incubated for 30 min before addition of RNA. The RNA-protein complexes were analyzed by electrophoresis on a 10% nondenaturing polyacrylamide gel in TBE buffer, pH 8.3, at 10 V cm^{-1}.

UV crosslinking

Proteins at 1 µM (final concentration) were mixed with 100 nM (final concentration) of 5′ 6-FAM-labeled poly$(U)_{30}$ RNA (Biomers) in a buffer containing 50 mM Tris/HCl pH 7.5, 50 mM

ammonium sulfate and 5 mM magnesium chloride in a total reaction volume of 20 µl. The samples to be crosslinked were placed on a precooled rack on ice, and irradiated by a 254 nm UV light source (Stratalinker 2400) at a total energy dose of 2400 mJ/cm^2. Subsequently, 2x denaturing protein loading buffer was added and the samples resolved on a 4-12% NuPAGE Bis-Tris gel (Thermo Fisher Scientific). The fluorescence of the 6-FAM label on the RNA was detected using a Typhoon RGB Biomolecular Imager (GE Healthcare), followed by Coomassie staining.

Pulldown assays

Purified MBP or MBP-tagged peptides (30 µg) were incubated with 30 µl amylose resin slurry (New England Biolabs) in binding buffer (50 mM Tris/HCl pH 7.5, 150 mM NaCl) at 4 °C. Following 1 h incubation, the beads were washed three times with binding buffer. Then, 20 µg CCR4-NOT complex was added and the mixture incubated at 4 °C for 1 h. Finally, the beads were washed three times with binding buffer, and the bound proteins eluted using binding buffer supplemented with 25 mM D-(+)-maltose and analyzed by SDS-PAGE.

Electron microscopy

For EM analysis, CCR4-NOT$_{FULL}$ was applied to EM grids coated with a continuous carbon layer at a concentration of 20 µg/ml and negatively stained using 0.75% (w/v) uranyl formate. The grids were analyzed on a Tecnai G Spirit TEM (Thermo Fisher Scientific) equipped with a TVIPs TemCam F416 4k CMOS camera. Mild crosslinking was performed during sucrose gradient centrifugation using the GraFix protocol[51] on a gradient of 10-30% (w/v) sucrose. The gradient was fractionated, and the fractions analyzed for protein content via a dot blot using an UltraCruz nitrocellulose membrane (Santa Cruz Biotechnology) and stained by amido black 10B (Sigma Aldrich) [0.1% (w/v) dissolved in 25% (v/v) isopropanol, 10% (v/v) acetic acid].

DATA AVAILABILITY

Source data for Fig. 1b, 2b-h, 3, 4, 5 and 6 and Supplementary Fig. 1b,d,f, 2b-g, 3, 4b,c,e, 5 and 6 are available with the paper online. All other data will be made available upon request.

ACKNOWLEDGMENTS

We dedicate this work to the memory of Elisa Izaurralde who passed away while this work was at the initial stage and we wish to gratefully acknowledge that the study was conceived and carried out in her laboratory. We remain deeply grateful for her enthusiasm, support, and encouragement. We thank Catrin Weiler and Sigrun Helms for excellent technical assistance, Ying Chen for providing several of the plasmids used for bacterial expression, Felix Sandmeir for initial help in establishing deadenylation assays, and members of the Izaurralde laboratory for many helpful discussions and comments on the manuscript. We are grateful to Imre Berger and Frederic Garzoni for expert advice and training in insect cell protein production supported by funding (to E.V.) from the European Commission's Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement N°283570). This work was supported by the Max Planck Society.

AUTHOR CONTRIBUTIONS

T.R. and E.V. conceived the study. T.R., E.V., C.C., Y.L., and S.M. carried out experiments and, together with S.R., analyzed the data. T.R. and E.V. wrote the manuscript; Y.L., C.C, and S.M. assisted in the manuscript and figure preparation. E.V. supervised the study as the principal investigator. All authors commented on the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Architecture and reconstitution of the human CCR4-NOT complex.

(a) Schematic representation of the human CCR4-NOT complex. Domains are indicated and known interactions between domains are shown by dashed lines. HEAT: domain consisting of α-helical HEAT-like repeats; MIF4G: HEAT-like domain with structural similarity to the middle domain of eIF4G; DUF: domain of unknown function 2363; TPR: predicted domain consisting of α-helical TPR-like repeats; ARM: domain consisting of α-helical armadillo repeats; EEP: endonuclease/exonuclease/phosphatase family member; LRR: Leucine-rich repeats; DEDD: exonuclease of the DEDD family; CC: predicted coiled-coil domain; NB: NOT-box domain.

(b) Coomassie-stained gel with the purified recombinant NOT10:NOT11, CCR4a:CAF1 and NOT1:NOT2:NOT3:CAF40 complexes used for modular reconstitution, and the resulting assembled CCR4-$NOT_{\Delta 10:11}$ and CCR4-NOT_{FULL} complexes. Source data are provided as a Source Data file.

(c) Schematic representation of the reconstituted complexes used in this study. CCR4-$NOT_{\Delta 10:11}$ lacks the NOT10:NOT11 heterodimer. In CCR4-NOT_{MINI} the N-terminal parts of NOT1, NOT2, and NOT3 are truncated. CCR4-NOT_{CORE} comprises just two structured domains of NOT1 with the docked exonucleases CAF1 and CCR4 as well as CAF40.

(d) Size exclusion chromatography (SEC) elution profile of the reconstituted CCR4-NOT_{FULL} complex with multi-angle laser light scattering (MALS) profile shown in red for the protein peak. Theoretical and experimentally-derived mass is indicated.

Figure 2. The contribution of the non-enzymatic modules towards activity.

(a) The sequence of the synthetic 7-mer-A_{20} RNA substrate, which consists of a 5′ end 6-carboxyfluorescein (6-FAM; green) fluorophore, a 7-mer RNA body (red) and a tail of 20 adenosines (A_{20}, black). Positional marker RNAs are also depicted.

(b) Deadenylation assay time course experiment with the 7-mer-A_{20} RNA substrate and the CCR4a:CAF1 exonuclease heterodimer in equimolar ratio (50 nM) demonstrating low catalytic activity.

(c) Time course assay as in (b), but with a 5-fold molar excess of CCR4a:CAF1 (250 nM) over 7-mer-A_{20} RNA (50 nM). Under these conditions, the deadenylation reaction is rapid.

(d-f) Deadenylation assays with 50 nM 7-mer-A_{20} RNA and equimolar concentrations of the CCR4-NOT_{FULL} (d), CCR4-$NOT_{\Delta 10:11}$ (e), and CCR4-NOT_{MINI} (f) complexes. All three complexes produce very similar deadenylation patterns and display enhanced activity compared to the CCR4a:CAF1 heterodimer under the same conditions (b).

(g) Deadenylation assay with 7-mer-A_{20} RNA and the CCR4-NOT_{CORE} complex in equimolar ratio (50 nM). The deadenylation reaction is slower than with the larger reconstituted complexes (d-f) but visibly enhanced compared to the CCR4a:CAF1 heterodimer (b).

(h) Electrophoretic mobility shift assay (EMSA) with 100 nM the 7-mer-A_{20} RNA and 1 µM CCR4-NOT_{FULL} (CCR4a^{E240A}:CAF1^{D40A}) and subcomplexes. The CCR4-NOT_{FULL} (CCR4a^{E240A}:CAF1^{D40A}) complex and the NOT10:NOT11 heterodimer both bind the RNA as visible from the shift of the RNA band to apparent higher molecular weight in the respective lanes. The upshifted protein-RNA complexes did not enter the gel.

Source data for panels (b-h) are provided as a Source Data file. All assays shown in this figure and other figures are supported by technical as well as biological replicates.

Figure 3. Comparison of deadenylation activities of CAF1 and CCR4a.

(a-b) Deadenylation assays with the 7-mer-A_{20} RNA substrate (50 nM) and a five-fold molar excess of the CCR4a:CAF1 exonuclease heterodimer (250 nM) containing inactivating mutations in either CAF1 (a; D40A mutation) or CCR4a (b; E240A mutation), respectively. Both mutations decrease the rate of deadenylation compared to wildtype (Fig. 2c), and the CAF1^{D40A} mutation (a) improves the selectivity of the exonuclease heterodimer for adenosine.

(c-d) Deadenylation assays with equimolar concentrations (50 nM) of 7-mer-A_{20} RNA and CCR4-NOT_{FULL} complexes with the catalytically inactivating mutations in CAF1 (c) and CCR4a (d), respectively. The CAF1^{D40A} mutation mildly decreases deadenylation activity and leads to increased adenosine specificity compared to wildtype (Fig. 2d), while the CCR4a^{E240A} mutation results in a drastically slower deadenylation reaction.

Source data are provided as a Source Data file.

Figure 4. A peptide motif from *Drosophila* Bag-of-marbles inhibits deadenylation.

(a) EMSA with the 7-mer-A_{20} RNA and the CAF40 module in the absence or presence of MBP-tagged Bam CBM (wildtype or the double mutant). Only Bam CBM^{WT}, but not Bam CBM^{MUT} efficiently competed with RNA for CAF40 binding.

(b) Titration experiment with 50 nM of the 7-mer-A_{20} RNA and the CCR4-NOT_{CORE} complex, and the indicated concentrations (0.05-5.0 µM) of either MBP, which served as a negative control, or MBP-tagged Bam CBM^{WT}. A 10-fold excess (500 nM) of Bam CBM^{WT} was sufficient to almost completely inhibit deadenylation. The reactions were stopped after 32 minutes.

(c) Time course assay with the 7-mer-A_{20} RNA and the CCR4-NOT_{CORE} complex in equimolar ratio (50 nM). In addition, a 50-fold molar excess (2.5 µM) of either MBP, which served as a negative control, or MBP-tagged Bam CBM^{WT} or Bam CBM^{MUT} (containing L17E and M24E mutations) were included.

(d-f) Titration experiments with 50 nM of the 7-mer-A_{20} RNA and the CCR4-NOT_{MINI} (d), CCR4-NOT_{FULL} (e) or CCR4-$NOT_{\Delta 10:11}$ (f) complexes, respectively, and the indicated concentrations (0.05-5.0 µM) of either MBP, which served as a negative control, or MBP-tagged Bam CBM^{WT}. The reactions were stopped after 32 minutes.

Source data are provided as a Source Data file.

Figure 5. Motifs from multiple CCR4-NOT recruitment factors inhibit deadenylation.

(a-c) Deadenylation assays with equimolar concentrations (50 nM) of the 7-mer-A_{20} substrate and the CCR4-NOT_{MINI} (a), CCR4-NOT_{FULL} (b), or CCR4-$NOT_{\Delta 10:11}$ (c) complexes, respectively. 2.5 µM of either MBP (as a negative control) or MBP-tagged Bam CBM^{WT}, Roquin-C or NOT4 CBM, respectively, were tested with the reconstituted complexes. All three motifs inhibited the deadenylation reaction of all three complexes but the extent of the inhibition was variable. The CCR4-NOT complexes lacking the NOT10:NOT11 heterodimer were strikingly more sensitive to the inhibition by the peptide motifs.

Source data are provided as a Source Data file.

Figure 6. The length of the segment preceding the poly(A) tail but not its sequence impacts on deadenylation.

(a) Deadenylation assay with the CCR4-NOT_{FULL} complex and the 13+7-mer-A_{20} RNA substrate in equimolar ratio (50 nM). This substrate contains an additional unstructured 13-mer segment 5′ to the 7-mer-A_{20} RNA. The deadenylation of this substrate is markedly enhanced compared to the 7-mer-A_{20} substrate (Fig. 2d) and did not proceed beyond the first two consecutive non-A nucleotides.

(b) Deadenylation assay with the same 13+7-mer-A_{20} substrate (50 nM) and a five-fold molar excess of the CCR4a:CAF1 exonuclease heterodimer (250 nM). The rate of deadenylation does not differ to that with the 7-mer-A_{20} substrate indicating that the exonuclease heterodimer does not discriminate between these substrates (Fig. 2c).

(c) Deadenylation assay with the CCR4-$NOT_{\Delta 10:11}$ complex and the 13+7-mer-A_{20} RNA substrate in equimolar ratio (50 nM).

(d) Deadenylation assay with the CCR4-NOT_{FULL} complex and the 13+7-mer-A_{20} substrate (both 50 nM) with higher resolution of the early time points compared to (a).

(e) Deadenylation assay with the CCR4-NOT_{FULL} complex and a 20-mer-A_{20} substrate where the seven nucleotides immediately preceding the poly(A) tail were different from the 13+7-mer-A_{20} substrate in equimolar ratio (50 nM).

(f) Deadenylation assay with the CCR4-NOT_{FULL} complex and the 13-SL-7-mer-A_{20} substrate containing a short stem-loop structure within the 20-nucleotide region upstream of A_{20} in equimolar ratio (50 nM) (see also Supplementary Table 4).

Source data are provided as a Source Data file.

Figure 7. A proposed mechanistic model for the switch from bulk to targeted deadenylation.

(a) The CCR4-NOT complex possesses intrinsic affinity towards nucleic acids which allows it to bind to and deadenylate bulk mRNAs in a manner independent of the sequence of the RNA. In addition to the active sites of the nucleases, also the CAF40 module, the NOT module, and the NOT10:NOT11 module can directly interact with the substrate and contribute to a different

extent to the stimulation of deadenylation, with CAF40 being the dominant stimulatory site under the conditions of our in vitro assays.

(b) Upon stimulation, RNA-binding proteins (RBPs) can recruit the CCR4-NOT complex to specific transcripts via direct protein-protein interactions and trigger the preferred deadenylation of those transcripts. In addition, RBPs carrying a CAF40-binding motif can mask the RNA-binding site on CAF40, thereby gaining further competitive advantage for deadenylation of their bound mRNA over other, non-specific transcripts in the cytoplasmic mRNA pool.

REFERENCES

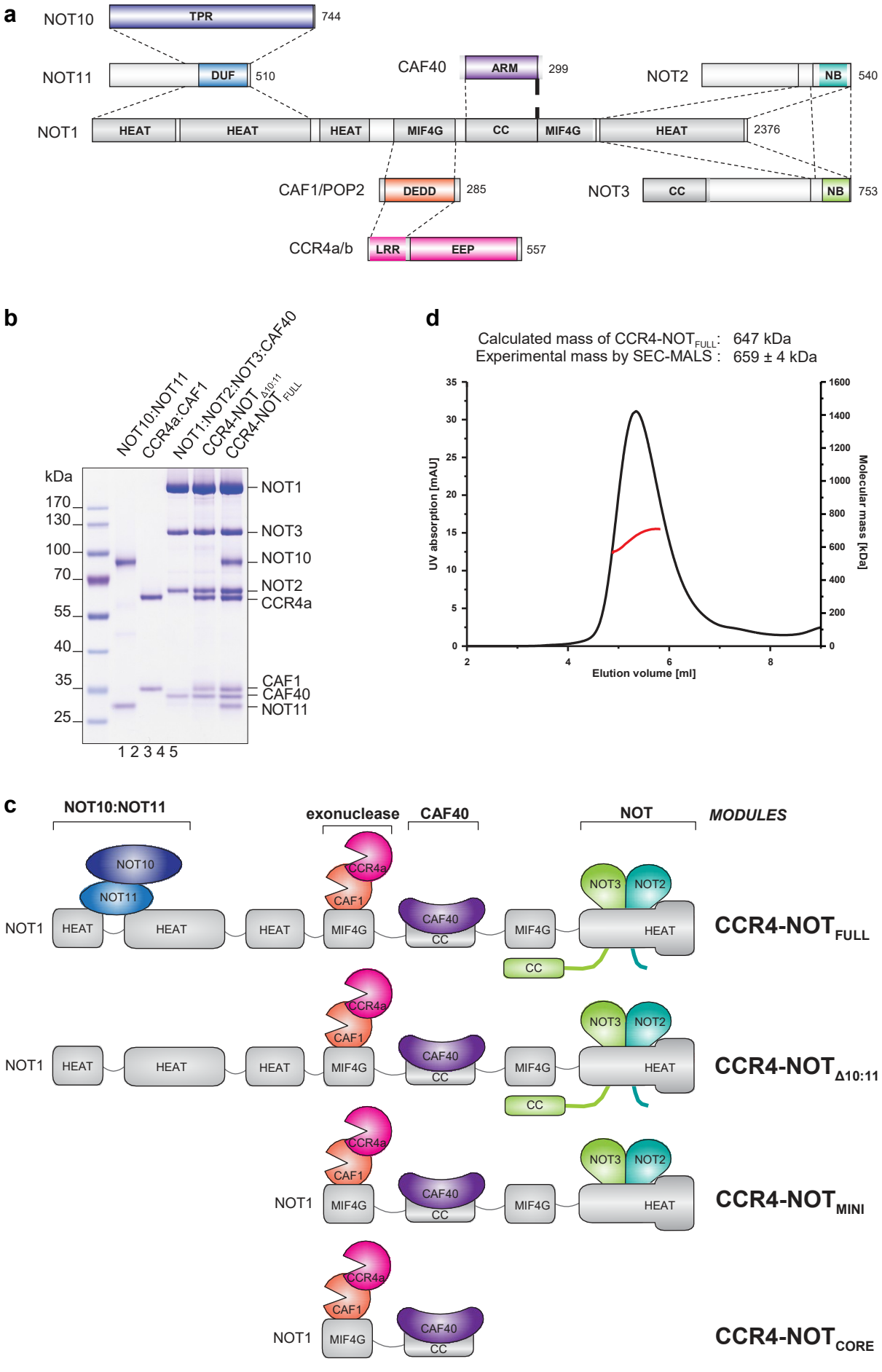
1. Wahle, E. Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.* **270**, 2800–2808 (1995).
2. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
3. Lima, S. A. *et al.* Short poly(A) tails are a conserved feature of highly expressed genes. *Nat. Struct. Mol. Biol.* **24**, 1057–1063 (2017).
4. Temme, C., Simonelig, M. & Wahle, E. Deadenylation of mRNA by the CCR4-NOT complex in Drosophila: molecular and developmental aspects. *Front. Genet.* **5**, 143 (2014).
5. Collart, M. A. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* **7**, 438–454 (2016).
6. Yamashita, A. *et al.* Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.* **12**, 1054–1063 (2005).
7. Yi, H. *et al.* PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol. Cell* **70**, 1081–1088.e5 (2018).
8. Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* **44**, 120–133 (2011).
9. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).
10. Bhandari, D., Raisch, T., Weichenrieder, O., Jonas, S. & Izaurralde, E. Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes Dev.* **28**, 888–901 (2014).
11. Raisch, T. *et al.* Distinct modes of recruitment of the CCR4-NOT complex by Drosophila and vertebrate Nanos. *EMBO J.* **35**, 974–990 (2016).
12. Van Etten, J. *et al.* Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.* **287**, 36370–36383 (2012).
13. Sgromo, A. *et al.* A CAF40-binding motif facilitates recruitment of the CCR4-NOT complex to mRNAs targeted by Drosophila Roquin. *Nat. Commun.* **8**, 14307 (2017).
14. Chekulaeva, M. *et al.* miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* **18**, 1218–1226 (2011).
15. Fabian, M. R. *et al.* miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat. Struct. Mol. Biol.* **18**, 1211–

964 1217 (2011).
965 16. Sarshad, A. A. *et al.* Argonaute-miRNA Complexes Silence Target mRNAs in the
966 Nucleus of Mammalian Stem Cells. *Mol. Cell* **71**, 1040–1050.e8 (2018).
967 17. Maillet, L., Tu, C., Hong, Y. K., Shuster, E. O. & Collart, M. A. The essential function of
968 Not1 lies within the Ccr4-Not complex. *J. Mol. Biol.* **303**, 131–143 (2000).
969 18. Temme, C. *et al.* Subunits of the Drosophila CCR4-NOT complex and their roles in
970 mRNA deadenylation. *RNA* **16**, 1356–1370 (2010).
971 19. Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S. & Izaurralde, E. NOT10 and
972 C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto
973 the NOT1 N-terminal domain. *RNA Biol.* **10**, 228–244 (2013).
974 20. Sgromo, A. *et al.* Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-
975 NOT complex to elicit repression of mRNA targets. *RNA* **24**, 381–395 (2018).
976 21. Albert, T. K. *et al.* Isolation and characterization of human orthologs of yeast CCR4-
977 NOT complex subunits. *Nucleic Acids Res.* **28**, 809–817 (2000).
978 22. Chen, J. *et al.* Purification and characterization of the 1.0 MDa CCR4-NOT complex
979 identifies two novel components of the complex. *J. Mol. Biol.* **314**, 683–694 (2001).
980 23. Lau, N.-C. *et al.* Human Ccr4-Not complexes contain variable deadenylase subunits.
981 *Biochem. J* **422**, 443–453 (2009).
982 24. Tucker, M., Staples, R. R., Valencia-Sanchez, M. A., Muhlrad, D. & Parker, R. Ccr4p is
983 the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in
984 Saccharomyces cerevisiae. *EMBO J.* **21**, 1427–1436 (2002).
985 25. Nousch, M., Techritz, N., Hampel, D., Millonigg, S. & Eckmann, C. R. The Ccr4-Not
986 deadenylase complex constitutes the main poly(A) removal activity in C. elegans. *J. Cell*
987 *Sci.* **126**, 4274–4285 (2013).
988 26. Temme, C., Zaessinger, S., Meyer, S., Simonelig, M. & Wahle, E. A complex containing
989 the CCR4 and CAF1 proteins is involved in mRNA deadenylation in Drosophila. *EMBO*
990 *J.* **23**, 2862–2871 (2004).
991 27. Piao, X., Zhang, X., Wu, L. & Belasco, J. G. CCR4-NOT deadenylates mRNA
992 associated with RNA-induced silencing complexes in human cells. *Mol. Cell. Biol.* **30**,
993 1486–1494 (2010).
994 28. Webster, M. W. *et al.* mRNA Deadenylation Is Coupled to Translation Rates by the
995 Differential Activities of Ccr4-Not Nucleases. *Mol. Cell* **70**, 1089–1100.e8 (2018).
996 29. Bai, Y. *et al.* The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically
997 and functionally separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.* **19**, 6642–
998 6651 (1999).
999 30. Albert, T. K. *et al.* Identification of a ubiquitin-protein ligase subunit within the CCR4-
1000 NOT transcription repressor complex. *EMBO J.* **21**, 355–364 (2002).

1001 31. Mauxion, F., Prève, B. & Séraphin, B. C2ORF29/CNOT11 and CNOT10 form a new
1002 module of the CCR4-NOT complex. *RNA Biol.* **10**, 267–276 (2013).
1003 32. Fabian, M. R. *et al.* Structural basis for the recruitment of the human CCR4-NOT
1004 deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* **20**, 735–739 (2013).
1005 33. Chen, Y. *et al.* A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct
1006 links between miRNA target recognition and silencing. *Mol. Cell* **54**, 737–750 (2014).
1007 34. Mathys, H. *et al.* Structural and biochemical insights to the role of the CCR4-NOT
1008 complex and DDX6 ATPase in microRNA repression. *Mol. Cell* **54**, 751–765 (2014).
1009 35. Stowell, J. A. W. *et al.* Reconstitution of Targeted Deadenylation by the Ccr4-Not
1010 Complex and the YTH Domain Protein Mmi1. *Cell Rep.* **17**, 1978–1989 (2016).
1011 36. Horiuchi, M. *et al.* Structural basis for the antiproliferative activity of the Tob-hCaf1
1012 complex. *J. Biol. Chem.* **284**, 13244–13255 (2009).
1013 37. Wang, H. *et al.* Crystal structure of the human CNOT6L nuclease domain reveals strict
1014 poly(A) substrate specificity. *EMBO J.* **29**, 2566–2576 (2010).
1015 38. Petit, A.-P. *et al.* The structural basis for the interaction between the CAF1 nuclease and
1016 the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res.*
1017 **40**, 11058–11072 (2012).
1018 39. Boland, A. *et al.* Structure and assembly of the NOT module of the human CCR4-NOT
1019 complex. *Nat. Struct. Mol. Biol.* **20**, 1289–1297 (2013).
1020 40. Raisch, T., Sandmeir, F., Weichenrieder, O., Valkov, E. & Izaurralde, E. Structural and
1021 biochemical analysis of a NOT1 MIF4G-like domain of the CCR4-NOT complex. *J.*
1022 *Struct. Biol.* **204**, 388–395 (2018).
1023 41. Garces, R. G., Gillon, W. & Pai, E. F. Atomic model of human Rcd-1 reveals an
1024 armadillo-like-repeat protein with in vitro nucleic acid binding properties. *Protein Sci.* **16**,
1025 176–188 (2007).
1026 42. Sari, D. *et al.* The MultiBac Baculovirus/Insect Cell Expression Vector System for
1027 Producing Complex Protein Biologics. *Adv. Exp. Med. Biol.* **896**, 199–215 (2016).
1028 43. Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research
1029 toolbox for multiprotein complexes. *Trends Biochem. Sci.* **37**, 49–57 (2012).
1030 44. Bhaskar, V. *et al.* Structure and RNA-binding properties of the Not1-Not2-Not5 module
1031 of the yeast Ccr4-Not complex. *Nat. Struct. Mol. Biol.* **20**, 1281–1288 (2013).
1032 45. Webster, M. W., Stowell, J. A. W., Tang, T. T. L. & Passmore, L. A. Analysis of mRNA
1033 deadenylation by multi-protein complexes. *Methods* **126**, 95–104 (2017).
1034 46. Niinuma, S., Fukaya, T. & Tomari, Y. CCR4 and CAF1 deadenylases have an intrinsic
1035 activity to remove the post-poly(A) sequence. *RNA* **22**, 1550–1559 (2016).
1036 47. Pavanello, L., Hall, B., Airhihen, B. & Winkler, G. S. The central region of CNOT1 and
1037 CNOT9 stimulates deadenylation by the Ccr4-Not nuclease module. *Biochem. J* **475**,

3437–3450 (2018).
48. Maryati, M., Airhihen, B. & Winkler, G. S. The enzyme activities of Caf1 and Ccr4 are both required for deadenylation by the human Ccr4-Not nuclease module. *Biochem. J* **469**, 169–176 (2015).
49. Jonstrup, A. T., Andersen, K. R., Van, L. B. & Brodersen, D. E. The 1.4-A crystal structure of the S. pombe Pop2p deadenylase subunit unveils the configuration of an active enzyme. *Nucleic Acids Res.* **35**, 3153–3164 (2007).
50. Keskeny, C. *et al.* A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4-NOT complex. *Genes Dev.* **33**, 236–252 (2019).
51. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
52. Nasertorabi, F., Batisse, C., Diepholz, M., Suck, D. & Böttcher, B. Insights into the structure of the CCR4-NOT complex by electron microscopy. *FEBS Lett.* **585**, 2182–2186 (2011).
53. Ukleja, M. *et al.* The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. *Nat. Commun.* **7**, 10433 (2016).
54. Lim, J. *et al.* Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science* **361**, 701–704 (2018).
55. Shirai, Y.-T., Suzuki, T., Morita, M., Takahashi, A. & Yamamoto, T. Multifunctional roles of the mammalian CCR4-NOT complex in physiological phenomena. *Front. Genet.* **5**, 286 (2014).
56. Bulbrook, D. *et al.* Tryptophan-Mediated Interactions between Tristetraprolin and the CNOT9 Subunit Are Required for CCR4-NOT Deadenylase Complex Recruitment. *J. Mol. Biol.* **430**, 722–736 (2018).
57. Donnelly, M. I. *et al.* An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr. Purif.* **47**, 446–454 (2006).
58. Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. *J. Struct. Biol.* **175**, 178–188 (2011).
59. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).

Raisch et al. Figure 1



Raisch et al. Figure 2

a

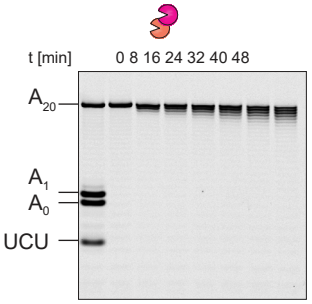
A_{20} 5′-6-FAM-UCUAAAUAAAAAAAAAAAAAAAAAAAA-3′

A_1 5′-6-FAM-UCUAAAUA-3′

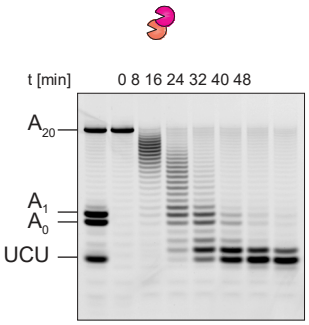
A_0 5′-6-FAM-UCUAAAU-3′

UCU 5′-6-FAM-UCU-3′

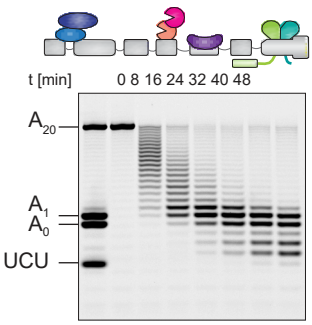
b CCR4a:CAF1 [1:1 protein:RNA]



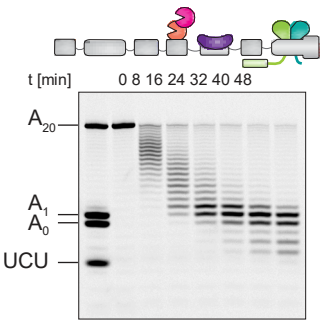
c CCR4a:CAF1 [5:1 protein:RNA]



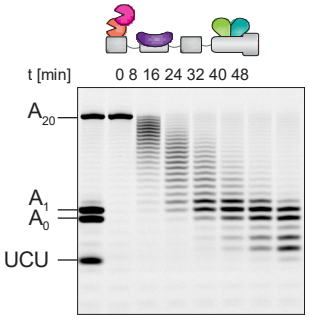
d CCR4-NOT$_{FULL}$



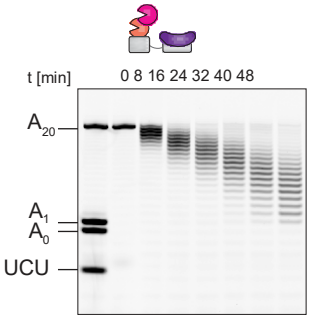
e CCR4-NOT$_{\Delta 10:11}$



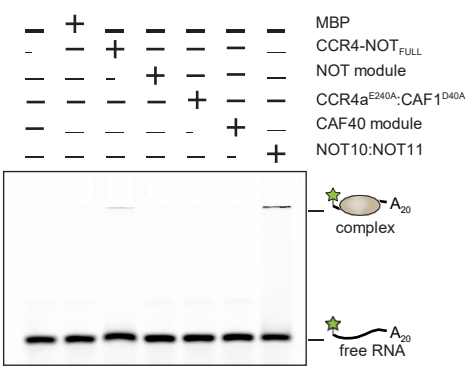
f CCR4-NOT$_{MINI}$



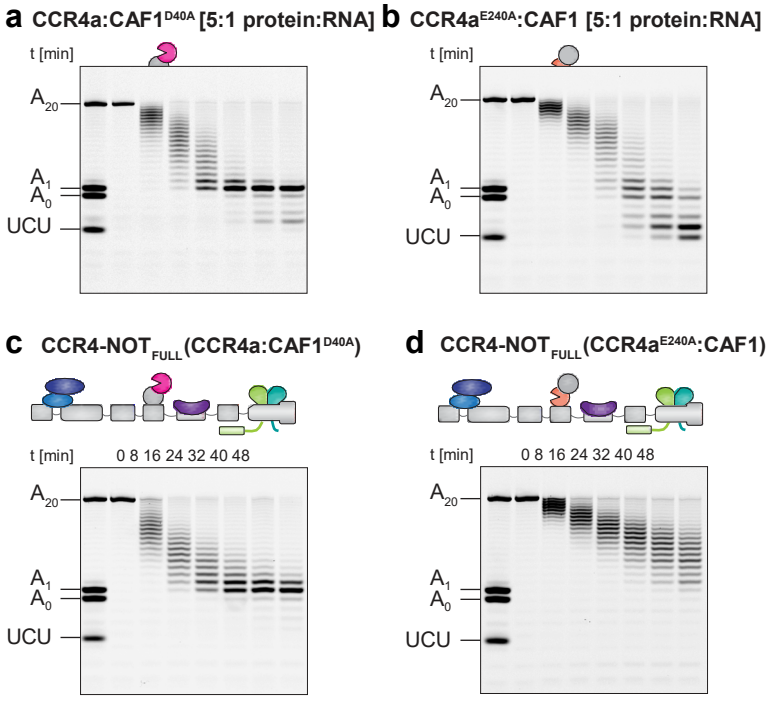
g CCR4-NOT$_{CORE}$



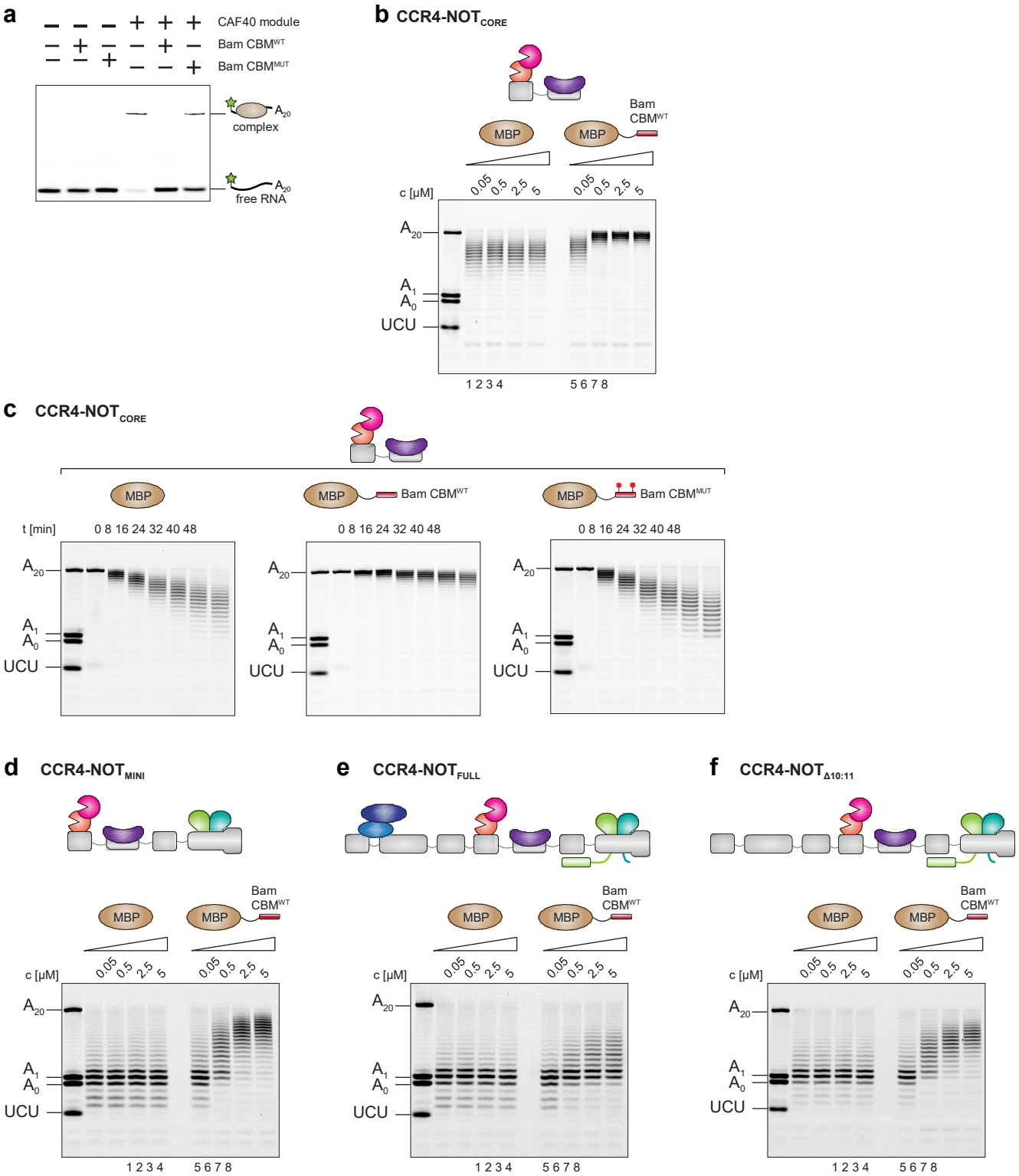
h



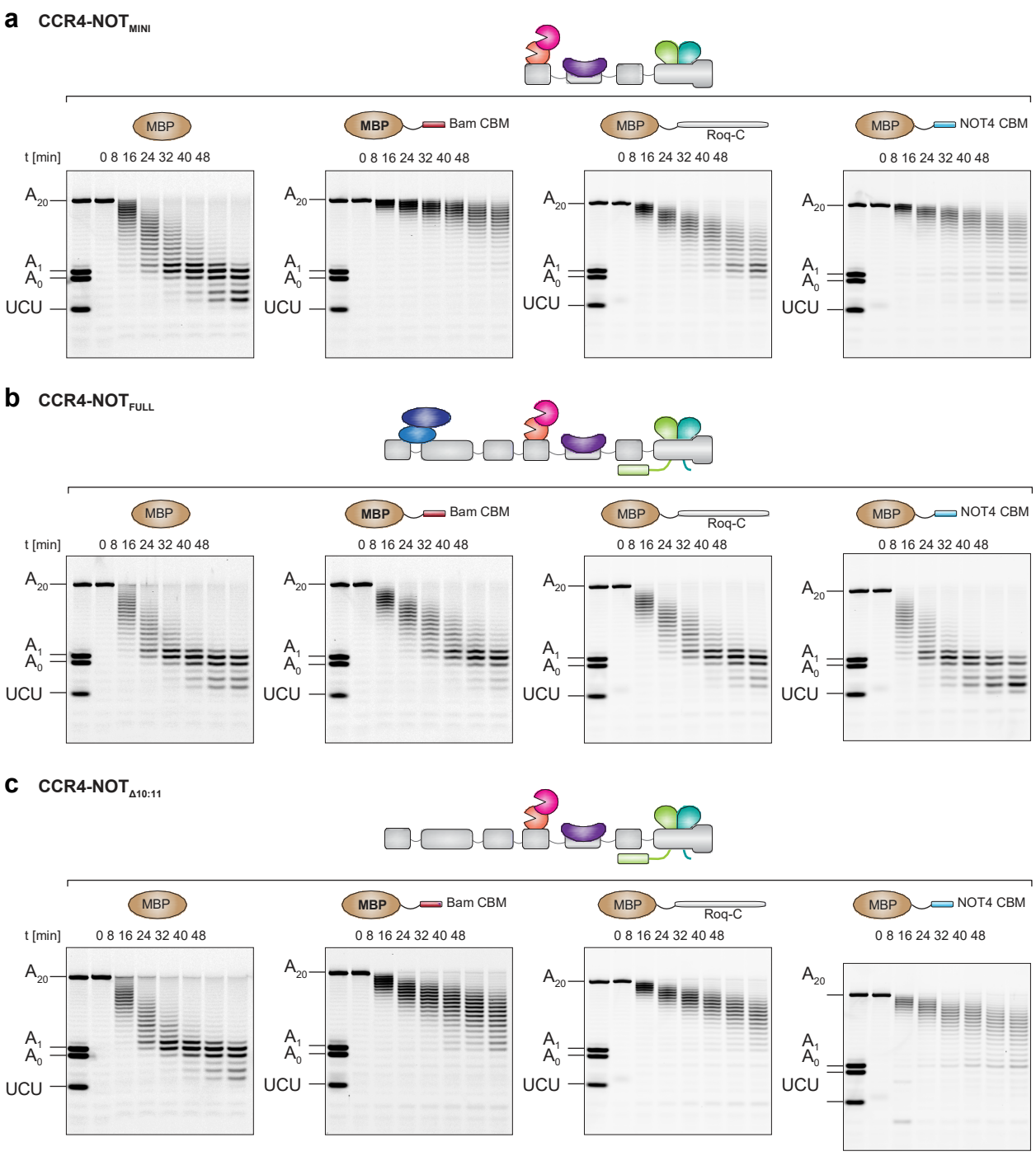
Raisch et al. Figure 3



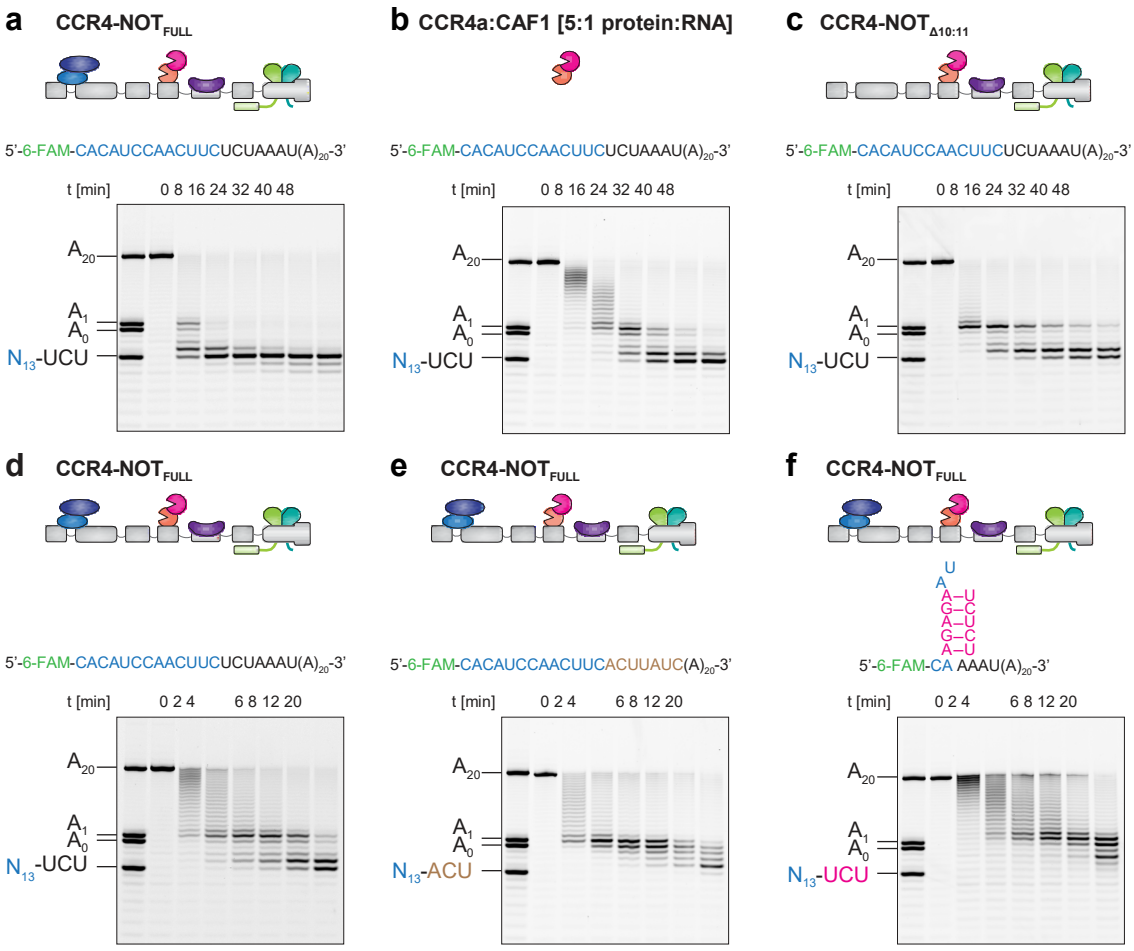
Raisch et al. Figure 4



Raisch et al. Figure 5

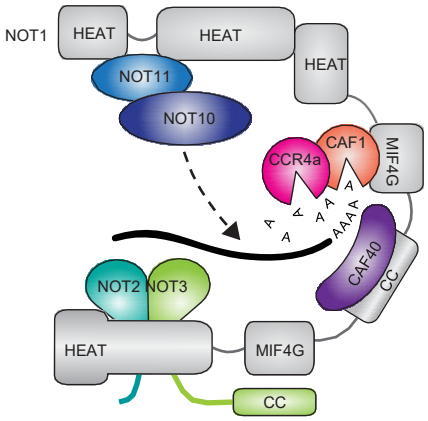


Raisch et al. Figure 6

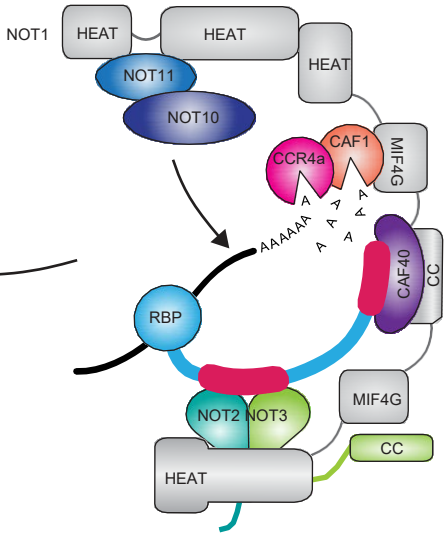


Raisch et al. Figure 7

a Bulk, nonspecific deadenylation



b Targeted deadenylation



SUPPLEMENTARY INFORMATION

Reconstitution of recombinant human CCR4-NOT reveals molecular insights into regulated deadenylation

Raisch et al.

Supplementary Table 1. Subunits of the human CCR4-NOT complex.

Subunit	UniProt ID	No. of residues	Molecular mass (kDa)	Synonyms
NOT1	A5YKK6-1	2376	266.9	CNOT1
NOT2	Q9NZN8-1	540	59.7	CNOT2
NOT3	O75175-1	753	81.9	CNOT3
CAF40	Q92600-1	299	33.6	CNOT9, RCD-1
CCR4a	Q9ULM6-1	557	63.3	CNOT6
CAF1	Q9UIV1-1	285	32.7	CNOT7
NOT10	Q9H9A5-1	744	82.3	CNOT10
NOT11	Q9UKZ1-1	510	55.2	CNOT11, C2ORF29
Total		6064	675.6	

Supplementary Table 2. Plasmids used in this study.

Plasmid	Protein	Residues	Tag	Protease site	Restriction sites	Comments
pACEBac-NOT1-CAF40						Acceptor plasmid for recombination with pIDK-NOT2-NOT3
	NOT1	1-2376	His_{10} (N)	TEV	-	
			$StrepII_2$ (C)	-		
	CAF40	1-299	-	-	-	
pIDK-NOT2-NOT3						Donor plasmid for recombination with pACEBac-NOT1-CAF40
	NOT2	1-540	-	-	XhoI/NcoI	
	NOT3	1-753	-	-	KpnI	
pACEBac-NOT1-CAF40-NOT2-NOT3						Bacmid for production of NOT1:NOT2:NOT3:CAF40 complex. Created by Cre-Lox recombination of pACEBac-NOT1-CAF40 and pIDK-NOT2-NOT3. For reconstitution of CCR4-NOT_{FULL} and CCR4-$NOT_{\Delta 10:11}$
	NOT1	1-2376	His_{10} (N)	TEV	-	
			$StrepII_2$ (C)	-		
	CAF40	1-299	-	-	-	
	NOT2	1-540	-	-	XhoI/NcoI	
	NOT3	1-753	-	-	KpnI	
pMCSG19c-CCR4a						Used for production of CCR4a:CAF1 dimer.

	CCR4a	1-557	MBP (N)	TVMV	KpnI/BamHI	For reconstitution of all larger complexes. MBP is proteolytically cleaved *in vivo*.
			His_6 (N)	TEV		
pET28b-SUMO-CAF1						Used for production of CCR4a:CAF1 dimer. For reconstitution of all larger complexes
	CAF1	285	SUMO-His_6 (N)	TEV	BamHI/HindIII	
pMCSG19c-CCR4a^{E240A}						Used for production of CCR4a^{E240A}:CAF1 dimer. For reconstitution of all larger complexes. MBP is proteolytically cleaved *in vivo*. Generated from wildtype plasmid using site-directed mutagenesis
	CCR4a	1-557	MBP (N)	TVMV	KpnI/BamHI	
			His_6 (N)	TEV		
pET28b-CAF1^{D40A}						Used for production of CCR4a:CAF1^{D40A} dimer. For reconstitution of all larger complexes Generated from wildtype plasmid using site-directed mutagenesis
	CAF1	285	SUMO-His_6 (N)	TEV	BamHI/HindIII	
pnYC-pM-NOT10_25-707						Used for production of NOT10:NOT11 dimer. For reconstitution of CCR4-NOT_{FULL}
	NOT10	25-707	MBP (N)	HRV-3C	XhoI/BamHI	
pnEA-NOT11_257-498-vH						Used for production of NOT10:NOT11 dimer. For reconstitution of CCR4-NOT_{FULL}
	NOT11	257-498	His_6 (C)	TEV	XhoI/BamHI	
pnYC-pM-NOT1_1093-2376						*Sgromo et al., 2017* For reconstitution of CCR4-NOT_{MINI}
	NOT1	1093-2376	MBP (N)	HRV-3C	XhoI/BamHI	
pnEA-H-NOT3_607-753-pM-NOT2_344-540-pH-CAF40_19-285						For reconstitution of CCR4-NOT_{MINI}

	NOT3	607-753	His_6 (N)	-	NdeI/BamHI	
	NOT2	344-540	MBP (N)	HRV-3C	NdeI/BamHI	
	CAF40	19-285	His_6 (N)	HRV-3C	XhoI/BamHI	
pnYC-pM-NOT1_1093-1607						For reconstitution of CCR4-NOT_{CORE}
	NOT1	1093-1607	MBP (N)	HRV-3C	XhoI/BamHI	
pnEA-CAF40_19-285						For reconstitution of CCR4-NOT_{CORE}
	CAF40	19-285	-	-	XhoI/BamHI	
	NOT1	1-682	-	-	-	
pnYC-pM-NOT1_1351-1588						For production of CAF40 module
	NOT1	1351-1588	MBP (N)	HRV-3C	XhoI/BamHI	
pnEA-pH-CAF40_19-285						*Chen et al., 2014* For production of CAF40 module
	CAF40	19-285	His_6 (N)	HRV-3C	XhoI/BamHI	
pnYC-pM-NOT1_1833-2361						*Bhandari et al., 2014* For production of NOT module
	NOT1	1833-2361	MBP (N)	HRV-3C	XhoI/BamHI	
pnEA-pH-NOT3_607-748-pM-NOT2_350-540						*Bhandari et al., 2014* For production of NOT module
	NOT3	607-748	His_6 (N)	HRV-3C	XhoI/BamHI	

	NOT2	350-540	MBP (N)	HRV-3C	XhoI/BamHI	
pnYC-vM-DmBam-CBM						*Sgromo et al., 2018*
	Bam	13-36	MBP (N)	TEV	XhoI/AvrII	
pnYC-vM-DmBam-CBM-Mut						Contains L17E and M24E mutations Generated from wildtype plasmid using site-directed mutagenesis
	Bam	13-36	MBP (N)	TEV	XhoI/AvrII	
pnYC-pM-Roquin1-C						*Sgromo et al., 2018*
	Roquin1	501-1133	MBP (N)	HRV-3C	AflII/AvrII	
pnEA-pM-NOT4-CBM						*Keskeny et al., 2019*
	NOT4	400-427	MBP (N)	HRV-3C	XhoI/NheI	
pnEA-pM-CAF40_18-293						For UV crosslinking
	CAF40	18-293	MBP (N)	HRV-3C	XhoI/BamHI	

Supplementary Table 3. Complexes investigated in this study.

CCR4-NOT$_{FULL}$				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	Sf21 insect cells	2408	270.7 kDa	Full-length, N-terminal His_{10} tag, C-terminal $StrepII_2$ tag
NOT2	Sf21 insect cells	540	59.7 kDa	Full-length
NOT3	Sf21 insect cells	753	81.9 kDa	Full-length
CAF40	Sf21 insect cells	299	33.6 kDa	Full-length
CCR4a	*E. coli* BL21(DE3) Star	558	63.4 kDa	Full-length, N-terminal His_6 tag (proteolytically removed)
CAF1	*E. coli* BL21(DE3) Star	286	32.8 kDa	Full-length, N-terminal His_6-SUMO tag (proteolytically removed)
NOT10	*E. coli* BL21(DE3) Star	689	76.4 kDa	Residues 25-707, N-terminal MBP tag (proteolytically removed)
NOT11	*E. coli* BL21(DE3) Star	253	28.8 kDa	Residues 257-498, C-terminal His_6 tag (proteolytically removed)
Total		5786	647.3 kDa	
CCR4-NOT$_{\Delta 10:11}$				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	Sf21 insect cells	2408	270.7 kDa	Full-length, N-terminal His_{10} tag, C-terminal $StrepII_2$ tag
NOT2	Sf21 insect cells	540	59.7 kDa	Full-length

NOT3	Sf21 insect cells	753	81.9 kDa	Full-length
CAF40	Sf21 insect cells	299	33.6 kDa	Full-length
CCR4a	*E. coli* BL21(DE3) Star	558	63.4 kDa	Full-length, N-terminal His_6 tag (proteolytically removed)
CAF1	*E. coli* BL21(DE3) Star	286	32.8 kDa	Full-length, N-terminal His_6-SUMO tag (proteolytically removed)
Total		4844	542.1 kDa	
CCR4-NOT$_{MINI}$				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	*E. coli* BL21(DE3) Star	1292	146.9 kDa	Residues 1093-2376, N-terminal His_6 tag (proteolytically removed)
NOT2	*E. coli* BL21(DE3) Star	201	23.3 kDa	Residues 344-540, N-terminal MBP tag (proteolytically removed)
NOT3	*E. coli* BL21(DE3) Star	166	20.2 kDa	Residues 607-753, N-terminal His_6 tag
CAF40	*E. coli* BL21(DE3) Star	273	31.0 kDa	Residues 19-285, N-terminal His_6 tag (proteolytically removed)
CCR4a	*E. coli* BL21(DE3) Star	558	63.4 kDa	Full-length, N-terminal His_6 tag (proteolytically removed)
CAF1	*E. coli* BL21(DE3) Star	286	32.8 kDa	Full-length, N-terminal His_6-SUMO tag (proteolytically removed)
Total		2776	317.6 kDa	

CCR4-NOT$_{CORE}$				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	*E. coli* BL21(DE3) Star	521	59.2 kDa	Residues 1093-1607, N-terminal MBP tag (proteolytically removed)
CCR4a	*E. coli* BL21(DE3) Star	558	63.4 kDa	Full-length, N-terminal His_6 tag (proteolytically removed)
CAF1	*E. coli* BL21(DE3) Star	286	32.8 kDa	Full-length, N-terminal His_6-SUMO tag (proteolytically removed)
CAF40	*E. coli* BL21(DE3) Star	270	30.7 kDa	Residues 19-285
Total		1635	186.1 kDa	

NOT10:NOT11 heterodimer				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT10	*E. coli* BL21(DE3) Star	689	76.4 kDa	Residues 25-707, N-terminal MBP tag (proteolytically removed)
NOT11	*E. coli* BL21(DE3) Star	253	28.8 kDa	Residues 257-498, C-terminal His_6 tag (proteolytically removed)
Total		942	105.2 kDa	

CCR4a:CAF1 exonuclease heterodimer				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
CCR4a	*E. coli* BL21(DE3) Star	558	63.4 kDa	Full-length, N-terminal His_6 tag (proteolytically removed)
CAF1	*E. coli* BL21(DE3) Star	286	32.8 kDa	Full-length, N-terminal His_6-SUMO tag (proteolytically removed)

Total		844	96.2 kDa	
CAF40 module				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	*E. coli* BL21(DE3) Star	244	27.6 kDa	Residues 1351-1588, N-terminal MBP tag (proteolytically removed)
CAF40	*E. coli* BL21(DE3) Star	273	31.0 kDa	Residues 19-285, N-terminal His_6 tag (proteolytically removed)
Total		517	58.6 kDa	
NOT module				
Subunit	**Expression host**	**Residues**	**Mol. weight**	**Comment**
NOT1	*E. coli* BL21(DE3) Star	535	61.5 kDa	Residues 1833-2361, N-terminal MBP tag (proteolytically removed)
NOT2	*E. coli* BL21(DE3) Star	197	22.9 kDa	Residues 350-540, N-terminal MBP tag (proteolytically removed)
NOT3	*E. coli* BL21(DE3) Star	148	18.2 kDa	Residues 607-748, N-terminal His_6 tag (proteolytically removed)
Total		880	102.6 kDa	

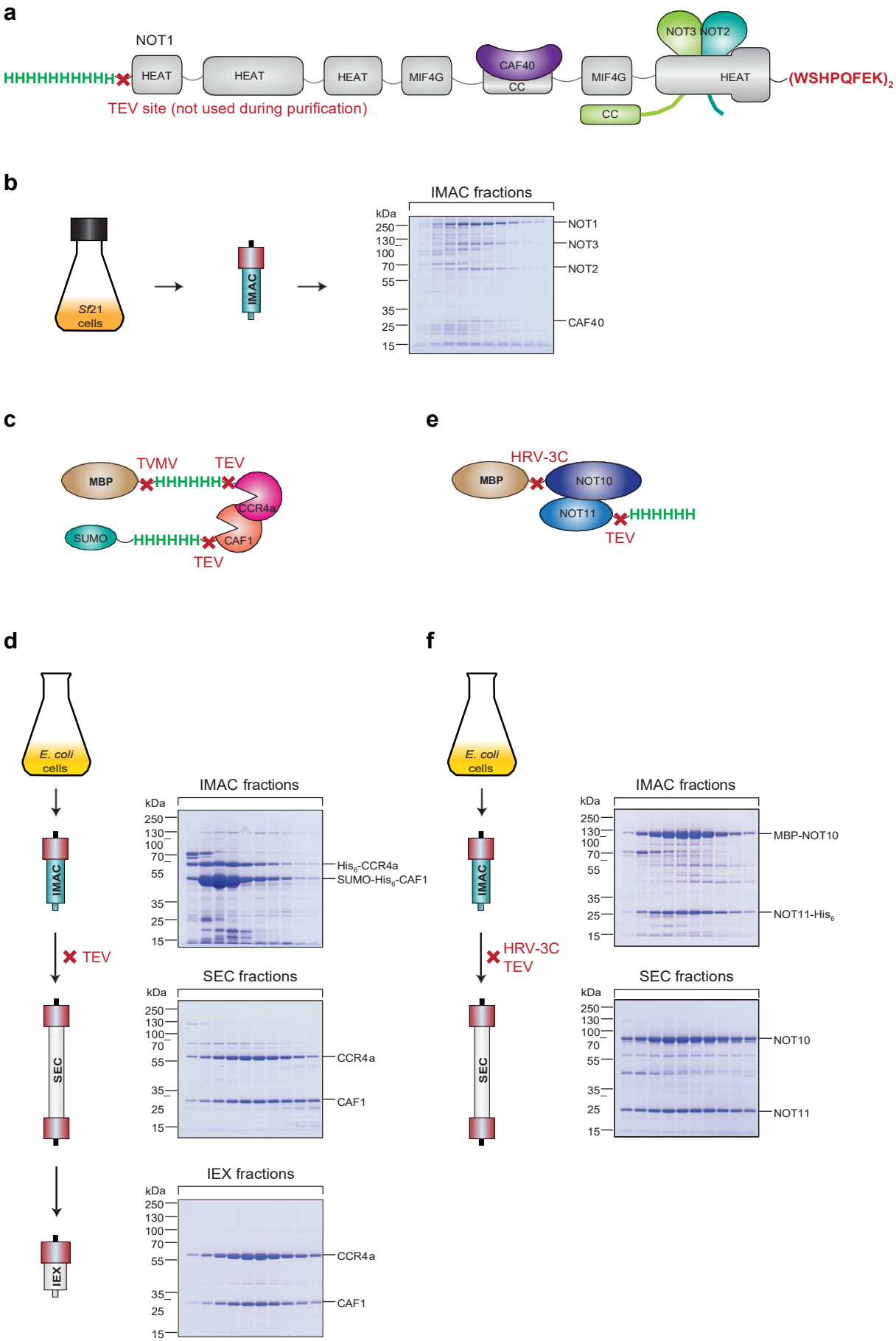
Supplementary Table 4. List of RNA substrates.

Name	Sequence	Comments
7-mer-A_{20}	UCUAAAUAAAAAAAAAAAAAAAAAAAAA	AAUAAA - sequence used by Wang et al., 2010
13+7-mer-A_{20}	CACAUCCAACUUCUCUAAAUAAA AAAAAAAAAAAAAAAAA	Underlined sequence is the 7-mer above. 13 bases predicted to be unstructured are 5' terminal to the 7-mer.
13-SL-7mer-A_{20}	CA**AGAGA**AUCC**UCUCU**AAAUAAA AAAAAAAAAAAAAAAAA	Residues highlighted in bold base-pair to form a stem-loop.
20-mer-A_{20}	CACAUCCAACUUCACUUAUCAAA AAAAAAAAAAAAAAAAA	Underlined sequence is altered compared to 7-mer-A_{20} with preceding 13 bases identical to 13-US-7-mer-A_{20}
4G-20-mer-A_{20}	CA***G***AUCCAA***G***UU***A***A***G***UUAU***G***AAA AAAAAAAAAAAAAAAAA	Derived from 20-mer-A_{20}; the G nucleotide substitutions are highlighted in bold italic

Supplementary Table 5. List of DNA oligonucleotide primers used in this study.

Name	Sequence (5′—3′)	Description
EV048	ATTATCTCCATGATCTATTAATATTCCGGGCAGCCGGATCTTCTAGGC	SV40/polh cassette assembly
EV049	GCCTAGAAGATCCGGCTGCCCGGAATATTAATAGATCATGGAGATAAT	SV40/polh cassette assembly
EV053	TGAACCTGAAACATAAAATGAATGC	SV40/polh cassette assembly
EV056	GCGCCCGATGGTGGGACG	SV40/polh cassette assembly
EV051	CGTCCCACCATCGGGCGCATGCACAGCCTGGCGACGGCTG	NOT9 forward
EV052	TCACTGAGGGGGCAGGGG	NOT9 reverse
EV155	GATCGTCGACATGGCCCACCACCACCACCACCACCACCACCACGAGAACCTGTACTTCCAGGGCATGAATCTTGACTCGCTCTCG	NOT1 forward
EV156	GATCGCGGCCGCGCATTCATTTTATGTTTCAGGTTCATCACTTCTCGAACTGGGGGTGGCTCCAGCCGCCGCTGCCGCCACTGGCACCTGTCCCTTCC	NOT1 reverse
EV157	GATCTCTAGATGAACCTGAAACATAAAATGAATGCAATT	NOT9 cassette forward
EV158	GATCAAGCTTTCACTGAGGGGGCAGGGGGATAC	NOT9 cassette reverse
EV149	GATCGGTACCATGGCGGACAAGCGCAAACTCCAAG	NOT3 forward
EV150	GATCGGTACCTCACTGGAGGTCCCGGTCCTCC	NOT3 reverse
EV147	GATCCTCGAGATGGTGAGGACTGATGGACATACATTATCT	NOT2 forward
EV148	GATCCCATGGTTAGAAGGCTTGCTGAGCAGGG	NOT2 reverse
EV075	TCTCATGCTGGAGTTCTTCGC	NOT2 cassette forward
EV076	GATCATCGATGGGTGGGGAAAAGGAAGAAAC	NOT2 cassette reverse
EV412	GATCGGATCCGAAAACCTTTACTTCCAGGGCATGCCAGCGGCAACTGTAGATCATA	CAF1 forward
EV413	GATCAAGCTTTCATGACTGCTTGTTGGCTTCCTC	CAF1 reverse
EV408	GATCGGTACCGAAAACCTTTACTTCCAGGGCATGCCCAAAGAAAAATACGAGCC	CCR4a forward
EV409	GATCGGATCCCTACCTCCTGCCAGGAAGGTGG	CCR4a reverse
TR367	TACGCCCTCGAGGATCAAGAGAAGGAGTTATCCAC	NOT10 forward
TR372	CGTAATGGATCCTTACTGATTCCTTTTGATGATCTGTAAG	NOT10 reverse
TR361	TACGCCCTCGAGGACAGCTCAGTTGCCTCTCAG	NOT11 forward
TR711	CGTAATGGATCCATCCAATGTCTTCAACAACCGG	NOT11 reverse

Supplementary Figure 1



Supplementary Figure 1. Production of subcomplexes used for CCR4-NOT$_{FULL}$ reconstitution.

(a) Schematic representation of the NOT1:NOT2:NOT3:CAF40 subcomplex in the same style as in Fig. 1c.

(b) Purification of the NOT1:NOT2:NOT3:CAF40 subcomplex. The four proteins were co-produced in *Sf*21 cells and isolated from the crude lysate by immobilized metal affinity chromatography (IMAC) via an N-terminal His_{10} tag on NOT1. Fractions containing the complex in sufficient purity were used for reconstitution with the other subcomplexes.

(c) Schematic representation of the CCR4a:CAF1 exonuclease heterodimer.

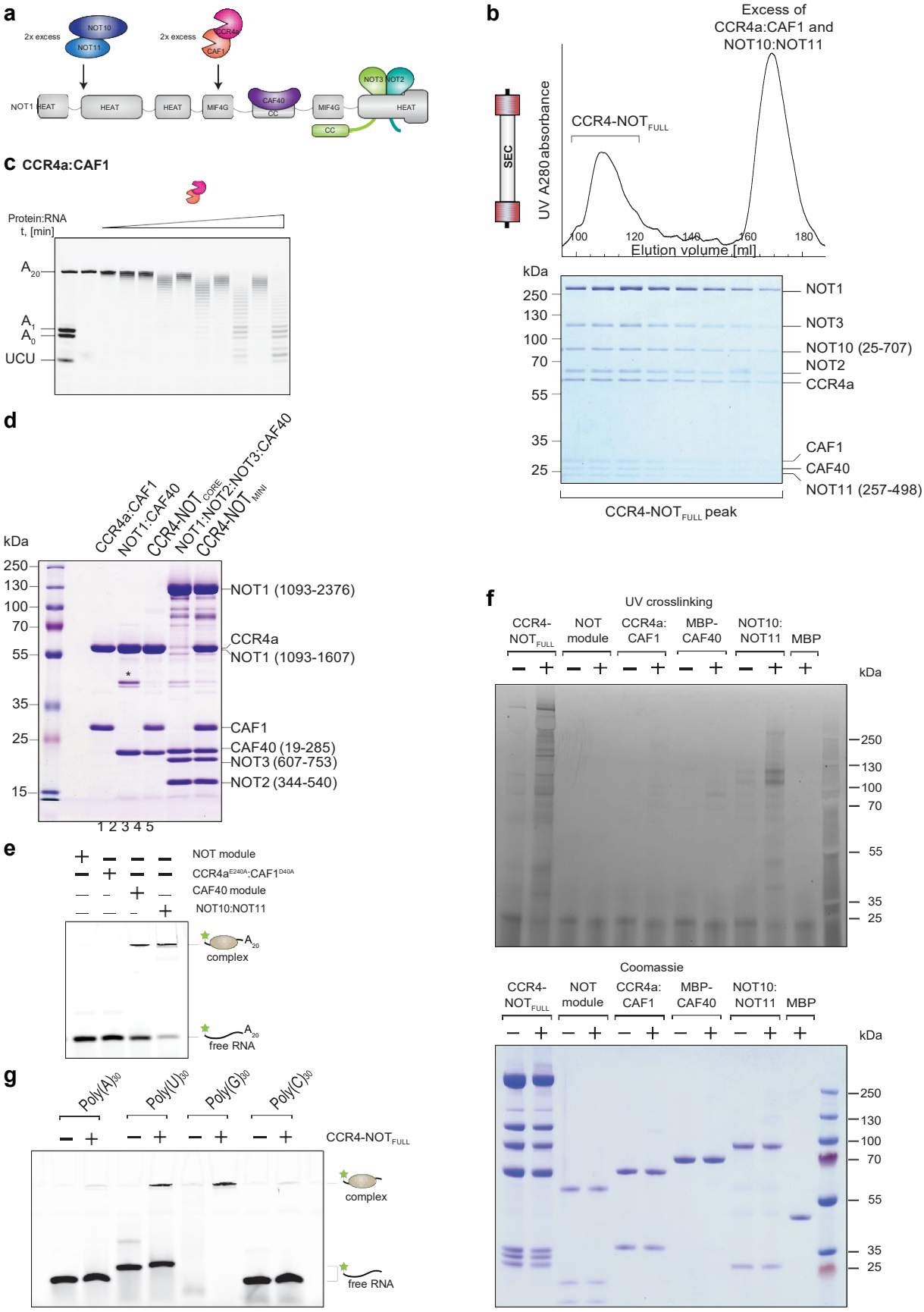
(d) Purification of the CCR4a:CAF1 heterodimer. Both proteins were co-produced in *E. coli*. The complex was isolated from the crude lysate by IMAC via His_6 tags on both proteins. Both tags were proteolytically removed and the complex separated from the tags by size exclusion chromatography (SEC). High-resolution anion exchange chromatography was used as the final purification step. Representative gels of all three purification steps illustrate the monitoring of the purification.

(e) Schematic representation of the NOT10:NOT11 heterodimer comprising NOT10 (residues 25-707) and NOT11 (residues 257-498).

(f) Purification of the NOT10:NOT11 heterodimer. Both proteins were co-produced in *E. coli*. The complex was isolated from the crude lysate by IMAC via a His_6 tag on NOT11. The His_6 tag from NOT11 and MBP tag from NOT10 were proteolytically removed and the heterodimer purified further by SEC. Representative gels of both purification steps illustrate the monitoring of the purification.

Source data for panels (b,d,f) are provided as a Source Data file.

Supplementary Figure 2



Supplementary Figure 2. Reconstitution of recombinant CCR4-NOT complexes.

(a) Schematic representation of reconstitution of the CCR4-NOT_{FULL} complex. The purified NOT1:NOT2:NOT3:CAF40 subcomplex (Supplementary Fig. 1a,b) was incubated with a two-fold molar excess of purified CCR4a:CAF1 (Supplementary Fig. 1c,d) and NOT10:NOT11 (Supplementary Fig. 1e,f) heterodimers.

(b) Purification of the reconstituted CCR4-NOT_{FULL} complex. After reconstitution (a), the assembled complex was separated from the excess of NOT10:NOT11 and CCR4a:CAF1 heterodimers by SEC. The purity and stoichiometry of the reconstituted CCR4-NOT_{FULL} was verified by SDS-PAGE.

(c) Titration experiment with 50 nM of the 7-mer-A_{20} RNA and the CCR4a:CAF1 dimer in the indicated molar ratios. Increasing the protein:RNA ratio leads to a progressive increase in deadenylation rate. For each ratio, two time points at 8 and 24 minutes, respectively, were analyzed.

(d) SDS-PAGE showing the purified recombinant CCR4a:CAF1 and NOT1:CAF40 complexes used for reconstitution of the CCR4-NOT_{CORE} complex, as well as the CCR4-NOT_{MINI} complex and the corresponding NOT1:NOT2:NOT3:CAF40 subcomplex used for reconstitution. Residual MBP co-purified with the NOT1:CAF40 complex is marked with an asterisk.

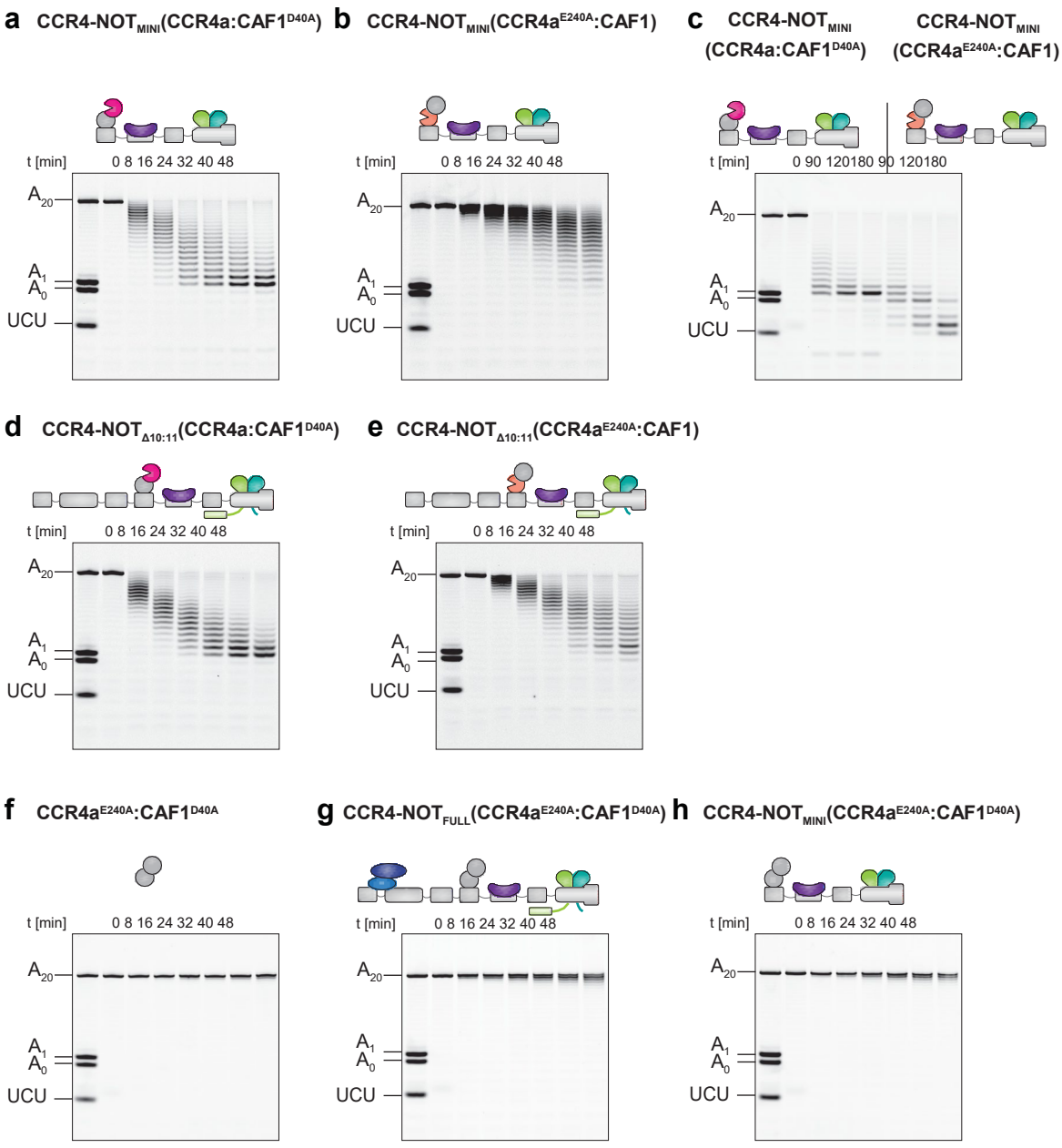
(e) Electrophoretic mobility shift assay (EMSA) with 100 nM of the 7-mer-A_{20} RNA and 25 µM CCR4-NOT subcomplexes. The CAF40 module and the NOT10:NOT11 heterodimer bind the RNA as visible from the shift of the RNA band to apparent higher molecular weight in the respective lane. The upshifted protein-RNA complex did not enter the gel.

(f) UV crosslinking experiment with 1 µM CCR4-NOT_{FULL} ($CCR4a^{E240A}$:$CAF1^{D40A}$) and subcomplexes, and 100 nM 6-FAM-labelled poly$(U)_{30}$ RNA analysed by denaturing SDS-PAGE. The lower panel shows the Coomassie-stained protein bands, and the upper panel the RNA substrate fluorescence of the same gel.

(g) EMSA with 100 nM of the four homopolymeric 30-mer RNAs and 1 µM CCR4-NOT_{FULL} ($CCR4a^{E240A}$:$CAF1^{D40A}$). The complex interacts with poly$(U)_{30}$ and poly$(G)_{30}$ substrates, and only weakly with poly$(A)_{30}$ and poly$(C)_{30}$.

Source data for panels (b-g) are provided as a Source Data file.

Supplementary Figure 3



Supplementary Figure 3. Comparison of deadenylation activities of CAF1 and CCR4a.

(a-b) Deadenylation assays with equimolar concentrations (50 nM) of the 7-mer-A_{20} RNA substrate and the CCR4-NOT_{MINI} complex containing inactivating mutations in either CAF1 (a; D40A mutation) or CCR4a (b; E240A mutation), respectively. The $CAF1^{D40A}$ mutation mildly decreases deadenylation rate and enhances selectivity for adenosine compared to wildtype (Fig. 2f), while the $CCR4a^{E240A}$ mutation drastically reduces the deadenylation rate.

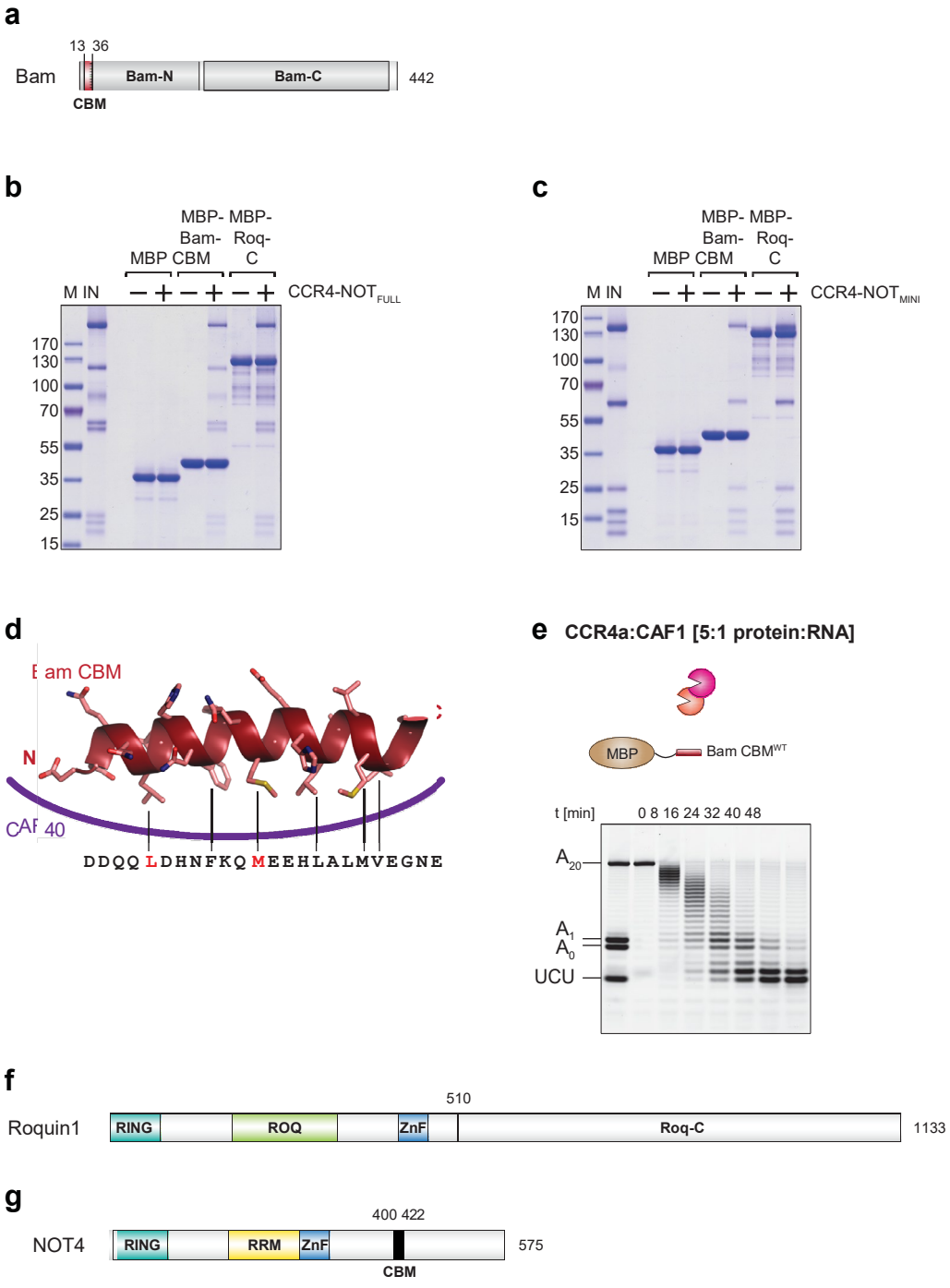
(c) Deadenylation assay with equimolar concentrations (50 nM) of the 7-mer-A_{20} RNA substrate and the CCR4-NOT_{MINI} complex containing inactivating mutations in either CAF1 (D40A mutation) or CCR4a (E240A mutation), respectively, incubated longer than in the experiments in (a) and (b). The complex containing the $CAF1^{D40A}$ mutation efficiently stops at A_1, whereas the complex with the $CCR4a^{E240A}$ mutation is able to degrade the substrate down to the UCU trinucleotide.

(d-e) Deadenylation assays with equimolar concentrations (50 nM) of 7-mer-A_{20} RNA and CCR4-$NOT_{\Delta 10:11}$ complexes which contain the same inactivating mutations in CAF1 (d) and CCR4a (e) as above. In the context of this complex, the $CAF1^{D40A}$ mutation mildly decreases deadenylation rate and leads to enhanced selectivity for adenosine (Fig. 2e), while the $CCR4a^{E240A}$ mutation strongly reduces deadenylation activity.

(f-h) Deadenylation assays with 50 nM 7-mer-A_{20} RNA and either 250 nM of the $CCR4a^{E240A}$:$CAF1^{D40A}$ double mutant nuclease dimer (f), or 50 nM of either CCR4-NOT_{FULL} (g) or CCR4-NOT_{MINI} (h) complexes reconstituted with the same double inactive nuclease dimer. In all three cases the complexes retain almost no catalytic activity.

Source data are provided as a Source Data file.

Supplementary Figure 4



Supplementary Figure 4. CCR4-NOT recruitment factors inhibit deadenylation.

(a) Schematic representation of *Drosophila* Bag-of-marbles (Bam). The protein consists of N- and C-terminal portions (Bam-N and Bam-C, respectively) and contains an N-terminal CAF40-binding motif (CBM)[20].

(b) MBP pulldown assay with either MBP or MBP-tagged Bam CBM or Roquin C-terminal fragment, respectively, and the CCR4-NOT_{FULL} complex. Both peptides specifically interact with CCR4-NOT_{FULL}.

(c) MBP pulldown assay with either MBP or MBP-tagged Bam CBM or Roquin C-terminal fragment, respectively, and the CCR4-NOT_{MINI} complex. Both peptides specifically interact with CCR4-NOT_{MINI}.

(d) Structure of the Bam CBM peptide bound to CAF40. The CBM folds into an amphipathic helix which binds to CAF40, here depicted by a purple line. The sequence of the CBM is shown below, with the two residues mutated in this study highlighted in red.

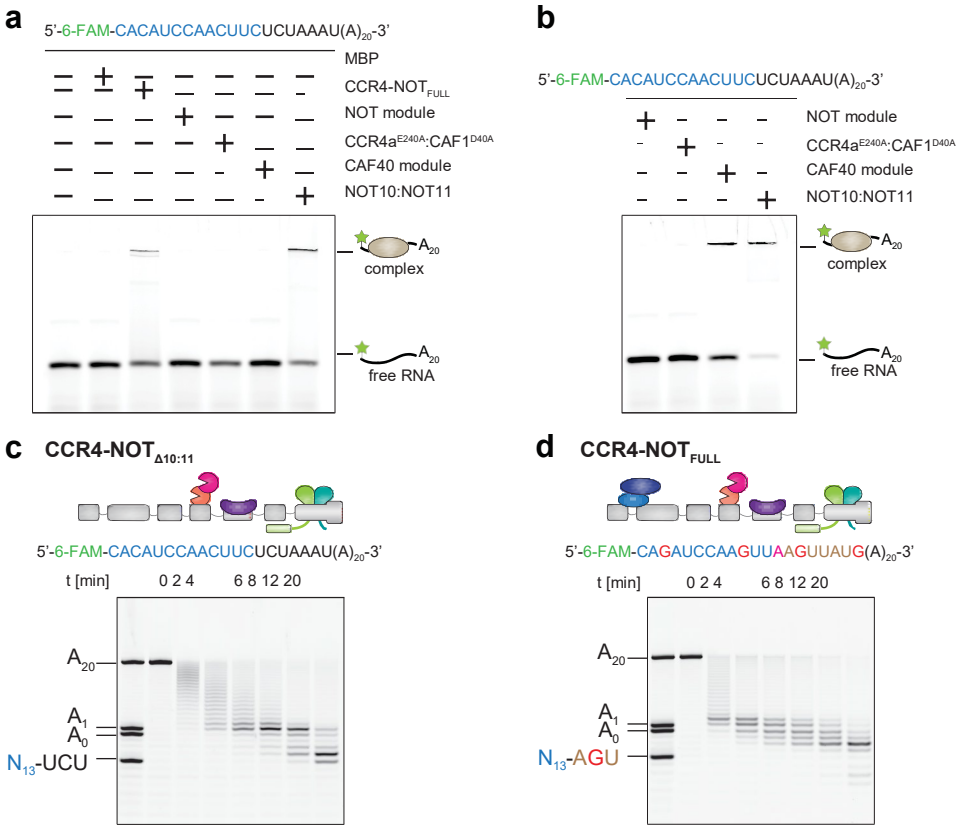
(e) Time course assay with 50 nM 7-mer-A_{20} RNA, 250 nM CCR4a:CAF1 heterodimer and 12.5 µM MBP-tagged Bam CBM^{WT}. The five-fold higher absolute concentration of the Bam peptide compared to the other experiments was chosen to achieve an equivalent 50-fold molar excess of the peptide over the nuclease dimer or nuclease-containing complex in all cases. The addition of the peptide does not affect the deadenylation activity of the exonuclease heterodimer. The equivalent experiment in the absence of the inhibitory Bam CBM^{WT} peptide is shown in Fig. 2c.

(f) Schematic representation of human Roquin1 consisting of a RING-type E3 ubiquitin ligase domain, a ROQ RNA-binding domain and a zinc finger domain (ZnF) in the N-terminal half of the protein, and an unstructured C-terminal tail which is known to interact with the NOT module and CAF40, even though no Bam-analogous CBM has been identified so far[13].

(g) Schematic representation of human NOT4 containing a RING-type E3 ubiquitin ligase domain, an RNA recognition motif domain (RRM) and a zinc finger domain (ZnF) in the N-terminal region of the protein. A CBM is present within the unstructured C-terminal region of the protein[50].

Source data for panels (b,c,e) are provided as a Source Data file.

Supplementary Figure 5



Supplementary Figure 5. CCR4-NOT efficiently binds and deadenylates RNAs with longer bodies.

(a) Electrophoretic mobility shift assay (EMSA) with 100 nM 13+7-mer-A_{20} RNA and 1 µM CCR4-NOT_{FULL} (CCR4a^{E240A}:CAF1^{D40A}) and subcomplexes. The CCR4-NOT_{FULL} (CCR4a^{E240A}:CAF1^{D40A}) complex and the NOT10:NOT11 heterodimer both bind the RNA as visible from the shift of the RNA band to apparent higher molecular weight in the respective lanes. The upshifted protein-RNA complexes did not enter the gel.

(b) Electrophoretic mobility shift assay (EMSA) with 100 nM 13+7-mer-A_{20} RNA and 25 µM CCR4-NOT subcomplexes. The CAF40 module and the NOT10:NOT11 heterodimer bind the RNA as visible from the shift of the RNA band to apparent higher molecular weight in the respective lane. The upshifted protein-RNA complex did not enter the gel.

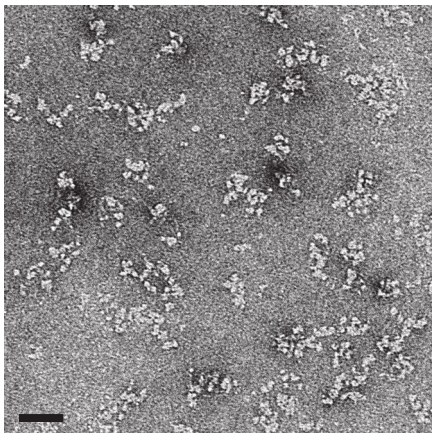
(c) Deadenylation assay with the CCR4-$NOT_{\Delta 10:11}$ complex and the 13+7-mer-A_{20} substrate (both 50 nM) with high resolution of the early time points.

(d) Deadenylation assay with an equimolar ratio (50 nM) of the CCR4-NOT_{FULL} complex and an RNA with a 20-mer body containing several guanosine nucleotides, with high resolution of the early time points.

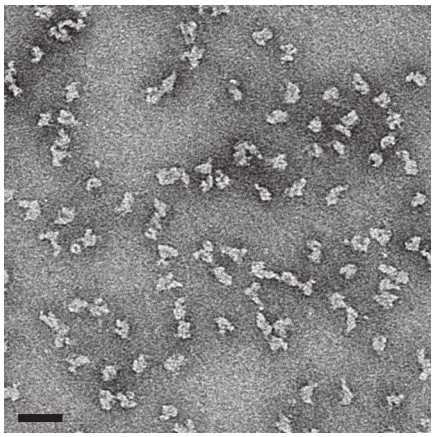
Source data are provided as a Source Data file.

Supplementary Figure 6

a Native complex



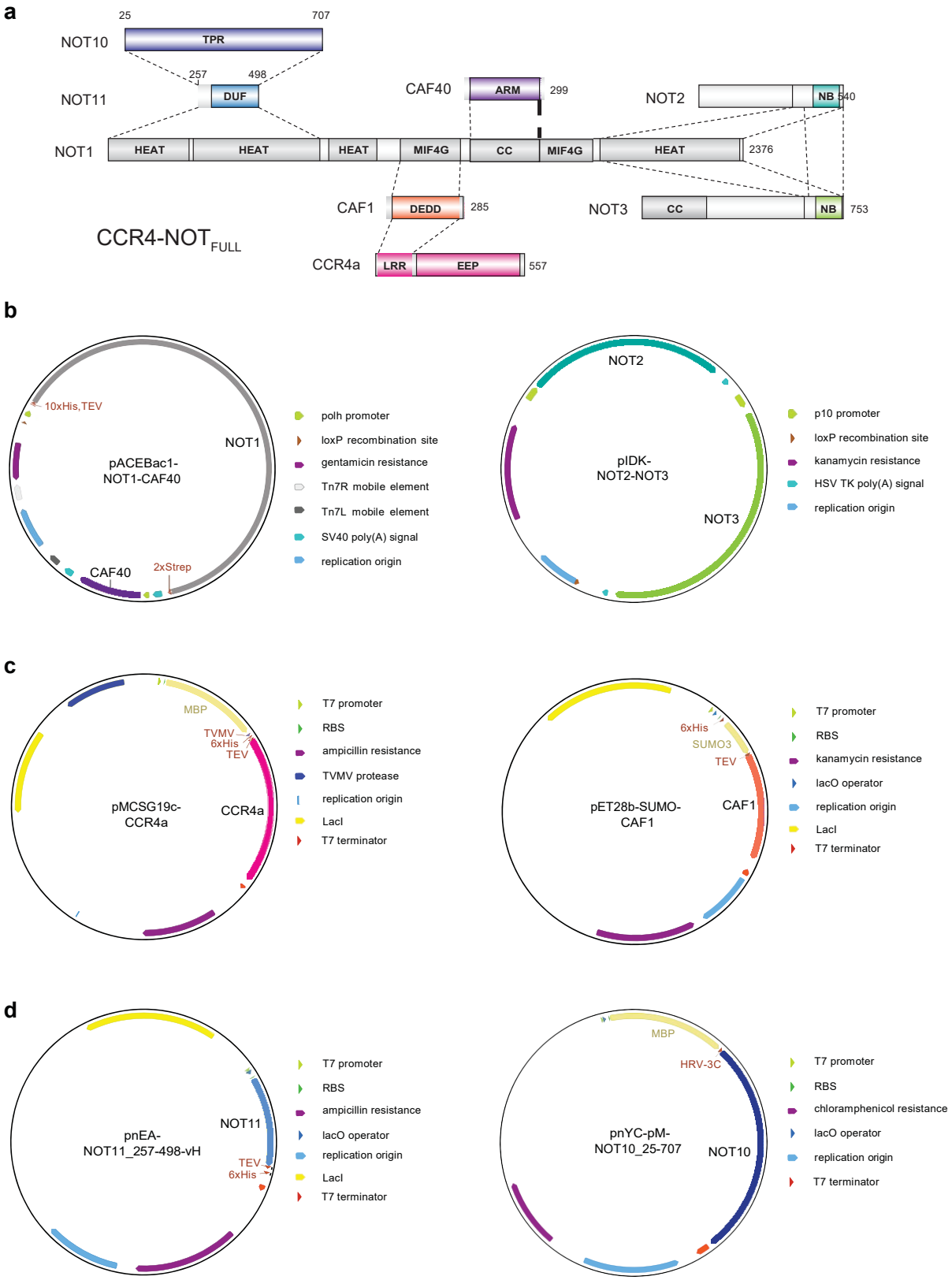
b Crosslinked complex



Supplementary Figure 6. Analysis of CCR4-NOT$_{FULL}$ by negative stain electron microscopy.

(a-b) Negative stain electron micrographs of reconstituted native (a) and crosslinked (b) CCR4-NOT$_{FULL}$, imaged on a Tecnai G Spirit electron microscope (Thermo Fisher Scientific) at a nominal magnification of 52,000. Scale bars: 50 nm. Crosslinking by glutaraldehyde during a sucrose gradient stabilizes the complex, but particles are still too heterogeneous for further characterization. Source data are provided as a Source Data file.

Supplementary Figure 7

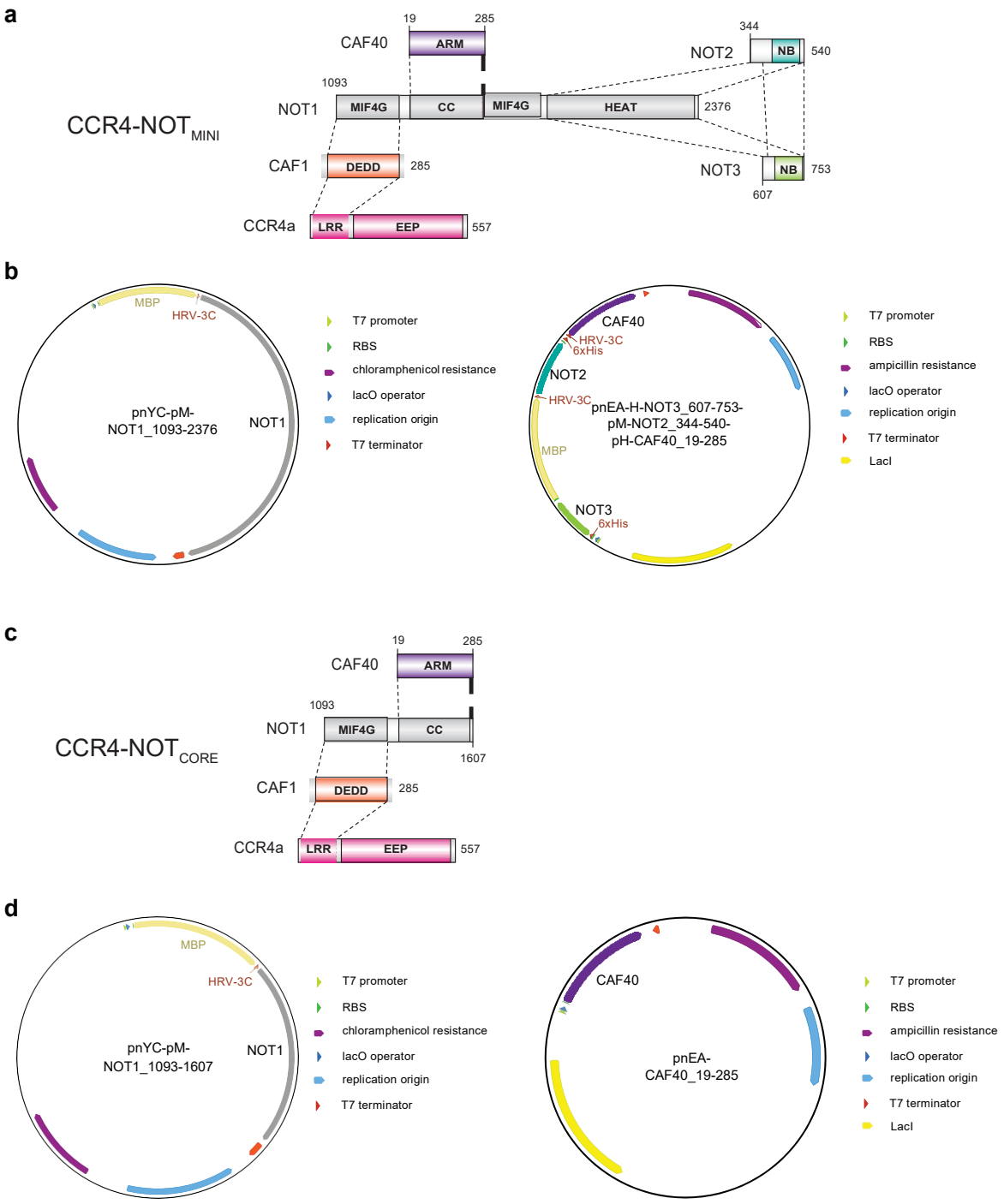


Supplementary Figure 7. Plasmids used for the production of the CCR4-NOT$_{FULL}$ complex.

(a) Schematic representation of the CCR4-NOT$_{FULL}$ complex in the same style as shown in Fig. 1a. All proteins were expressed as full-length constructs, apart from NOT10 (residues 25-707) and NOT11 (residues 257-498).

(b-d) Schematic representations of plasmids used for reconstitution of the CCR4-NOT$_{FULL}$ complex. pACEBac1-NOT1-CAF40 and pIDK-NOT2-NOT3 (b) were assembled using Cre-based recombination and used for generating a recombinant baculovirus for production of the NOT1:NOT2:NOT3:CAF40 subcomplex in *Sf*21 cells. *E. coli* were co-transformed with pMCS19c-CCR4a and pET28b-SUMO-CAF1 (c) for production of the CCR4a:CAF1 heterodimer, and pnEA-NOT11_257-498-vH and pnYC-pM-NOT10_25-707 (d) were co-transformed in *E. coli* for production of the NOT10:NOT11 heterodimer.

Supplementary Figure 8



Supplementary Figure 8. Plasmids used for the production of CCR4-NOT$_{MINI}$ and CCR4-NOT$_{CORE}$.

(a) Schematic representation of the CCR4-NOT$_{MINI}$ complex in the same style as shown in Fig. 1a. Construct boundaries of the subunits are indicated.

(b) Schematic representations of plasmids used for reconstitution of the CCR4-NOT$_{MINI}$ complex. *E. coli* were co-transformed with pnYC-pM-NOT1_1093-2376 and pnEA-H-NOT3_607-753-pM-NOT2_344-540-pH-CAF40_19-285 for production of the NOT1:NOT2:NOT3:CAF40 subcomplex.

(c) Schematic representation of the CCR4-NOT$_{CORE}$ complex in the same style as shown in Fig. 1a. The complex contained full-length CCR4a and CAF1, and NOT1 (residues 1093-1607) and CAF40 (residues 19-285)

(d) Schematic representations of plasmids used for reconstitution of the CCR4-NOT$_{CORE}$ complex. *E. coli* were co-transformed with pnYC-pM-NOT1_1093-1607 and pnEA-CAF40_19-285 for production of the NOT1:CAF40 subcomplex.