Molecular mechanisms of Ded1p phase separation

DISSERTATION ZUR ERLANGUNG DES

DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN

DER UNIVERSITÄT REGENSBURG

Vorgelegt von

Julian Hübner

Aus

Bad Soden am Taunus

Im Jahr

2023

Molecular mechanisms of Ded1p phase separation

DISSERTATION ZUR ERLANGUNG DES

DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)

DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN

DER UNIVERSITÄT REGENSBURG

Vorgelegt von

Julian Hübner

Aus

Bad Soden am Taunus

Im Jahr

2023

Das Promotionsgesuch wurde eingereicht am:

14.09.2023

Die Arbeit wurde angeleitet von:

Prof. Dr. Remco Sprangers

Unterschrift:

List of publications and manuscripts

Submitted / under revision

Ceciel Jegers, Titus M Franzmann, **Julian Hübner**, Jakob Schneider, Cedric Landerer, Sina Wittmann, Agnes Toth-Petroczy, Remco Sprangers, Anthony A Hyman, and Simon Alberti. A conserved and tunable mechanism for the temperature-controlled condensation of the translation factor Ded1p, *bioRxiv* (2022).

In preparation

Jan Philip Wurm, **Julian Hübner**, and Remco Sprangers. Systematic assessment of the conformational changes in DEAD-box RNA helicases.

Julian Hübner, Jan Philip Wurm, Valentin Thron, and Remco Sprangers. Tryptophan-Arginine sticker pairs determine Ded1p protein condensation.

The contributions of each author are indicated at the beginning of each chapter.

Contents

Abbreviations

1D, 2D, 3D	One-, two-, three-dimensional
A_{280}	Absorption at 280 nm
ADP	Adenosine diphosphate
ALS	Amyotrophic lateral sclerosis
AMP	Adenosine monophosphate
AMPPNP	Adenylyl imidodiphosphate
ATP	Adenosine triphosphate
ATP-γ-S	Adenosine 5'-(gamma-thiotriphosphate)
Cryo-EM	Cryogenic electron microscopy
c_{SAT}	Saturation concentration
CSP	Chemical shift perturbations
C. thermophilum	*Chaetomium thermophilum*
Ded1p	Definition of Essential Domain 1
ds	Double-stranded
DTT	Dithiothreitol
E. coli	*Escherichia coli*
eIF	Eukaryotic translation initiation factor
FID	Free induction decay
FUS	Fused in sarcoma
FRAP	Fluorescence recovery after photobleaching
GdmCl	Guanidinium hydrochloride
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HMQC	Heteronuclear multiple quantum coherence
H. sapiens	*Homo sapiens*
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IPTG	Isopropyl β-d-thiogalactopyranoside
LB	Lysogeny broth
LCR	Low complexity region
LCST	Lower critical solution temperature
LLPS	Liquid-liquid phase separation

m_7G	N7-methyl guanosine
MAGIC	Methyl assignment by graphic interference construct
mRNA	Messenger RNA
NAD^+/NADH	Nicotinamide adenine dinucleotide
Nickel-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD_{600}	Optical density at a wavelength of 600 nm
P-body	Processing body
PCR	Polymerase chain reaction
PDB	Protein data bank
PML	Promyelocytic leukemia
ppm	Parts per million
PRE	Paramagnetic relaxation enhancement
PTM	Post-translational modification
RecA	Recombinase A
RF	Radio frequency
RMSD	Root means square deviation
RNP	ribonucleoprotein
Rpm	Rounds per minute
S. cerevisiae	*Saccharomyces cerevisiae*
SEC	Size exclusion chromatography
SF	superfamily
SG	Stress granule
SOFAST	Band-selective optimized flip-angle short-transient
ss	Single-stranded
TDP-43	TAR DNA-binding protein-43
TEV	Tobacco etch virus
TROSY	Transverse relaxation-optimized spectroscopy
TSP	Trimethylsilylpropanoic acid
UCST	Upper critical solution temperature
UTR	Untranslated region

v/v	Volume per volume
w/v	Weight per volume
WT	Wild type

All amino acids were abbreviated with their one- or three-letter code. Nucleobases were abbreviated with their one-letter code.

Summary

Liquid-liquid phase separation (LLPS) describes a mechanism that results in the clustering of biomacromolecules, such as nucleic acids and proteins, into large microscopically visible foci. On a molecular level, LLPS is the result of a not fully characterized, redundant network of weak protein-protein and protein-RNA interactions. It is now widely understood that LLPS plays an important role in the formation of many membraneless organelles.

In this work, I used nuclear magnetic resonance (NMR) spectroscopy and biochemical assays to elucidate the intermolecular interactions that govern the LLPS of the yeast DEAD-box RNA helicase *Definition of Essential Domain 1* (Ded1p), which is a main component of stress granules (SG). The Ded1p protein consists of a highly conserved helicase core of two folded RecA domains that constitute the active site and additional N- and C-terminal disordered extensions.

Based on an *in vitro* LLPS assay, I established that LLPS of Ded1p is mainly mediated by its C-terminal extension. Using paramagnetic relaxation enhancement (PRE) and spin relaxation experiments, the molecular grammar that governs the driving force of LLPS was identified for the C-terminal region. Tryptophan and arginine residues were shown to be highly potent stickers that dynamically drive condensation of the C-terminal intrinsically disordered region (IDR) of Ded1p. My data also uncover a direct relationship between the mobility of individual amino acids and the importance of these residues for LLPS. Importantly, I established that these findings for the isolated IDR can be transferred to the LLPS behavior of the full-length protein. In addition, I identified interactions between the N- and C-terminal extensions and the folded helicase core that also contribute to LLPS.

In a second line of work, I used NMR titration experiments to study the interaction between the RecA core domains and their substrates adenosine triphosphate (ATP) and RNA. Based on five different DEAD-box helicases, I established a conserved energetic coupling between ATP and RNA binding, which is mediated by a conformational rearrangement of the RecA core domains. In the apo state, the RecA domains tumble independently, however, in the presence of both substrates, a transition to a rigid closed conformation takes place. The formation of this closed

helicase core state is required for helicase unwinding activity in all tested DEAD-box RNA helicases.

Taken together, this work highlights the importance of IDRs for Ded1p LLPS and increases our understanding of the molecular mechanisms of DEAD-box RNA helicases.

Zusammenfassung

Flüssig-flüssig-Phasentrennung (LLPS) ist ein Phänomen, bei dem sich Biomakromoleküle zu großen, mikroskopisch sichtbaren Tröpfchen zusammenlagern. Auf molekularer Ebene führt LLPS zur Bildung eines redundanten und bisher wenig charakterisierten Netzwerks, das aus schwachen Protein-Protein und Protein-RNA Wechselwirkungen besteht. Es ist inzwischen allgemein anerkannt, dass LLPS eine essentielle Rolle bei der Bildung von membranlosen Zellorganellen spielt.

In dieser Arbeit habe ich Kernspinresonanzspektroskopie (NMR) und biochemische Experimente genutzt, um die intermolekularen Wechselwirkungen zu untersuchen, die zu LLPS des Hefeproteins Ded1p führen, einer DEAD-box RNA Helikase, die ein Hauptbestandteil von Stressgranula ist. Ded1p besteht aus einem hochkonservierten Helikasekern, der von zwei gefalteten RecA-Domänen besteht, die das aktive Zentrum bilden, sowie N- und C-terminalen ungefaltete Erweiterungen.

Basierend auf *in vitro* LLPS-Experimenten konnte ich herausfinden, dass LLPS von Ded1p hauptsächlich durch die C-terminale Erweiterung angetrieben wird. Mit Hilfe von *paramagnetic relaxation enhancement* (PRE) und Spinrelaxationsexperimenten konnte die molekulare Grammatik, welche LLPS der C-terminalen Region beschreibt, aufgeklärt werden. Dabei konnte ich herausfinden, dass Tryptophan- und Argininreste hochpotente Sticker-Reste sind, welche die Bildung von Kondensaten aus der C-terminalen IDR von Ded1p dynamisch vorantreiben. Meine Daten belegen auch einen direkten Zusammenhang zwischen der Flexibilität einzelner Aminosäuren und der Bedeutung dieser Reste für LLPS. Ein weiteres wichtiges Ergebnis ist, dass sich das Phasentrennungsverhalten der isolierten C-terminalen IDR auch auf das Volllängeprotein Ded1p übertragen lässt. Darüber hinaus habe ich Wechselwirkungen zwischen den N- und C-terminalen Erweiterungen und dem gefalteten Helikasekern identifiziert, die ebenfalls zu LLPS beitragen.

In einer zweiten Studie untersuchte ich mit Hilfe von NMR-Titrationsexperimenten die Wechselwirkung zwischen den RecA-Kerndomänen und ihren Substraten ATP und RNA. Anhand von fünf verschiedenen DEAD-box RNA Helikasen konnte ich eine konservierte energetische Kopplung zwischen ATP- und RNA-Bindung nachweisen, die durch eine Konformationsänderung der RecA Kerndomänen vermittelt wird. Im apo-Zustand taumeln die RecA Domänen unabhängig voneinander, in Gegenwart

beider Substrate findet jedoch ein Übergang in eine starre geschlossene Konformation statt. Die Ausbildung dieses geschlossenen Helikase-Kernzustands ist bei allen untersuchten DEAD-box RNA Helikasen für die Helikaseaktivität notwendig.

Insgesamt unterstreicht diese Arbeit die Bedeutung intrinsisch ungeordneter Proteinregionen für LLPS von Ded1p und erweitert unser Verständnis der atomaren Mechanismen von DEAD-box RNA Helikasen.

Chapter 1: General introduction

1.1 The role of biological membranes and the formation of compartments in eukaryotic cells

The development of biological membranes was an essential step in the origin of life. One way cells achieved subcellular organization was by developing biological membranes that consist of a bilayer of lipid molecules in eukaryotic cells. These membranes also facilitated the formation of additional subcellular compartments known as cell organelles (1).

These so-called membrane-bound canonical cell organelles were first discovered in the 19th century using light microscopy. The nucleus, which carries the genetic information of eukaryotic cells, was the first organelle described by Brown in 1833 (2, 3). The phospholipid bilayer is impermeable to most biological molecules and gives rise to the interior and exterior of an organelle. An exchange of ions or macromolecules can only be achieved by highly specialized membrane transporters (4, 5). However, small molecules such as O_2 can still pass the membrane by diffusion (6). Each organelle has a specific cellular function, and the biological membrane allows an organelle the maintenance of the environment necessary for its specific proteins to perform their desired tasks (5). For example, the endoplasmic reticulum or Golgi apparatus distributes lipids, vesicles, and proteins to other organelles or to the cytoplasm, while in mitochondria the biological membrane maintains a proton gradient across the lipid bilayer, which is necessary for cellular respiration and the synthesis of ATP, the energy currency of life (7).

Taken together, the evolution of biological membranes provided life with a way to perform its diverse set of biological processes, and cell compartments allow eukaryotic cells to separate biological processes not only in space but also in time.

1.2 Formation of membrane-less organelles via liquid-liquid phase separation

1.2.1 Function of membrane-less cell organelles

Apart from the aforementioned membrane-bound cell organelles, another class of subcellular structures is found in the nucleoplasm and cytoplasm of cells. This second class of “organelles” is not enclosed by membranes but is formed via the condensation of specific proteins and nucleic acids. Nowadays, there is strong evidence that LLPS is the key principle behind the formation of most membrane-less organelles, which lack a physical barrier separating them from the surrounding medium (8). LLPS describes a phenomenon in which a homogeneous solution de-mixes into a dilute and a dense phase. The phase separation of biomacromolecules is driven by a complex and currently not fully characterized network of weak interactions between proteins and/or nucleic acids.

Cellular organelles in the nucleus and cytoplasm that lack a lipid bilayer have been intensively studied over the last decades (9, 10). Initially, it remained unclear whether membrane-less organelles served a purpose or whether they arose due to non-functional side effects. However, extensive studies have provided evidence that the formation of membrane-less organelles is fundamental for eukaryotic cells. In cells, these subcellular structures can form and condense by localizing specific proteins and nucleic acids in close proximity at very high concentrations, leading to condensation and separation from the surrounding medium (11). In this thesis, such membrane-less organelles will also referred to as (biological) condensates.

In fact, a membrane-less structure was also first described in the 1830s, when the nucleolus was observed in neuronal cells (12). Since then, more and more membrane-less organelles have been discovered in all eukaryotes, archaea, and also in some prokaryotes (**Figure 1.1**) (10, 13, 14).

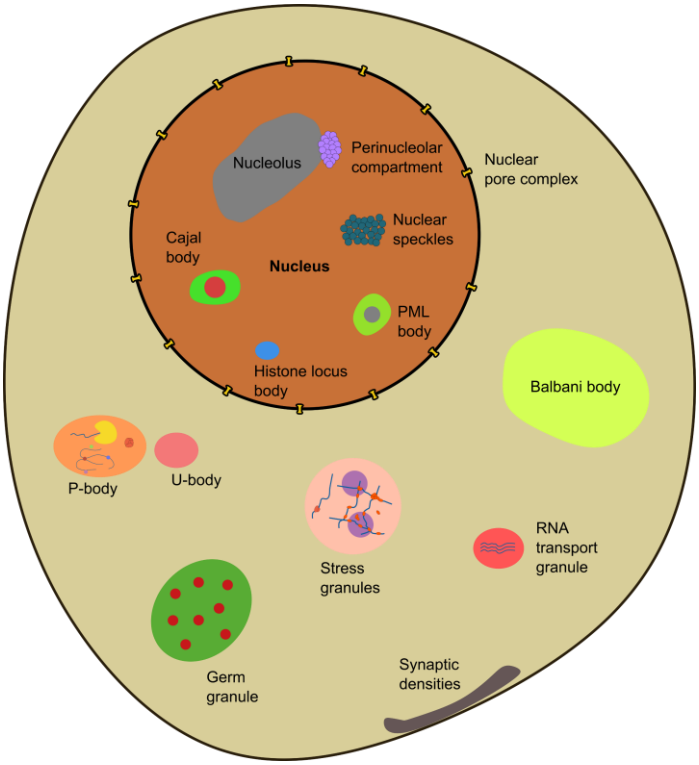


Figure 1.1: Schematic overview of biomolecular condensates in eukaryotic cells. Various membrane-less organelles can be found in the nucleus and cytoplasm of eukaryotic cells. While some condensates are present constitutively, such as P-bodies, other condensates, such as stress granules, form only under specific conditions. Note that not all shown condensates are found in every cell. Some condensates occur only in specific cells, such as Balbani bodies and germ granules, which are specific to germ cells. Apart from the nucleolus and Cajal bodies, other condensates with various functions are commonly found in the nucleus, e.g. nuclear speckles (15), perinucleolar compartments (16), promyelocytic leukemia (PML) bodies (17) or Histone locus bodies (18). Besides P-bodies and stress granules, Balbani bodies (19), U-bodies (20), germ granules (21), synaptic densities (22), and RNA transport granules (23) are commonly found in the cytoplasm. The figure was adapted from Banani *et al* and Hirose *et al* (10, 24).

Numerous membrane-less organelles play a role in the life cycle of RNAs, but there are other tasks that involve biological condensates. Condensation is also observed for signaling between cells, transcription, nuclear export and import mechanisms, assembly of cytoskeletal elements, and T cell activation (25–31). Condensation has also been observed to be important in initiating the DNA damage response (32).

A number of molecular mechanisms have been suggested for membrane-less condensates. Common examples of biomolecular condensates found in the nucleoplasm of eukaryotic cells are the nucleolus and Cajal bodies, both nuclear structures that are implicated in the biogenesis and regulation of ribonucleoproteins (RNPs) (**Figure 1.1**) (33–36). In the cytoplasm, processing bodies (P-body) and SGs

are frequently found and regulate the processing of messenger RNA (mRNA) and translation (37–40). P-bodies belong to a group of condensates that are constitutively present, while others, in this case SGs, only appear under certain environmental conditions such as stress (41).

Membrane-less organelles could serve as hubs to perform biochemical reactions effectively since enzymes, nucleic acids, and substrates are tightly packed and in close proximity. This also leads to an increase in local viscosity that is several orders of magnitude higher than in the surrounding medium, which could be beneficial for biochemical reactions (42, 43). Despite their differences in function, high-resolution microscopy has revealed that membrane-less organelles show similarities in assembly and shape. However, the size of the condensates can differ substantially from only a few nm up to several µm in oocytes of *Xenopus laevis* (44).

SGs are condensates that rapidly form in the cytoplasm when eukaryotic cells are exposed to environmental stress such as heat, starvation, UV-light, or certain chemicals (45). SGs recruit mRNAs that were prepared for translation together with other proteins of the translation initiation complex and thus inhibit protein synthesis by sequestering mRNA and translation factors (46). Typically, long mRNA transcripts encoding housekeeping proteins are recruited into SGs, whereas shorter and less complex mRNA transcripts encoding for proteins needed for stress response are not recruited into SGs (41, 47–50). When cellular stress ends, SGs disassemble and translation of stalled ribosome-mRNA complexes is reactivated (51). Cytoplasmic P-bodies contain parts of the mRNA degradation machinery, but it is not clear whether mRNA turnover is actually highly active in P-bodies (52–54). P-bodies and SGs are closely connected and even share a same set of proteins that can exchange between both condensates (55, 56).

1.2.2 Membrane-less cell organelles show characteristics of liquid-like behavior

Many studies revealed that membrane-less organelles exhibit a liquid-like character due to the fact that condensates can split or fuse, a behavior similar to an oil-water mixture (**Figure 1.2A**, **B**) (57, 58). In addition, condensates can contain subcompartments and two different condensates can fuse and co-exist (**Figure 1.2C**).

The different subcompartments of these multilayered liquids differ in their biophysical properties which arise from their varying components, and it has been postulated that sequential processes occur within a formed biological condensate (59, 60). Furthermore, particles can exchange their components with the surrounding nucleo- or cytoplasm, and components can move inside condensates, as it was shown in studies by following fluorescence recovery after photobleaching (FRAP) (**Figure 1.2D**). Still, it is proposed that some of the biological condensates do harbor solid-like cores that act as hubs for condensate assembly. These cores may transition from a liquid-like state upon condensate formation to a more solid-like state during condensate aging (61, 62).

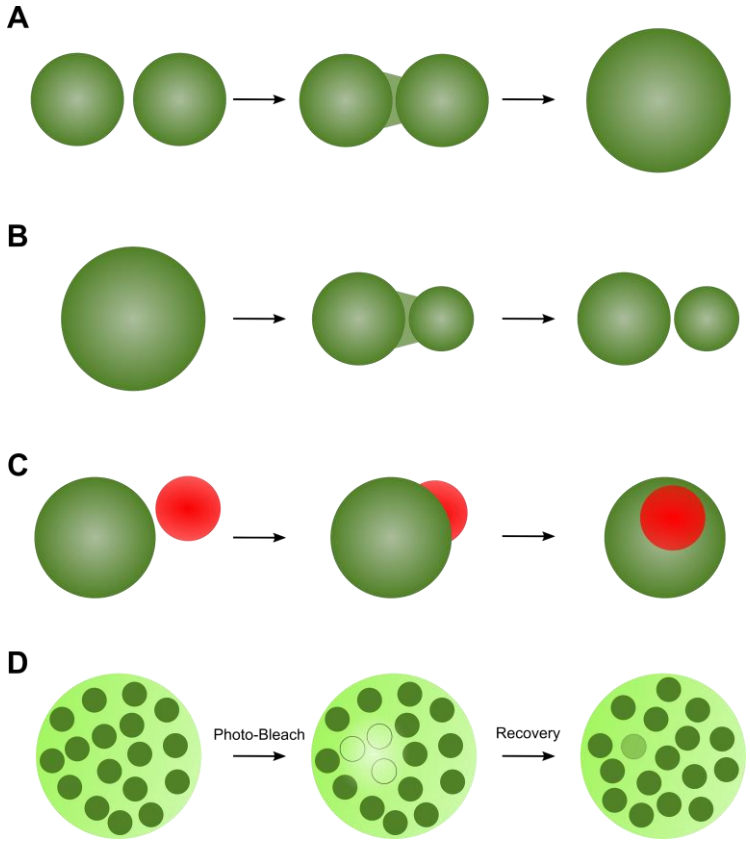


Figure 1.2: Membrane-less organelles show characteristics of liquid-like behavior. Phase-separated organelles that are made up of biomacromolecules exhibit liquid-like character. They are able to fuse (**A**) into larger droplets or can split (**B**) into several smaller droplets. Furthermore, subcompartments of two or more different condensates can fuse and co-exist (**C**). The movement of biomolecules inside a condensate can be examined by FRAP (**D**).

1.2.3 Molecular driving forces of liquid-liquid phase separation

For proteins, LLPS is often driven by motifs containing low complexity regions (LCR) that show little diversity in their amino acid composition (63). LCRs are usually IDRs that lack a unique three-dimensional structure and also contribute to LLPS (40, 64–66). It is essential to note here that LLPS is a fully reversible process and can therefore be easily discriminated from irreversible aggregation. In addition, it is now also understood that LLPS is actively regulated by the cell, e.g. by phosphorylation. Post-translational modifications (PTM) influence the participation of proteins in a condensate, as it is the case for the RNA helicase DDX4, which forms condensates upon methylation of specific arginine residues (67). PTMs provide cells with many ways of inducing and regulating condensate formation. How PTMs influence the LLPS network is further covered in a following chapter (Chapter 1.4.2).

The formation of membrane-less organelles that are not constitutively present can be induced by small changes in protein concentration, ionic strength, pH, or temperature (24, 68–70). LCRs or IDRs of LLPS-prone proteins can sense small changes in environmental conditions, making them a perfect tool for the formation of condensates.

Rapid changes in environmental conditions are common stress factors. With the rapid formation of membrane-less organelles, such as SGs, a cell can adapt quickly, often within seconds, to drastic environmental changes without having to resort to the biogenesis of proteins. However, the production of the machineries required to build and maintain the semi-permeable lipid bilayer is time-consuming and energy-demanding (71). Synthesizing additional membrane-bound organelles that are only needed during the stress response is energy and time consuming, neither of which is favorable when a cell is exposed to stress. Therefore, the formation of membrane-less organelles that are involved in stress response is favored.

1.2.4 Aging of biological condensates

In vitro studies have shown that some condensates change their physical properties over time (**Figure 1.3A, B**). This irreversible process is called maturation and leads to the formation of a gel-like structure that is no longer soluble (72). Studies of the aged and gel-like structures have been performed, but it is not yet fully understood what

structural rearrangements take place during the maturation process (73, 74). Nonetheless, maturation has been linked to neuronal diseases in living cells like amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (75–79). In order to prevent the harmful maturation process, cells must be capable to actively regulate LLPS and the formation of membrane-less organelles. Understanding the maturation process might help researchers to find future therapeutic targets that can be tackled (80). Besides neuronal diseases, aberrant LLPS is also widely associated with cancer formation (81).

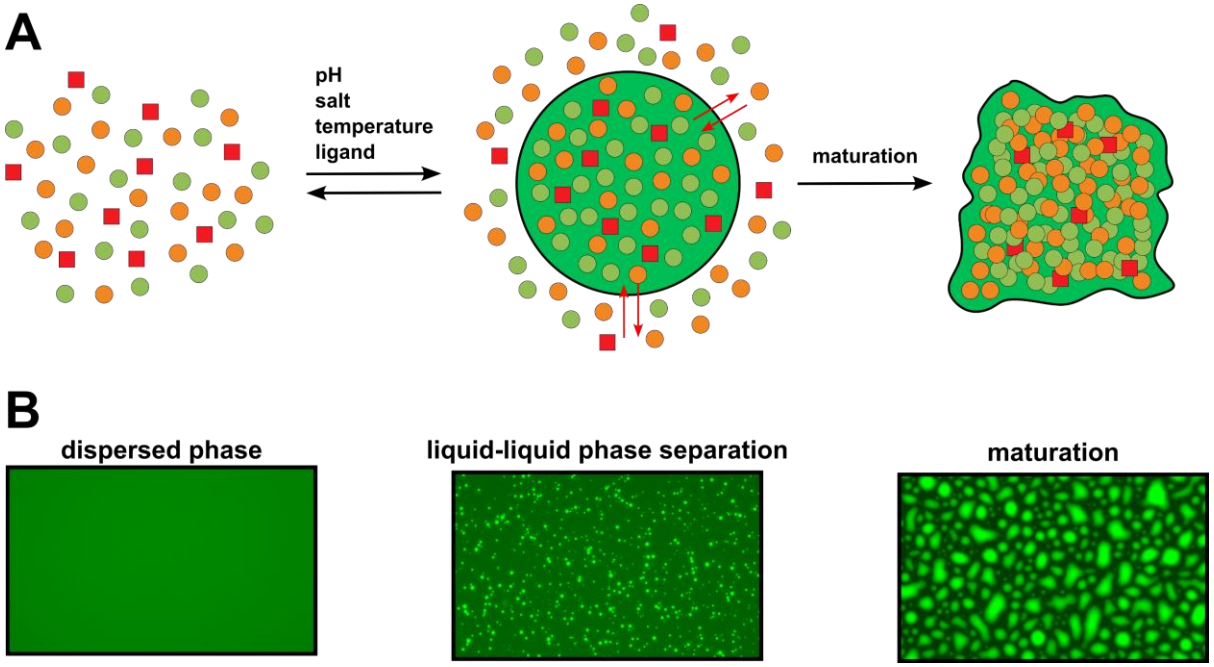


Figure 1.3: Formation and maturation of membrane-less organelles via liquid-liquid phase separation. (**A**) Condensates with a liquid-like nature form via regulated assembly of biomacromolecules like proteins and nucleic acids. Upon environmental stimuli, e.g. stress, the recruitment of specific molecules into an existing membrane-less organelle is triggered or new condensates are formed. The condensation is fully reversible, biomacromolecules can exchange with their environment, and organelles disassemble once the stress signal subsides in the case of SGs. However, it was shown in *in vitro* studies that some phase-separated proteins become gel-like over time. Therefore, cells must maintain regulatory mechanisms to prevent the maturation of phase-separated macromolecules that would otherwise be harmful to them. Shown here is the schematic formation of biological condensates via LLPS and maturation of droplets. Green and orange circles with the red squares represent proteins and nucleic acids. (**B**) Shown is an example of how a protein-RNA mixture can undergo phase separation. In-house fluorescence microscopy images of a well-mixed dispersed phase (left), LLPS state (middle), and the maturated state (right) of the protein Ded1p with fluorescent-labeled RNA. When protein is added, RNA is recruited into the droplets. With time, droplets made of RNA and protein start to mature and become gel-like.

1.2.5 Characterization of LLPS-prone proteins

In order to study the LLPS behavior of a protein *in vitro*, phase transition diagrams are determined under different conditions (8, 82). With phase diagrams, the phase separation behavior of a condensate is visualized and the co-existence region between the protein-light and protein-dense phase can be mapped (83). Usually, variables such as temperature, salt concentration, or pH are plotted against the biomolecule concentration for a protein to undergo LLPS in phase diagrams (**Figure 1.4**).

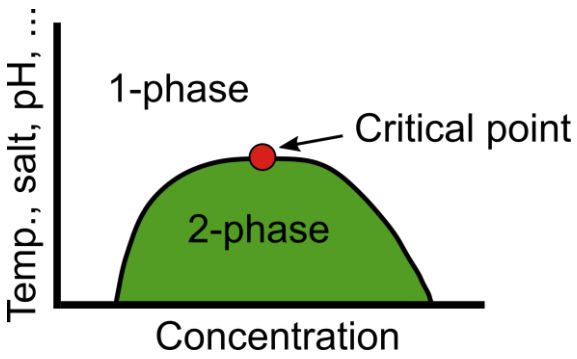


Figure 1.4: Schematic phase diagram. LLPS can be induced by several environmental changes such as temperature, salt concentration, pH, or by changes in the concentration of its molecular components such as proteins and nucleic acids. LLPS does not occur in the 1-phase regime. A transition from the homogeneous 1-phase into the 2-phase regime happens when a critical point of phase separation is reached for a given concentration (saturation concentration) of a biomolecule.

The resulting phase diagram can be fitted with different theoretical models (84). Based on the fit, the energy of mixing, a critical temperature in the case of temperature induced LLPS, or other thermodynamic parameters can be extracted. These parameters provide insights into the nature of involved interactions. To do so, different methods are often applied to determine the critical point of LLPS.

LLPS can be seen as turbidity with the naked eye or as droplets under the microscope (**Figure 1.3B**). The degree of LLPS can be quantified either by registering the absorbance of each mixture or by dynamic light scattering. Notably, the values can be biased since droplet size is not known and light scattering can also be caused by other particles and protein aggregation. A less error-prone approach to quantify the degree of LLPS is by determining the saturation concentration (c_{SAT}), the critical threshold concentration of a protein that is needed to induce phase separation under specific

conditions. This approach can then be complemented by single-molecule fluorescence resonance energy transfer and NMR spectroscopy to study interactions that lead to LLPS (79, 85, 86). For *in vivo* studies, proteins can be tagged with a fluorophore and fluorescence microscopy can be used to follow the recruitment of the protein of interest into a membrane-less organelle.

Taken together, there is growing evidence that the formation of most known membrane-less organelles takes place via LLPS. The formation is triggered by proteins that harbor multivalent interaction sites and/or nucleic acids. A good way to obtain insights into how phase separation is modulated and whether LLPS is observable under physiological conditions is by generating phase diagrams (87). LLPS might become a new fundamental principle in cell biology, but the underlying molecular basis of LLPS remains to be elucidated. However, how compartments select for certain macromolecular components remains elusive.

1.3 Driving forces that lead to liquid-liquid phase separation

LLPS is an energetically favorable and fully reversible process that leads to the transformation from a one-phase to a two-phase regime, resulting in a protein-dense and a protein-light phase. To achieve phase separation, the components of a membrane-less organelle must have a higher affinity to each other than they have with other molecules or the surrounding solvent, i.e. the Gibbs free energy of the system must be lower after phase separation. This is achieved by changing either the enthalpy or the entropy of a system. The enthalpic component of the free energy consists of all interaction potentials between the biomacromolecules and the solvent molecules and also between solvent-solvent and biomolecule-biomolecule interactions. On the other hand, the entropic component of the free energy consists of the conformational degrees of freedom of the biomacromolecules and the solvent. Thus, one can distinguish between enthalpy- and entropy-driven LLPS. The free energy of mixing ΔG_m is described by the following (Equation 1).

Equation 1 $$\Delta G_m = \Delta H_m - T \Delta S_m$$

where ΔHm and ΔSm are the enthalpy and entropy of mixing and T is the absolute temperature (88).

Studying the process that drives LLPS *in vivo* for single biomacromolecules and how each of them contributes to LLPS is challenging. Typically, a biological condensate consists of many RNAs and proteins. In contrast, isolated components can be analyzed with high precision *in vitro* and LLPS can be modeled (89). In condensates, nucleic acids and proteins form a fuzzy network via intermolecular protein-protein or protein-nucleic acid interactions. Proteins participating in this network usually contain two archetypes that promote network formation. First, folded protein domains that interact with other proteins or harbor binding sites for RNA (**Figure 1.5A**) (11). Second, long disordered protein regions or proteins that lack three-dimensional structure and have multiple binding sites for other biomacromolecules (**Figure 1.5B**) (67, 90).

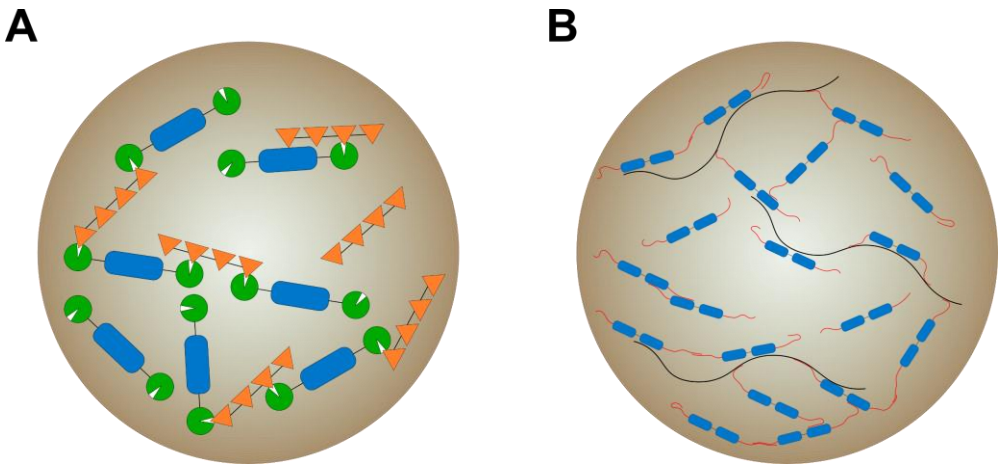


Figure 1.5: Network formation that drives condensation via liquid-liquid phase separation. (**A**) In a heterotypic system, weak and transient interactions between folded proteins with multivalent motifs cause LLPS. Such proteins contain repeats of weakly binding interaction domains. This is illustrated by the interaction between triangles and cycles, both representing folded domains in proteins. (**B**) Proteins harboring unfolded regions are also key players and their interactions cause LLPS. Often, interactions between IDRs are sufficient in order to drive LLPS, but RNA also contributes to this network. RNA is represented by a long, black, curved line that interacts with a multi-domain protein that harbors IDRs. The figure was adapted from Shin and Brangwynne (78).

When multiple motifs are participating in LLPS and multiple binding sites exist between macromolecules, a system is referred to as heterotypic (**Figure 1.5A**). In these systems, interactions often involve folded domains with motifs that mediate RNA or protein binding, but interactions can also be driven by disordered regions (91).

In vitro studies have shown that isolated IDRs are able to undergo LLPS without the requirement of additional components (92–94). For IDRs, phase separation is driven

by their multivalency. A “sticker-and-spacer” model is used as a concept for the phase separation behavior of intrinsically disordered proteins (IDPs)/IDRs (**Figure 1.6A**) (94–99). In this model, intra- and intermolecular electrostatic interactions, cation-π stacking, dipolar interactions, and π-π stacking drive enthalpically favored condensation (**Figure 1.6B-E**) (8, 100, 101). Stickers are residues that form non-covalent cross-links with other stickers and thereby contribute to LLPS. In particular, aromatic residues and arginines have been identified to serve as stickers that promote oligomer formation by forming multivalent interactions (102). Spacers are residues that do not contribute to LLPS but determine the position of stickers and in that way influence the driving forces of LLPS. Therefore, the distance between stickers created by spacers allows or suppresses the formation of cross-links. Sticker-sticker interactions are mainly influenced by the valence and patterning of residues along the sequence (94). Examples of this concept will be further described in the following chapter (Chapter 1.4.1).

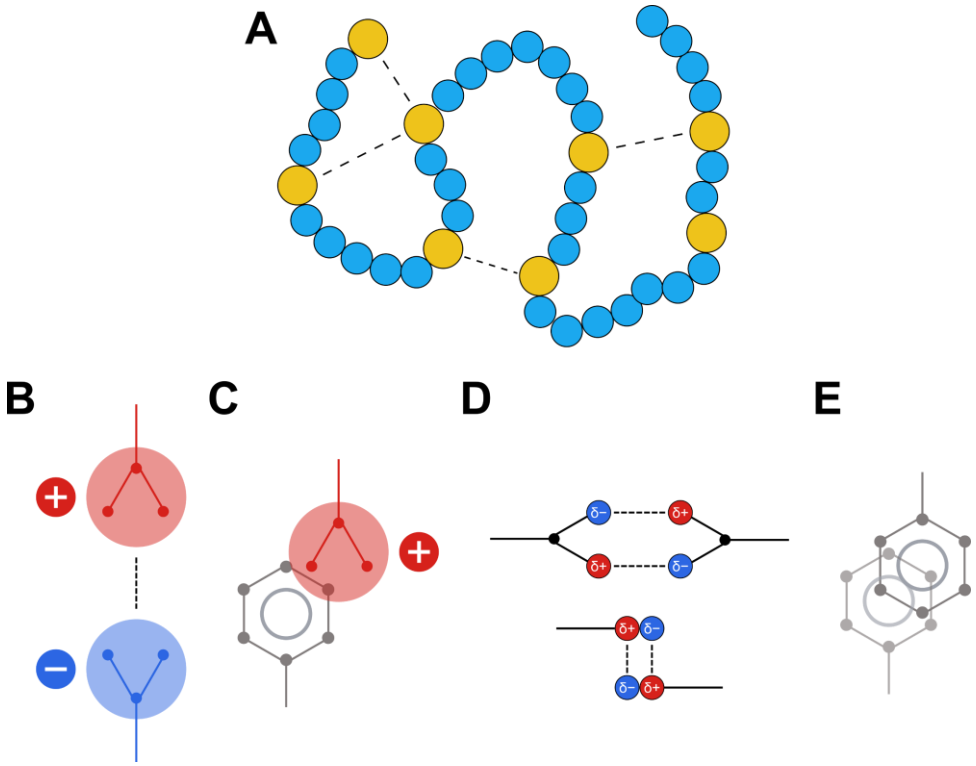


Figure 1.6: Liquid-liquid phase separation of intrinsically disordered regions as described by the sticker- and spacer-model. (**A**) Stickers (yellow circles) form non-covalent cross-links with other sticker residues. Non-covalent interactions are represented by dotted lines. Stickers are bridged by spacers (blue circles). Spacer residues do not contribute to LLPS but define the position of stickers. (**B-E**): Schematic representation of intra- and intermolecular interactions between sticker residues. Non-covalent cross-links are mediated by the following interactions: (**B**) Electrostatic; (**C**) cation-π; (**D**) dipole-dipole; (**E**) π-π stacking. The figure was adapted from Brangwynne *et al* (101).

As an example, for phase separation that is induced by a change in temperature, the molecular interactions that drive enthalpic LLPS become stronger as the temperature decreases (**Figure 1.7**). For such a system, an upper critical solution temperature (UCST) transition can be defined. Contrary to enthalpy-driven phase separation, LLPS at high temperature has also been observed for peptides with high hydrophobicity when less charged residues are present (88). An example of this are elastin-like peptides that contain a high fraction of hydrophobic amino acids which results in de-mixing at high temperature (103). Here, the de-mixing is caused by an increase in entropy due to the release of water molecules from the hydration shell of the side chains of hydrophobic residues (104). For this to occur, both the content but also the patterning of hydrophobic residues in a protein sequence must be finetuned. Hydrophobic compaction can be seen as an example of de-mixing at high temperature. In the case of LLPS, hydrophobic compaction has an inverse temperature dependency, meaning that LLPS is favored with increasing temperature. This so-called lower critical solution temperature (LCST) transition is entropic in nature (**Figure 1.7**). Such effects have been well characterized in studies of synthetic polymers (105, 106). Interestingly, IDPs can be designed in a way that they either exhibit LCST or UCST based on the replacement of certain residues (105). While many biological IDPs have a UCST phase behavior, LCST behavior cannot be excluded (107).

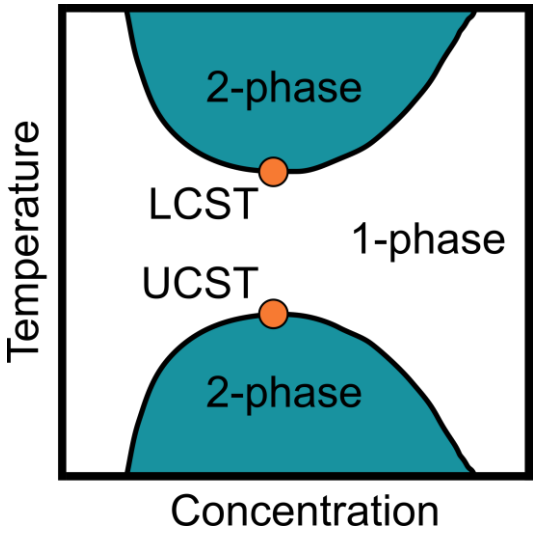


Figure 1.7: Simplified schematic liquid-liquid phase separation diagram. LLPS diagram of a biomolecule with LCST and UCST, respectively. LCST is mainly driven by entropic effects. LLPS is observed with increasing temperature. For a biomolecule with UCST, enthalpic effects are predominant. Here, electrostatic interactions, cation-π stacking, dipole-dipole interactions, and π-π stacking lead to phase separation at lower temperatures.

In brief, LLPS is largely driven by the association of biomolecules that is mediated by a multitude of weak interactions (100, 108, 109).

1.4 The role of molecular interactions in phase separation of intrinsically disordered proteins

1.4.1 Intrinsically disordered protein regions promote LLPS

A common feature of RNA-binding proteins that participate in LLPS networks is that they often share large patches of IDRs or are fully disordered (110, 111). IDPs are highly abundant in nature and often contain an amino acid sequence of low complexity. The abundance of disordered protein regions increases the more complex an organism gets (112). Almost one third of all amino acids in the human proteome are part of IDRs or IDPs (113). Even though IDRs and IDPs lack a rigid three-dimensional structure, some of them can be functional and their structural flexibility allows them to form multivalent interaction sites (90, 114, 115). For decades, the important biological functions of IDPs have been neglected when proteins were studied, and it was a paradigm that the specific functionality of proteins is conveyed only by their three-dimensional structure. The concept that a folded protein structure is necessary for its biological function became outdated with more and more studies that revealed that IDPs can complement the function of ordered proteins (116). IDPs are of highly dynamic nature and do not adopt a single structure in solution, but can be represented by a conformational ensemble (117). Being unstructured allows IDPs to bind multiple targets by providing multiple interaction motifs, and thereby IDPs function as hubs in protein-protein and protein-RNA interaction networks (118, 119). On the other hand, some protein regions contain only a small subgroup of amino acids and they often consist of highly repetitive sequences that are enriched in polar, aromatic, and charged residues (120). IDPs and LCRs play a role in various functions and are important components of the cellular signaling machinery. They are involved in most regulatory functions in the cell, such as transcription, translation, and the cell cycle (121, 122).

In the following, several examples of how IDRs drive condensation are highlighted for which it was shown that IDRs drive condensation. Aromatic side chains were identified to play an important role in the interplay between IDRs. Cation-π interactions via side

chain stacking of arginines and phenylalanines were shown to be the major driving force that causes LLPS for the N-terminal IDR of the DEAD-box RNA helicase DDX4 (108), the IDR of isoform A of human hnRNPA1 (95, 123), and NMR revealed that interactions between arginines and aromatic side chains contribute to LLPS as well as the condensed phase of CAPRIN1 (124).

A substantial amount of research has been performed on the IDR (residue 1-214) of fused in sarcoma (FUS) and its LLPS behavior (32, 92, 94, 109). While the wild-type (WT) IDR can undergo LLPS in isolation, a mutant in which all aromatic residues were substituted with non-aromatic amino acids no longer underwent phase separation. On the other hand, increasing the aromatic content in FUS increased the LLPS propensity. Interactions involving tyrosines or phenylalanines, which are common stickers in the IDRs of RNA-binding proteins, are not energetically equivalent. π-π interactions via Tyr stacking were found to be stronger than interactions of Phe side chains (95).

Interestingly, the contribution of tryptophan to LLPS remains elusive and only a few examples are found in the literature, which are described in the following. In the case of the condensate forming C-terminal IDR of the TAR DNA-binding protein (TDP-43), no prominent LLPS motif is present. Still, a critical feature of this IDR is that it harbors a central alpha-helical element containing a tryptophan residue that is critical for LLPS (125). The substitution of this tryptophan with glycine was found to abolish phase separation (126). In total, the C-terminal IDR contains three tryptophan residues that are the most important elements for LLPS, while other aromatic residues contribute to the degree of LLPS to a lesser extent (125). The helical region of the C-terminal IDR of TDP-43 harbors several ALS-related mutations, and LLPS was disrupted when these disease-causing mutations were introduced (127). Thus, this observation establishes a potential link between ALS and misregulation of LLPS.

In a study of phase separation of resilin-like polypeptides, tryptophan residues were found to be more potent in forming LLPS-induced contacts than tyrosine or phenylalanine residues (128). Similar to tyrosine, tryptophan was also found to participate in intermolecular interactions leading to LLPS of designed minimalistic peptides (129).

Aromatic residues are not the only key stickers that were identified in IDRs. It was also shown that arginine residues act as key players in the formation of phase separation.

In studies of FUS, the N-terminal IDR of DDX4, and LAF-1, the substitution of arginines with lysines led to a significant reduction of the LLPS propensity which further underscores the importance of π-related interactions in the formation of LLPS (108, 130, 131). Increasing the content of arginine resulted in a higher degree of LLPS, similar to the behavior observed for the content of aromatic residues described for FUS. The number of stickers, in the above examples of aromatic and arginine residues, is proportional to the c_{SAT} needed to undergo LLPS. Lysine is still capable of forming cation-π interactions, but arginine is capable to form both cation-π and π-π interactions, making it a more potent sticker residue (70, 100). The terminal guanidinium group of arginine, a non-aromatic sp^2-hybridized residue, has quasi-aromatic properties and thereby allows for π-π interactions (132). On the other hand, the ammonium group of the lysine side chain is a fully hydrated spherical ion (133). Pairs of ammonium groups formed by lysine-lysine interactions are less stable in water compared to arginine-arginine interactions, making them unfavorable (134). LLPS caused by arginines can be independent of aromatic residues, and due to its nature, arginine-rich peptides, but not lysine-rich peptides demonstrate LLPS at high salt concentrations (133). In addition to their role in IDRs for LLPS, RNA binding is mostly facilitated by positively charged residues such as lysine and arginine.

1.4.2 Regulation of LLPS via post-translational modifications

With PTMs of RNPs and especially IDPs, cells can directly influence LLPS. Long-range interactions mediated by IDPs can be directly enhanced or weakened by PTMs. Phosphorylation or methylation of arginines is a major features in regulating LLPS behavior (67, 93, 135, 136).

It has also been shown that the degree of phosphorylation of tyrosine residues on the disordered cytoplasmic tail of nephrin regulates the LLPS transition of the nephrin/NCK/N-WASP system (11, 137). Other PTMs that modulate LLPS have been described in the literature, e.g. NEDD8 conjugation and SUMOylation (138, 139). Also, binding of ubiquitin and poly(ADP-ribose) are effecting LLPS (140–142). Taken together, PTMs have been demonstrated to be an effective regulation mechanism for LLPS formation. Dysregulation of PTMs might also lead to the formation of amyloid-like protein aggregates that are implicated in neurodegenerative diseases (143).

In summary, IDRs provide a way for proteins to sense small changes in ionic strength, pH, or temperature. Such small changes are usually sufficient to trigger droplet formation, highlighting the important role of disordered protein regions and disordered proteins in the regulation of LLPS. Still, further studies of IDRs are needed to improve our knowledge of phase separation.

1.5 DEAD-box RNA helicase and the yeast helicase Ded1p

1.5.1 DEAD-box RNA helicases: an overview

For almost all cellular processes that involve nucleic acids, helicases are essential (144). This is especially the case for RNAs, which are prone to form stable, misfolded structures that prevent function or are even harmful (145). During the catalytic cycle, helicases make use of the energy stored in nucleoside triphosphates to catalyze separation of double-stranded (ds) RNA or DNA or to resolve misfolded structures or protein-nucleic acid complexes (144).

Helicases are found in all three domains of life and can be divided into six superfamilies (SF) (146, 147). The largest superfamilies are SF1 and SF2, which share the same conserved motifs (144, 148) and are functional as monomers and dimers (146, 149). The SF2 superfamily can be further divided into ten subfamilies, of which DEAD-box proteins are the largest subgroup (150). Their name derives from the conserved sequence motif Asp-Glu-Ala-Asp (DEAD), which is responsible for ATP hydrolysis. It is also characteristic for this family that the central catalytic core is composed of at least 12 conserved motifs (**Figure 1.8A**) (151).

On a structural level, DEAD-box RNA helicases consist of a highly conserved helicase core consisting of two folded domains (**Figure 1.8B**) (152, 153). Structurally, both domains resemble the fold of the bacterial recombination protein named recombinase A (RecA) (146). Both RecA domains are covalently connected via a linker (154). Sites for RNA- and ATP binding are shared by both RecA domains.

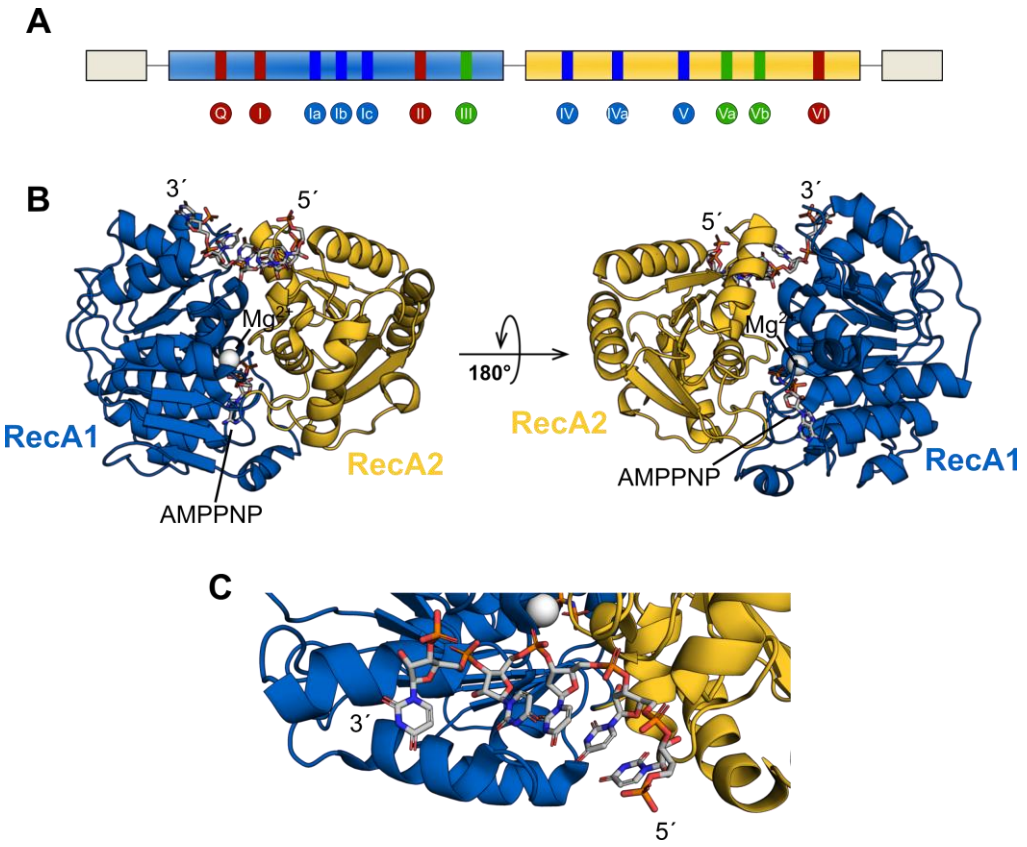


Figure 1.8: Domain architecture and structure of DEAD-box RNA helicases. (**A**) The light blue box represents the RecA1 domain, while the yellow box represents the RecA2 domain. Both domains form the folded helicase core. Grey boxes flanking the core domains represent the N- and C-terminal IDRs. Red, blue and green rectangles represent conserved motifs in the folded core. Red: motifs are responsible for ATP binding and hydrolysis; green: motifs coordinate nucleic acid and nucleotide binding; blue: these motifs are located in the nucleic acid binding site. The image was adapted from Putnam and Jankowsky (155). (**B**) Crystal structure of the helicase core of the DEAD-box protein Vasa of *Drosophila melanogaster* in complex with RNA and non-hydrolysable AMPPNP (PDB ID: 2DB3) (156). Vasa is illustrated as cartoon. The first RecA domain (RecA1) is shown in blue while the second RecA domain (RecA2) is represented in yellow. RNA and AMPPNP are shown in stick representation. Mg^{2+} is depicted as a white sphere. (**C**) Close-up view of RNA-Vasa contacts. Single-stranded RNA is bound to Vasa via the backbone. Bases point away from the helicase.

In addition to the folded domains, the helicase core is often flanked by N- and C-terminal intrinsically disordered extensions (157, 158). Both disordered extensions contain interaction sites for other proteins and, compared to the helicase core, their sequence shows a very high variability and they differ in length (159). In addition, both disordered extensions can contain DNA- or RNA-binding domains, binding sites for proteins, or oligomerization sites (160, 161). Beside the highly conserved helicase core that consists of the two RecA domains, many DEAD-box RNA helicases contain additional folded domains that contribute to their desired function which often harbor

additional RNA binding sites. Examples for such helicases are DbpA or the C-terminal domain of Mss116p (162, 163).

DEAD-box enzymes play an important role for the RNA metabolism such as unwinding of RNA duplexes or remodeling of RNA-protein complexes in an ATP-dependent manner (164). In addition to their unwinding function, DEAD-box RNA helicases can also participate in strand annealing and they can bind to RNA, which gives them a function as RNA clamps. DEAD-box RNA helicases are non-processive and show no specificity towards any RNA sequence (165).

In the apo state, both RecA domains do not interact with one another and instead tumble independently (155, 158). This state is referred to as the open state. Upon binding of RNA and ATP, a closed state is formed in which both RecA domains undergo a structural rearrangement and come together to form a bipartite RNA binding cleft (154, 166). For this RNA-bound state, the first structural information was obtained when non-hydrolysable ATP analogues such as adenylyl imidodiphosphate (AMPPNP) were used (150, 167). However, ATP hydrolysis is not needed for RNA unwinding but for the release of bound RNA from the helicase (168).

Still, unpaired RNA regions are often required for effective binding to the helicase. Local strand separation starts with the core of DEAD-box proteins binding directly to five nucleotides of an RNA strand (**Figure 1.8C**) (162, 169, 170). This interaction takes place at the phosphate backbone of the RNA, which explains why unwinding is independent of the nucleotide sequence. The unique substrate specificity of DEAD-box RNA helicases is more likely specified by their N- and C-terminal extensions.

The unwinding cycle of DEAD-box RNA helicases is represented in **Figure 1.9**. The interaction causes a bending of one RNA strand, resulting in a “kink” structure. Consequently, a few base pairs are separated from the duplex. In this way, the remaining duplex is destabilized, and the other base pairs dissociate as well. The helicase has completed a full cycle when RNA, adenosine diphosphate (ADP), and Pi are released, after which the helicase is then prepared for another unwinding cycle. However, not all duplexes can be unwound, and most DEAD-box RNA helicases can only unwind RNA duplexes that contain less than 10-12 base pairs (171). The rate of unwinding becomes slower the longer and more stable a duplex is (172). Despite the

insights into the unwinding cycle (**Figure 1.9**), the details of how DEAD-box RNA helicases perform strand separation are structurally not fully understood.

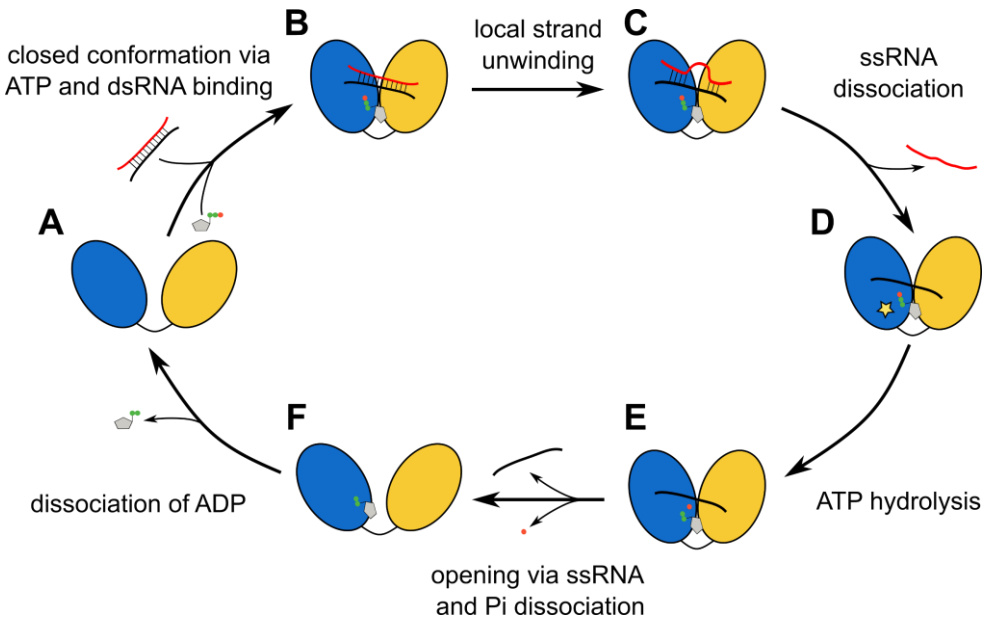


Figure 1.9: Schematic representation of the unwinding cycle of DEAD-box RNA helicases. The RecA1 domain is shown in blue, the RecA2 domain is represented in yellow. (**A**) The helicase is in its open state, the domains tumble independently. (**B**) Upon binding of dsRNA and ATP, both RecA domains of the folded helicase core undergo structural rearrangements and a transition from the open to the closed state takes place. (**C**) Binding of RNA results in local strand unwinding, which leads to the "kink" structure. (**D**) A single strand of RNA dissociates from the complex. (**E**) ATP hydrolysis. (**F**) The release of Pi causes a transition from the closed to the open state, leading to RNA dissociation. The now open conformation has a low affinity for RNA and ADP, leading to the dissociation of ADP.

1.5.2 Function of the DEAD-box RNA helicase Ded1p

The DEAD-box RNA helicase Ded1p is a 604 amino acid protein involved in translation initiation in *Saccharomyces cerevisiae* (*S. cerevisiae*) (**Figure 1.10**) (173–175). The presence of Ded1p is essential for the viability of yeast cells (176). During translation initiation, the cap-binding protein, eukaryotic translation initiation factor (eIF) 4E, binds to the N7-methyl guanosine (m_7G) cap structure of the mRNA. Together with eIF4G and eIF4A, another DEAD-box RNA helicase involved in translation, eIF4E forms the heterotrimeric complex eIF4F. This complex, in concert with additional factors, then recruits the 40S subunit of the ribosome, which subsequently scans the mRNA for the start codon. Once the start codon is reached, the 60S subunit of the ribosome is recruited and translation begins. Ded1p is capable of unwinding secondary structures in the 5' untranslated regions (UTRs) of the mRNA (160, 177). This facilitates the

scanning of the ribosome for the start codon. Ded1p is recruited to the scanning ribosome where the N-terminal IDR of Ded1p interacts with eIF4A and eIF4E, while the C-terminal IDR interacts with the RNA 3-binding domain of eIF4G (160). In addition to translation initiation, the human ortholog of Ded1p, DDX3, is further linked to other biological processes, including the regulation of the immune system (178). It was also shown that viruses can hijack DDX3 for their replication cycle and that tumorigenesis is often associated with mutated DDX3 (179–183).

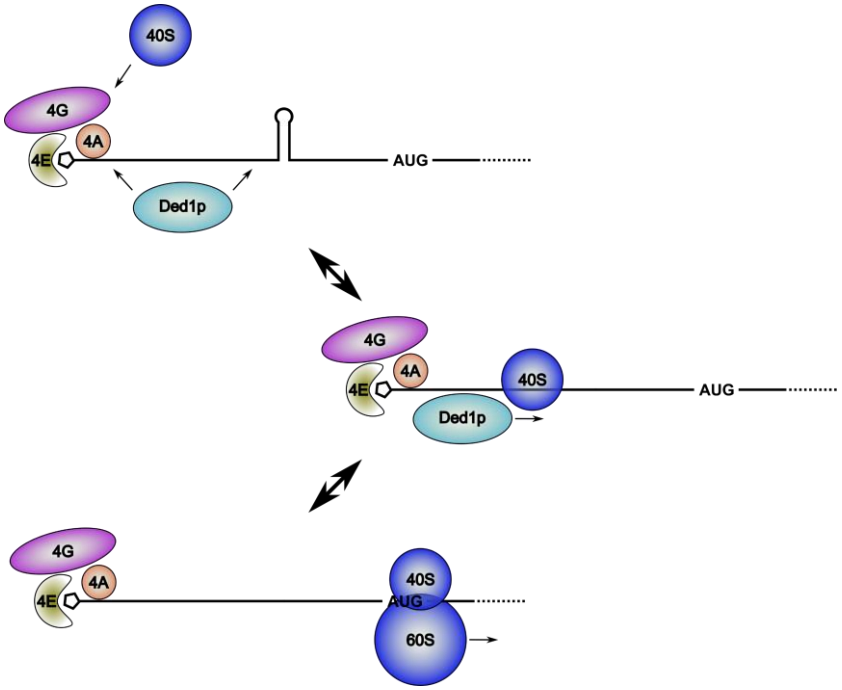


Figure 1.10: Scheme of the initial steps of translation initiation. The eIF4F complex, consisting of eIF4A-eIF4E-eIF4G, is recruited to the 5´end of an mRNA. The cap structure of the mRNA is bound to the cap binding protein eIF4E, which associates with the DEAD-box RNA helicase eIF4A and the scaffold protein eIF4G. The m_7G cap structure is shown as a pentagon. Ded1p interacts with eIF4A and unwinds the secondary structure elements located in the 5´ UTR of the mRNA. This facilitates scanning of the 40S ribosomal subunit for a start codon which is followed by the formation of an actively translating ribosome once a start codon is reached.

Ded1p facilitates the translation of housekeeping genes under normal growth conditions (46). At the same time, Ded1p is also a main component of SGs, and Ded1p activity is regulated via the formation of SGs (49, 184, 185). Upon heat-induced stress, Ded1p is recruited into SGs and binds housekeeping mRNAs that are also recruited into SGs. Subsequently, the translation of housekeeping genes is down-regulated while simultaneously mRNAs that do not contain secondary structure elements and

are encoding for heat-shock proteins are still translated (46, 49). Furthermore, *in vitro* studies have shown that Ded1p alone can undergo LLPS without RNA or additional SG proteins (46). Similar to other phase-separating proteins, the IDRs of Ded1p contribute to LLPS. The removal of either the N- or C-terminal IDR led to an almost complete disruption of phase separation or altered the c_{SAT} of the construct.

While the structure of the human Ded1p ortholog DDX3 has been solved by the Doudna group (186), a structure of the yeast helicase Ded1p is not yet available. Alphafold2 predicts a structure model for Ded1p (**Figure 1.11**) (187, 188). This model shows interactions between the N-terminal IDR and the RecA2 domain, while another interaction site is predicted for the C-terminal IDR and the RecA1 domain.

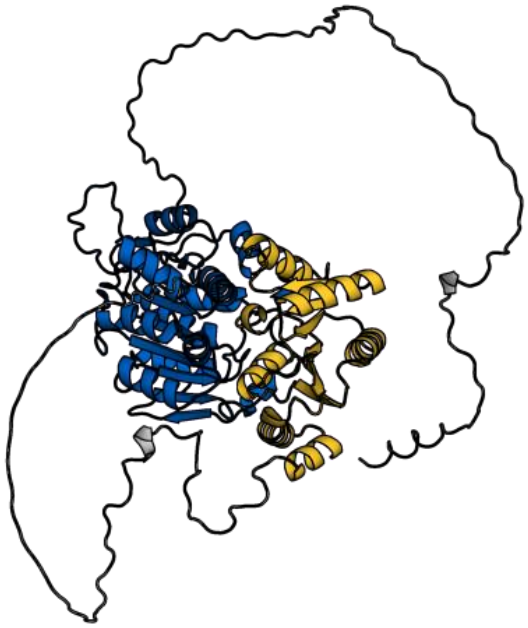


Figure 1.11: AlphaFold2 model of the DEAD-box RNA helicase Ded1p from *S. cerevisiae*. The RecA1 domain is colored blue while the RecA2 domain is highlighted in yellow. N- and C-terminal IDRs are flanking the folded helicase core and are shown in grey. PyMol was used for the representation of this model (189).

In summary, members of the DEAD-box RNA helicase family are found in all three domains of life. They participate in the life cycle of RNAs ranging from synthesis to decay (**Figure 1.12**). ATP is used for the unwinding of short RNA duplexes by local strand separation. The DEAD-box RNA helicase Ded1p plays an essential role in

translation initiation in *S. cerevisiae* and is also a main component of SGs. How Ded1p is recruited into SGs remains unclear.

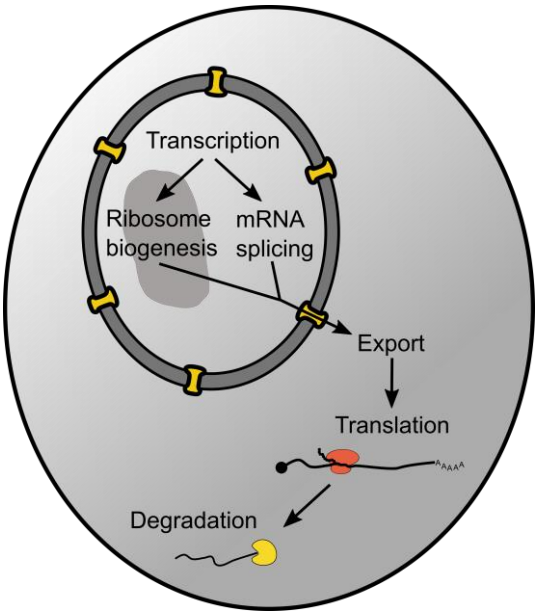


Figure 1.12: DEAD-box RNA helicases are required for many cellular processes. DEAD-box RNA helicases catalyze reactions that are involved in almost all processes of the RNA life cycle. The image was adapted and modified from Linder (153).

1.6 Nuclear magnetic resonance spectroscopy

1.6.1 Theoretical background of nuclear magnetic resonance spectroscopy

NMR spectroscopy is a powerful method for studying biomacromolecules in solution. It allows to analyze structure, interaction partners, and dynamics (on the picosecond to second time scale) of biomolecules with atomic resolution under near physiological conditions. Binding events with affinities up to the mM range can be investigated, and NMR allows to unravel correlations between biomolecular dynamics and the activity of a biomolecule (190).

NMR makes use of a special intrinsic angular momentum of atomic nuclei: the spin (191). The spin of an atomic nucleus is associated with a small magnetic moment. The spin angular momentum of a nucleus depends on the total number of protons and neutrons. NMR-active nuclei that are commonly observed in biological applications were chosen because they have a spin of ½ (^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P). The spin angular momentum $\vec{I}$ gives rise to a magnetic dipole moment $\vec{\mu}$ and the proportionality constant

linking these two quantities is called the gyromagnetic ratio γ, which depends on the type of nucleus (Equation 2) (192).

Equation 2 $$\vec{\mu} = \gamma \vec{I}$$

NMR relies on this magnetic dipole moment because it can interact with an external magnetic field and thereby local information about the chemical environment of nuclei can be obtained. Outside of a magnetic field, spins have a random orientation (**Figure 1.13A**). However, the magnetic dipole moments of spins orient themselves when an external magnetic field B_0 is applied (**Figure 1.13B**). The z-component of the magnetic dipole moment can orient either parallel or antiparallel to the B_0 field. When a system is in equilibrium, the two populations are not equal, but are distributed according to the Boltzmann distribution (Equation 3). As a result, the lower energetic state, the α state (dipole moment parallel to B_0), is higher populated than the energetically less favored β state (dipole moment antiparallel to B_0).

Equation 3 $$\frac{N(\beta)}{N(\alpha)} = exp(\frac{-\Delta E}{kT})$$

where ΔE stands for the energy difference between state α and β, k represents the Boltzmann constant, and T is the temperature (K). $N(\alpha)$ and $N(\beta)$ represent the number of nuclei that occupy the higher and lower energetic state, respectively.

In NMR, the observed signal is linked to the population difference between states α and β. This energy difference, and therefore also the population difference, is small compared to kT. The energy difference ΔE between two spin states depends on the magnetic field strength and the gyromagnetic ratio of the nucleus (Equation 4).

Equation 4 $$\Delta E = E_\beta - E_\alpha = \gamma \bar{h} B_0$$

where E_α and E_β stand for the energy of the higher and lower energetic state, respectively. $\bar{h}$ represents the Planck constant divided by 2π.

The sum of all magnetic dipole moments gives rise to the bulk magnetization M, which in equilibrium is parallel to the external magnetic field B_0 (**Figure 1.13B**). Furthermore, all individual spin vectors precess around the magnetic field B_0 at an angle θ

(**Figure 1.13C**) with their characteristic Larmor frequency ω_0 that depends on the external magnetic field B_0 and the gyromagnetic ratio (Equation 5).

Equation 5 $$\omega_0 = \gamma B_0$$

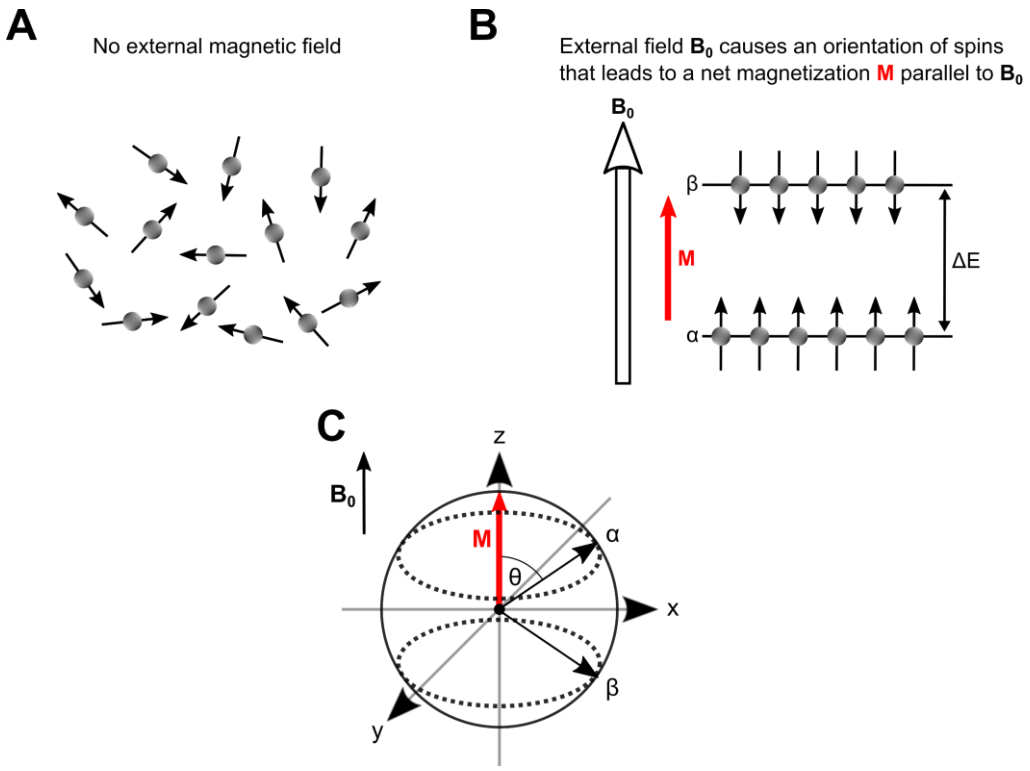


Figure 1.13: Simplified principle of NMR spectroscopy. All spins of the nuclei, shown here as grey filled circles, are represented as vectors in a three-dimensional space. (**A**) In the absence of an external magnetic field, the spins have a random orientation that leads to a net magnetization of zero. (**B**) Net magnetization is induced by applying an external field B_0. At equilibrium, there is a small preference for an alignment of the spins with the magnetic field. The *α* state is energetically favored over the *β* state which results in a slight excess of parallel aligned magnetic moments. The vector sum over the whole sample results in the bulk magnetization M along the direction of B_0. The *α* and *β* states are filled according to the Boltzmann distribution and give rise to two different energy levels. (**C**) The magnetic moment of individual spins precesses around the external magnetic field with an angle *θ*.

To perform an NMR experiment, the bulk magnetization must be rotated from the z-axis into the transverse x,y-plane so that the precession of the bulk magnetization around the external magnetic field can be measured. To achieve this, a 90° radio-frequency (RF) pulse is applied by setting up an oscillating magnetic field B_1 orthogonal to the B_0 field with the Larmor frequency of the nuclei (**Figure 1.14A**). The B_1 field is applied for only a short time (usually on the order of µs) and is therefore

referred to as a pulse. As a consequence, the magnetization is transferred to the transverse x,y-plane. Subsequently, all spins begin to precess around the external magnetic field with their Larmor frequency. This characteristic frequency depends on the local chemical environment of each nucleus. The oscillation in the transverse plane caused by the precession of the magnetization induces a current in a receiver coil, the free induction decay (FID), which is recorded over time.

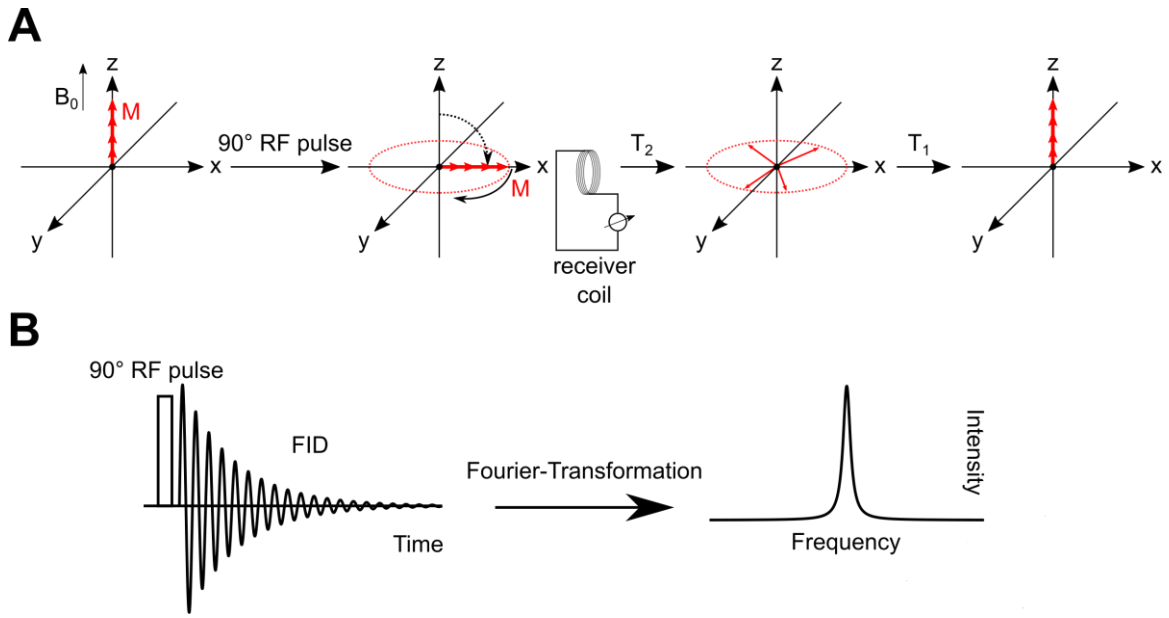


Figure 1.14: Simplified and schematic representation of a pulse-acquire NMR experiment. (**A**) The different population of the *α* and *β* states results in a bulk magnetization, which is visualized here as a red arrow along the z-axis with the external magnetic field B_0. The bulk magnetization is transferred from the z-axis into the x,y-plane by applying a 90° pulse. In the transverse x,y-plane, the individual spins precess with their Larmor frequencies. The oscillation in the transverse plane caused by the precession of the magnetization induces a current in a receiver coil. Due to R_1 and R_2 relaxation processes, M returns to its equilibrium state which causes a decay of transversal magnetization. (**B**) After a 90° pulse, which gives rise to the precessing bulk magnetization in the x,y-plane, a current is induced in the receiver coil. The FID is then transferred from the time-domain into the frequency-domain by Fourier-Transformation.

This time-dependent signal decreases exponentially with time since different relaxation processes take place that causes dephasing of spins. This process is known as transverse or spin-spin relaxation T_2 ($1/R_2$) which describes the loss of coherent magnetization during the rotation in the transverse plane. Mainly dipolar interactions with nearby spins, B_0 inhomogeneity, and chemical shift anisotropy contribute to the loss of coherence. The characteristic relaxation rate R_2 can be determined by measuring the linewidth of a signal at half the peak intensity. The R_2 relaxation rate is

strongly influenced by the tumbling rate of a biomacromolecule. In general, large proteins experience high relaxation rates due to a slow tumbling rate. This is due to the fact that the energy and the Larmor frequency are dependent of the orientation of a molecule in relation to B_0. This results in a dephasing of the signals that is averaged out by the tumbling of the molecule. The faster the tumbling, the better the averaging and R_2 rates become slower. This can be observed as peak broadening that is more severe for large proteins than for small ones. The characteristic time for the magnetization to be restored to equilibrium is its longitudinal T_1 ($1/R_1$) relaxation, also known as spin-lattice relaxation. T_1 is significantly longer than T_2 for larger molecules such as proteins and typically defines how long one must wait before a new NMR experiment can be performed. Depending on the nucleus and sample conditions, this process can take up to seconds. Recording R_1 and especially R_2 rates are commonly used techniques in bio-NMR to draw conclusions regarding protein dynamics.

The acquired time-dependent FID is transferred from the time domain into the frequency domain by Fourier-transformation (**Figure 1.14B**). As a result, the resonance frequencies of the nuclei are visualized by the characteristic NMR resonances that emerge. These peaks emerge at their characteristic Larmor frequencies, which report on the chemical environment of each nucleus. For a nucleus x, the chemical shift δ is defined by the following equation (Equation 6).

Equation 6 $$\delta(x) = \frac{\omega_x - \omega_{ref}}{\omega_{ref}} * 10^6$$

where ω_{ref} is defined as the frequency of a reference nucleus and ω_x is the Larmor frequency of the nucleus. Since units are given in parts per million (ppm), the chemical shift value is multiplied by 10^6.

Figure 1.14B shows a one-dimensional (1D) pulse-acquire NMR experiment that can be recorded for, e.g. ^1H. ^1H is a stable spin ½ isotope and thus NMR-active with a natural abundance of 99.99% making it an important reporter in NMR experiments.

There are several ways to improve the signal-to-noise ratio of an experiment, which can result in sharper signals. The magnetic field can be increased so that the energy difference between the energetic states is also increased, a high sample concentration can be used, the temperature can be increased in the case of thermostable

biomacromolecules which results in a faster tumbling rate, and experiments can be repeated many times while the signal is averaged.

1.6.2 NMR spectroscopy of biomacromolecules

For biomacromolecules, 1D ^{1}H NMR spectra already provide information about the protein. However, larger proteins typically have thousands of protons. To resolve overlapping signals and to obtain more site-specific information, the inclusion of a second or even a third NMR dimension is necessary. In the case of proteins, ^{13}C and ^{15}N are commonly used NMR-active isotopes. Since both share a low natural abundance of only 1.1% for ^{13}C and 0.36% for ^{15}N, an enrichment of both isotopes by specific labeling schemes is necessary in order to make use of these nuclei in NMR experiments (193).

In two-dimensional (2D) NMR experiments, an intricate sequence of RF pulses and delays is applied to the NMR sample, allowing the chemical shifts of two spins to be correlated. While in a 1D experiment a plot of intensity against frequency is shown, the intensity is plotted against two frequencies in a 2D experiment, one on each axis. This has the advantage that peak overlap occurs much less frequently because the signals are dispersed in a plane. In the case of ^{15}N-labeled proteins, 2D NMR experiments are commonly conducted that correlate N-H spins in the protein backbone by the magnetization transfer from ^{1}H to ^{15}N of the amide group. For example the ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiment gives rise to a single peak for every amino acid of a protein, expect proline, and is used as a fingerprint spectrum.

By adding a third dimension, usually ^{13}C for proteins, and combining a set of triple-resonance experiments, resonances in the 2D spectrum can be assigned and linked to their corresponding residue.

In addition to the transfer of magnetization via chemical bonds (J-coupling), one can also make use of the Nuclear Overhauser effect (NOE) for which magnetization is transferred between nuclei through space. By making use of this technique, information about the distance between different protons up to 5-10 Å can be revealed.

Using this approach, it is possible to determine the structures of biomacromolecules such as proteins, RNAs, and DNAs.

Heteronuclear NMR experiments with uniformly ^{13}C and ^{15}N labeled protein are limited to proteins up to 25-30 kDa in size. This is due to the vast number of NMR-active nuclei and the very fast R_2 relaxation which originates from the slow tumbling rate and leads to severe peak broadening rendering such experiments unfeasible. One way to decrease peak broadening is protein deuteration, which decreases the dipolar interactions (194). Another technique that can be utilized is selective isotope labeling of methyl groups with ^{1}H and ^{13}C ($^{13}C^{1}H_3$) in combination with uniform deuteration of the protein (195, 196). Methyl groups are ideal reporters for studying the dynamics and interactions of large proteins since they are mostly present in the hydrophobic core of proteins or at binding surfaces. $^{13}C^{1}H_3$ labeled methyl groups can be introduced into threonine (T), isoleucine (I), leucine (L), valine (V), methionine (M), or alanine (A) side chains. To produce proteins with such an ILVMA labeling scheme, one can make use of the metabolic pathways of a particular amino acid in bacteria that are used for recombinant protein expression (197). To do this, labeled precursors are added to a bacterial expression culture at specific time points before protein overexpression is induced, which will result in the incorporation of methyl labels at different positions. With this strategy and by recording methyl transverse relaxation optimized spectroscopy (TROSY) spectra, well-resolved 2D ^{1}H-^{13}C NMR spectra can be recorded due to the beneficial relaxation properties of methyl groups for proteins that are bigger than 30 kDa (198).

Still, the assignment of methyl groups in large protein complexes can be challenging. One way to obtain methyl group assignment is by employing a “divide and conquer” strategy where large complexes are divided into smaller domains (196). Smaller domains usually show a better spectral quality than the large complex which makes it therefore easier to assign resonances. Methyl group resonances can be assigned to the smaller domains and the assignments are then transferred to the large complex. However, this is only possible when the subunits are independent of each other.

In addition, single residues can be mutated to structurally related residues. By comparing NMR spectra before and after mutagenesis, the resonance of the substituted residue can be easily identified as it is missing from the spectrum of the

mutated protein. A third way of obtaining methyl group assignment is to employ computer software that combines known structural information with experimentally determined NEO distance data. An example for this is the methyl assignment by graphic inference construct (MAGIC) algorithm (199). This algorithm makes use of three-dimensional (3D) methyl Nuclear Overhauser Enhancement Spectroscopy (NOESY) experiments and combines these data with the known structure to calculate the assignments of the resonances based on the local connection of the methyl groups in the protein structure.

Since proteins are not static but flexible biomacromolecules, dynamics and conformational changes are essential for their biological function. Unfolded proteins, which lack a unique three-dimensional structure, cannot be studied by X-ray crystallography or cryogenic electron microscopy (cryo-EM) because both techniques are limited to static structures. NMR is one of the suitable biophysical techniques to study the association of IDRs and low complexity domains in LLPS (86). Although IDRs are highly dynamic in nature, they are still capable to form transient and intermolecular interactions that are linked to their biological function. This can be followed by NMR with atomic resolution (200). In some cases, it could be shown by NMR that IDRs adapt largely the same ensemble of conformations in the dispersed phase and in the condensed phase (201, 202).

To probe transient and weak long-range (self-)interactions of IDRs, which are difficult to detect with conventional NMR experiments, PRE NMR can be employed. The PRE effect arises from dipolar interactions between the unpaired electron of a paramagnetic probe, hereafter referred to as a spin-label, with the observed nucleus. As a result, the R_2 relaxation rate is enhanced which can be detected by an increase in the line width of the peaks. By making use of this technique, long-range distances up to 10-25 Å can be probed (203). A spin-label can be introduced site-specifically by conjugation to a cysteine residue.

Taken together, solution NMR is a versatile tool for obtaining structural information of biomacromolecules under conditions that are close to their natural environment. Employing specific labeling of methyl groups together with methyl TROSY spectroscopy, large protein complexes can be studied by NMR. On the other hand, NMR is also a powerful biophysical technique that has been applied to study

LLPS-prone proteins, which often contain large, disordered regions. For these proteins, NMR can be utilized to visualize short-lived and transient long-range interactions that mainly contribute to the formation of the complex LLPS network.

1.7 Aims of this thesis

My goal was to elucidate the molecular mechanism of DEAD-box RNA helicases and to gain a more detailed understanding of the molecular interactions that lead to LLPS of Ded1p.

In **Chapter 2**, NMR titration experiments combined with fluorescence anisotropy binding experiments and an ATPase assay were used to study the conformational changes of the core domains of DEAD-box RNA helicases upon RNA and nucleotide binding. In addition, RNA binding was investigated, and the formation of the closed core state was followed.

Chapter 3 and **Chapter 4**, a suite of different NMR experiments of the folded domains and IDRs of the DEAD-box RNA helicase Ded1p were used in combination with an LLPS assay to investigate the molecular grammar that governs the driving force of LLPS of Ded1p. This approach was further complemented by X-ray crystallography.

Chapter 2: Systematic assessment of the conformational changes in DEAD-box RNA helicases

Contributions: This project was originally started by Philip Wurm and he performed parts of the NMR titration experiments. I finished the project by recording the remaining NMR spectra and performing all the fluorescence anisotropy and ATPase assays. The results of this chapter are in preparation for a publication. Therefore, this chapter will be largely identical in text and figures to the manuscript:

Jan Philip Wurm, **Julian Hübner**, and Remco Sprangers. Systematic assessment of the conformational changes in DEAD-box RNA helicases, *in preparation*

2.1 Introduction

The correct folding of an RNA is critical for its function, whereas aberrant but stable secondary structures can interfere with function or become harmful to a cell (145). In order to prevent incorrect folding, RNA helicases unwind RNA duplexes and can remodel protein-RNA complexes, making DEAD-box RNA helicases essential for cells (144). DEAD-box RNA helicases represent the largest family of RNA helicases and are ubiquitously found in all aspects of RNA metabolism (146, 147). This class of enzymes unwind RNA duplexes, typically up to 12-16 base pairs long, in an ATP-dependent manner, irrespective of the RNA sequence (171, 174). Structurally, DEAD-box RNA helicases consist of a functional core formed by two RecA-like domains (152, 153). The two RecA domains are covalently connected via a short, flexible linker (154). For DEAD-box RNA helicases, the helicase core harbors the conserved motifs that are necessary for ATP binding and hydrolysis as well as for RNA binding (**Supplemental Figure 2.1**) (155). However, studies of the full-length proteins have shown that RNA binding differs depending on nucleotides and is not similar for all DEAD-box RNA helicases (155).

In addition, the core of DEAD-box RNA helicases is usually flanked by auxiliary N- and C-terminal domains (157, 158, 162, 163). Such extensions can vary in length from only a few up to hundreds of amino acids. Extensions are mostly intrinsically disordered and allow recruitment of the helicase to its target, and show only high sequence conservation among protein orthologs (159). Therefore, the role of the

individual RecA domains and especially the role of the auxiliary domains in RNA and nucleotide binding remains to be elucidated.

In the absence of substrate RNA and ATP, the helicase core adopts an open conformation in which both domains tumble independently (155, 158). Upon binding of RNA and ATP, the helicase core undergoes a conformational rearrangement in which ATP is sandwiched between both RecA domains and one strand of the RNA duplex is bound to the core (**Figure 1.9**) (154, 166). In the closed conformation, the RNA duplex is destabilized and the unbound RNA strand dissociates spontaneously. Upon ATP hydrolysis, ADP and single-stranded (ss) RNA are released, and the helicase returns to its open conformation. This highlights that ATP hydrolysis is only required for product release (168). Still, how the initial interactions between the helicase and the substrate RNA are mediated remains largely elusive. ADP and ATP are important modulators of RNA binding and, vice versa, RNA interactions with the helicase influence nucleotide binding. Most biochemical studies use ATP analogues instead of ATP. However, how the nature of different types of nucleotide analogues affects RNA binding appears to vary among DEAD-box proteins (204–207).

In the last two decades, several structures of the DEAD-box helicase core in the RNA-bound state with ADP or non-hydrolysable ATP analogues have been solved (150, 156, 167, 186, 208, 209). Almost all structures capture the post-unwound state of the helicase for which a ssRNA is bound. In all these structures, the helicase interacts with the phosphate backbone of the RNA (162, 169, 170). Furthermore, several structures of DEAD-box helicases have been solved in the absence of RNA. For such helicases, the reaction product adenosine monophosphate (AMP) is found in the ATP binding pocket (210). This indicates that some helicases also have an affinity for AMP.

Instead of ATP, nucleotide analogues are commonly used in biochemical studies, including AMPPNP, in which the terminal bridge oxygen is replaced by an NH group (211) or adenosine 5´-(gamma-thiotriphosphate) (ATP-γ-S). In addition, protein structures in which ADP is in a complex with BeF_3^- or AlF_3^- have also been reported (208, 212). However, most studies do not take into account that ATP analogues are not able to fully mimic the chemical properties of ATP, since modifications change their chemical properties compared to ATP and can also affect the geometry of the analogues (213). Although ATP mimics are considered to be non-hydrolysable, this

may not strictly be the case and resistance against hydrolysis must be tested experimentally for each nucleotide in studies with a new DEAD-box RNA helicase. This was shown in a study with the bacterial DEAD-box RNA helicase DbpA, for which ATP-γ-S was not hydrolysis-resistant (208). Furthermore, a case study of the *Helicobacter pylori* DnaB helicase demonstrated that AMPPNP, as well as ATP-γ-S were not resistant to hydrolysis (213). In addition, it was observed that different ATP analogues could interfere with the functionality of the protein, and that due to the different conformations that these analogues can adopt, they can fail to induce different states in the hydrolysis cycle (213). These observations indicate that the right choice of an ATP analogue is highly dependent on the protein under investigation.

Here, we systematically assessed whether the two RecA domains of different DEAD-box RNA helicases tumble independently in the absence of RNA and nucleotides. Furthermore, we searched for a suitable ATP analogue that can be used for structural studies of DEAD-box RNA helicases. To do so, we investigated if the closed core state for five helicase core constructs (eIF4A and Dhh1 from *Chaetomium thermophilum* (*C. thermophilum*), Ded1p and Dbp5 from *S. cerevisiae*, and DDX3 from *Homo sapiens* (*H. sapiens*)) that all lack their auxiliary flanking domains (**Figure 2.1**) is formed in the presence of several ATP analogues.

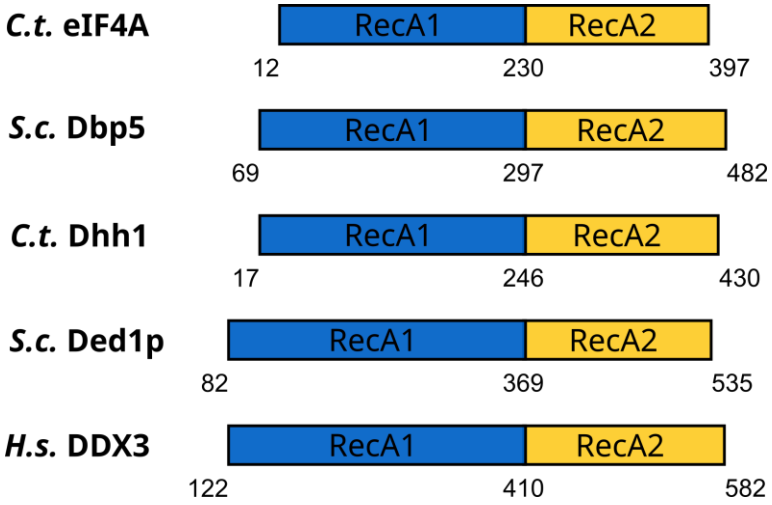


Figure 2.1: Schematic domain architecture of the five DEAD-box RNA helicases studied here. Only the helicase core, consisting of two RecA-like domains, is studied. The helicases differ slightly in their domain size.

All helicase constructs harbor the conserved motifs that are necessary for the binding and hydrolysis of ATP and for RNA binding (**Supplemental Figure 2.1**). eIF4A is a minimalistic helicase that only consists of the helicase core with a very short N-terminal extension (214). It was the first DEAD-box RNA helicase that had been structurally studied and is part of the translation initiation complex. Ded1p and its human ortholog DDX3 are also involved in translation initiation in eukaryotic cells as they unwind secondary structures in the 5' UTR of mRNAs (46, 177). Dbp5 is a DEAD-box protein that is associated with the nuclear core complex and has an essential role for the mRNA export from the nucleus (215). Finally, Dhh1 functions in translation suppression and the decay of mRNA as an important regulator of mRNA decapping (216).

In this study, we employed NMR titration experiments in combination with fluorescence anisotropy binding experiments and ATPase assays. Using fully deuterated proteins for which the Ile-δ1, Leu-δ1, Leu-δ2, Val-γ1, Val-γ2, and Met-ε methyl groups were $^{13}CH_3$ labeled, methyl TROSY NMR spectra were recorded. This allows us to follow the formation of the closed state upon RNA and nucleotide binding of the isolated helicase cores.

Our results demonstrate that RNA and nucleotide binding affinities do not vary significantly between the core structures of the different DEAD-box RNA helicases, but we were able to show that AMPPNP and ADP-BeF_3 differ notably in their ability to induce the closed core conformation in the presence of ss- and dsRNA.

2.2 Methods

2.2.1 Protein expression and purification

All genes for the core constructs (RecA1-RecA2) of the different host proteins were cloned into pETM-11 plasmid vectors that carried an N-terminal tobacco etch virus (TEV) cleavable His_6-tag. The plasmid was transformed into chemically competent *Escherichia coli* (*E. coli*) BL21(DE3) codon plus cells (Stratgene).

To produce unlabeled protein, cells were grown at 37°C while shaking in lysogeny broth (LB) medium until an optical density at a wavelength of 600 nm (OD_{600}) of 0.6-0.8 was reached. Next, protein expression was induced by the addition of 1 mM isopropyl

β-d-thiogalactopyranoside (IPTG). Cells were shifted to 20°C and incubated for 12-18 h, after which they were harvested by centrifugation (6,000 g for 15 min at 20°C). To obtain ILVM-labeled protein, 25 mL of H_2O based M9 medium was inoculated with 0.5 mL of LB overnight culture and incubated at 37°C until an OD_{600} of 0.8 was reached. Cells were gently harvested by centrifugation (4,000 g for 10 min at 20°C) and were further used to inoculate 100 mL D_2O based M9 medium containing 4 g/L $^{2}H^{12}C$-glucose to an OD_{600} of 0.15. The M9 medium was incubated at 37°C overnight, diluted the next day with fresh M9/D_2O medium to a final volume of 800 mL, and the cells were further incubated at 37°C until an OD_{600} of 0.8 was reached. Next, 60 mg/L 2-ketobutyric acid-(4-^{13}C,3,3-d_2) and 100 mg/L 2-Keto-3-methyl-butyric acid-(dimethyl-$^{13}C_2$, 3-d) dissolved in 200 mL M9/D_2O medium, and 100 mg/L L-methionine (methyl-$^{13}CH_3$) were added one hour prior to induction. Protein expression was induced by the addition of 1 mM IPTG, cells were shifted to 25°C, and harvested by centrifugation (6,000 g for 15 min at 25°C) after 12-18 h.

Cell pellets were resuspended in 25 mL Buffer A (25 mM NaPi pH 7.4, 500 mM NaCl, 1 mM dithiothreitol (DTT), 10 mM imidazole) supplemented with 1:1000 volume per volume (v/v) Triton-X-100, 1 mg/mL lysozyme, and 0.1 mg/mL DNase I. Resuspended cells were lysed by sonication and cell debris was removed by centrifugation (39,191 g for 30 min at 4°C). The supernatant was filtered using a 0.45 µm syringe filter and was then applied to a gravity flow nickel-nitrilotriacetic acid (Ni-NTA) column that has been equilibrated with Buffer A. The column was then washed with five column volumes of Buffer A and 10 mL high salt washing buffer (25 mM NaPi pH 7.4, 2 M NaCl, 1 mM DTT) to elute protein-bound RNA. Next, proteins were eluted with Buffer B (25 mM NaPi pH 7.4, 500 mM NaCl, 1 mM DTT, 300 mM imidazole). The N-terminal His_6-tag was removed by TEV cleavage while the sample was dialyzed against dialysis buffer (25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.3, 125 mM NaCl, 1 mM DTT) overnight at room temperature. The cleaved N-terminal His_6-tag was removed by a second Ni-NTA column that was equilibrated with dialysis buffer. The flow-through was concentrated to a volume of ~2 mL, and the protein was further purified by size exclusion chromatography (SEC) using a 16/600 Superdex S75 column that was equilibrated with SEC buffer (25 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT). Protein-containing fractions were pooled, concentrated, and the final protein concentration was determined based on the absorption at 280 nm (A_{280}). All

proteins were frozen in liquid nitrogen and stored at -80°C until they were used for further experiments. All protein constructs that were used in this study are listed in **Supplemental Table 2.1**.

2.2.2 Nuclear magnetic resonance spectroscopy

All NMR experiments were performed at 25°C or 30°C on Bruker 600 MHz and 800 MHz Avance Neo spectrometers equipped with either nitrogen-cooled (600 MHz) or helium-cooled (800 MHz) triple resonance cryoprobes. To record methyl-TROSY NMR spectra, the SOFAST-heteronuclear single quantum coherence (HMQC) pulse sequence was employed (217). All NMR spectra were processed using either Topspin 4.02 or NMRPipe 9.6 (218) and were displayed using NMRDraw.

The final composition of the NMR sample was SEC buffer supplemented with 5 mM $MgCl_2$ and 5% (v/v) D_2O for frequency locking. Concentrations of ILVM-labeled protein ranged from 35 to 100 µM. For RNA titration experiments, 12mer ssRNA (5'-GGAGAGAGAAGG-3') or 12mer dsRNA (5'-GGAGAGAGAAGG-3' hybridized with 5'-CCUUCUCUCUCC-3') were added in 1.25-fold molar excess over the protein. All RNAs were obtained by Eurofins or IDT. To form the AMPPNP bound state, 2 mM AMPPNP was added. To form an ADP-bound state, 2 mM ADP was added. Following this, the ADP-BeF_3 bound state was induced by subsequent addition of 2 mM BeF_2 and 10 mM NaF.

2.2.3 Fluorescence anisotropy

Fluorescence anisotropy measurements were performed in SEC buffer supplemented with 5 mM $MgCl_2$ and 0.002% (v/v) Triton-X-100. In all experiments, the RNA concentration was 25 nM. To determine affinities of the helicase cores to ssRNA, a 5'-fluorescein labeled 12mer 5'-GGAGAGAGAAGG-3' ssRNA was used. Fluorescence anisotropy was recorded in the absence and presence of nucleotides. Either 3 mM AMPPNP or, for ADP-BeF_3, 3 mM ADP, 3 mM BeF_2, and 10 mM NaF were added to the reaction mixture. For RNA affinity measurements in the open conformation of the helicase, no nucleotide was added. Experiments were recorded in technical triplicates using a TECAN spark plate reader at 25°C. The excitation and

emission wavelengths were 485 nm and 535 nm, respectively. The dissociation constant of the protein/RNA complex, K_D, is obtained from the anisotropy A employing a one-site binding model (Equation 7).

Equation 7 $$A = A_0 + \Delta A * \left(\frac{[R] + [P] + K_D}{2} - \sqrt{\left(\frac{[R] + [P] + K_D}{2}\right)^2 - [P] * [R]} \right)$$

where A_0 is the anisotropy of the free RNA, ΔA represents the increase in anisotropy due to binding, [R] is the concentration of fluorescein labeled RNA, and [P] is the protein concentration. Data were analyzed using in-house written Matlab scripts.

2.2.4 ATPase activity assay

The ATPase activity of the different DEAD-box RNA helicase core constructs in the presence of a 12mer dsRNA dsRNA (5'-GGAGAGAGAAGG-3' hybridized with 5'-CCUUCUCUCUCC-3') was determined by making use of a coupled pyruvate kinase/lactate dehydrogenase assay (219, 220). In this assay, the production of ADP is linked to the oxidation of nicotinamide adenine dinucleotide (NADH) which can be followed by monitoring the absorption of NADH at 340 nm. A reaction mixture with a total volume of 150 µL contained 20 µM protein, 5 µM of a 12mer dsRNA, 450 µM NADH, 1.5 mM pyruvate, 10 U/mL pyruvate kinase/lactate dehydrogenase (from rabbit muscle, Sigma Aldrich), and 5 mM $MgCl_2$ in SEC buffer. Samples were prepared in 96-well plates and were subsequently incubated for 5 min at 25°C. Experiments were recorded in technical triplicates on a TECAN spark plate reader. ATPase activity was initiated by the addition of 3 mM ATP and oxidation of NADH was followed by detecting the loss of absorption every 20 s at 340 nm. The linear regime of the absorption time course was fitted using a linear equation and the hydrolysis rates were calculated based on the following equation using an in house written Matlab scripts (Equation 8).

Equation 8 $$k_{obs} = -\frac{\Delta A}{\Delta t} * \left(\frac{1}{K_{path}}\right) * \left(\frac{1}{[P]}\right)$$

where k_{obs} is the ATPase activity, ΔA/Δt represents the slope of the linear fit, K_{path} is the molar absorption coefficient of NADH at 340 nm for the path length in the 96-well plate (2.22 mM^{-1}) (208), and [P] is the protein concentration.

2.3 Results

2.3.1 The two RecA-like domains tumble independently in solution

Solution NMR was applied to investigate whether the two RecA-like domains of the studied helicase cores (eIF4A and Dhh1 from *C. thermophilum*, Ded1p and Dbp5 from *S. cerevisiae*, and DDX3 from *H. sapiens)* tumble independently in the absence of substrate RNA and ATP. To probe this, ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled helicase core domains as well as of the isolated RecA1 and RecA2 domains were recorded and compared (**Figure 2.2** and **Supplemental Figure 2.2**). The sum of the spectra of the isolated RecA1 and RecA2 spectra largely resemble the spectrum of the helicase core in all cases, indicating that both domains tumble independently in solution and do not make any specific contacts. In the case of Dhh1 and Ded1p, a few resonances of the full-length core construct showed no corresponding resonance in any of the two individual domains. Most likely, the disappeared resonances correspond to residues close to the domain border between RecA1 and RecA2.

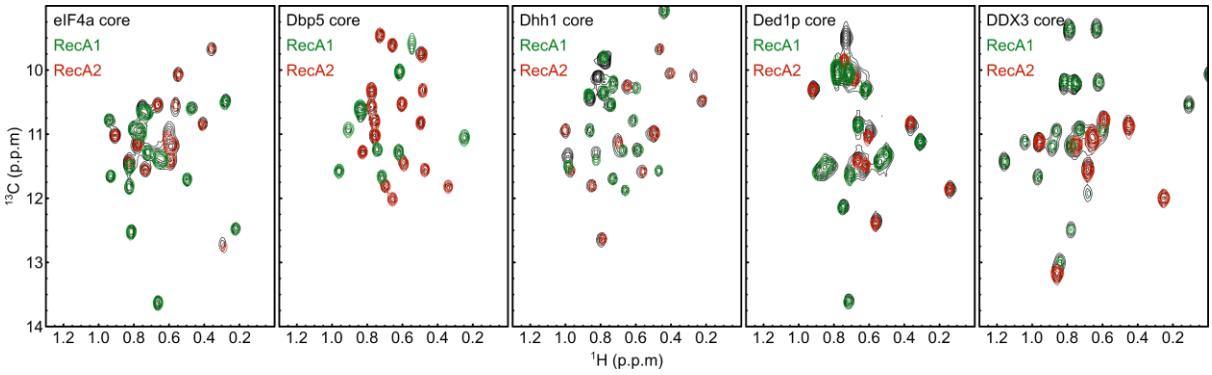


Figure 2.2: RecA-like domains tumble independently in solution. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core (black), RecA1 (green), and RecA2 (red), respectively.

2.3.2 ADP is bound by the RecA1 domain

In a next step, ADP was added to the core complexes. The binding of a nucleotide can be followed by monitoring changes of the chemical shift of methyl group resonances. Addition of ADP resulted in chemical shift perturbations (CSP) for all helicase cores (**Figure 2.3** and **Supplemental Figure 2.3**). Mostly residues of the RecA1 domain are affected by ADP binding, while residues of the RecA2 domain show less CSPs, demonstrating that the RecA1 domain interacts with ADP. This is expected since

nucleotide binding motifs are located in the RecA1 domain based on **Supplemental Figure 2.1**.

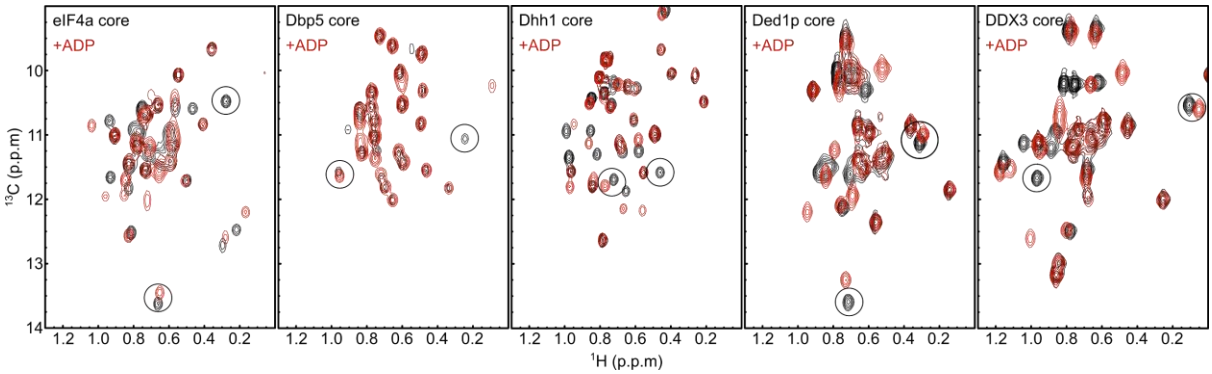


Figure 2.3: ADP is bound by the RecA1 domain. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of ADP. As an example, circles highlight some of the affected signals of RecA1.

2.3.3 AMPPNP is bound by the same residues that bind ADP

AMPPNP is an ATP analogue that is frequently used to stabilize the RNA-bound state of DEAD-box RNA helicases, but has also been used in studies in the absence of RNA. Upon the addition of AMPPNP, only minor CSPs were observed (**Figure 2.4** and **Supplemental Figure 2.4**). The CSPs can be mapped to the same residues that were affected by the addition of ADP (**Figure 2.3**) and all belong to the RecA1 domain. This indicates that ADP and AMPPNP both share the same binding site. Nevertheless, the CSPs are less pronounced for AMPPNP than for ADP. The addition of more AMPPNP does not lead to a stronger extent of the CSPs (**Supplemental Figure 2.5**).

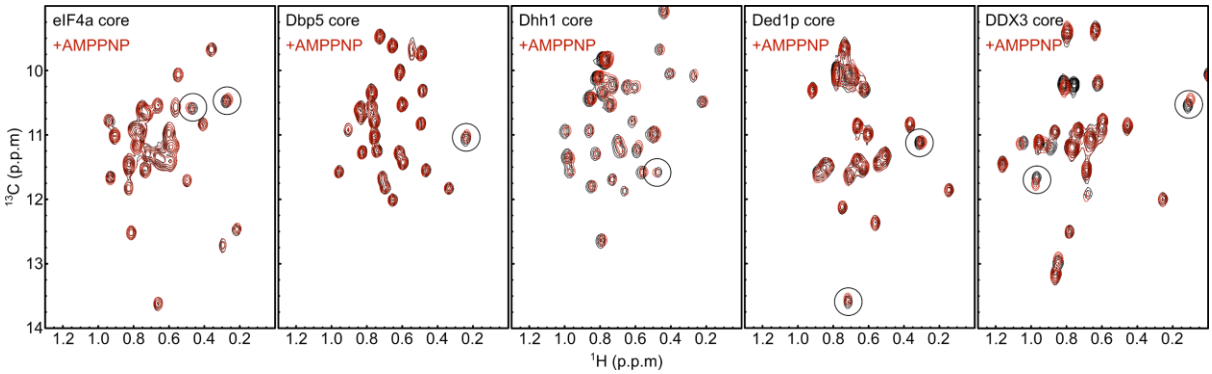


Figure 2.4: AMPPNP is bound by the RecA1 domain. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of AMPPNP. As an example, circles highlight some of the affected signals of RecA1.

2.3.4 Helicase cores do not significantly bind single-stranded RNA in the absence of a nucleotide

Having established that ADP and AMPPNP bind to the same residues, we further investigated the RNA binding of the DEAD-box RNA helicases. NMR spectra revealed that the isolated helicase cores have almost no affinity for a 12mer ssRNA. No CSPs are observable in the case of eIF4A, Dbp5, and Dhh1, whereas small CSPs are detected for Ded1p and DDX3 (**Figure 2.5** and **Supplemental Figure 2.6**). The CSPs for Ded1p and DDX3 are both linked to residues of the RecA1 domain, indicating very weak RNA binding. Interestingly, most of the CSPs can be mapped to the same residues that were affected by the addition of ADP and AMPPNP located in the RecA1 domain and would suggest that the observed RNA binding is to the ATP binding site and not to the RNA binding site.

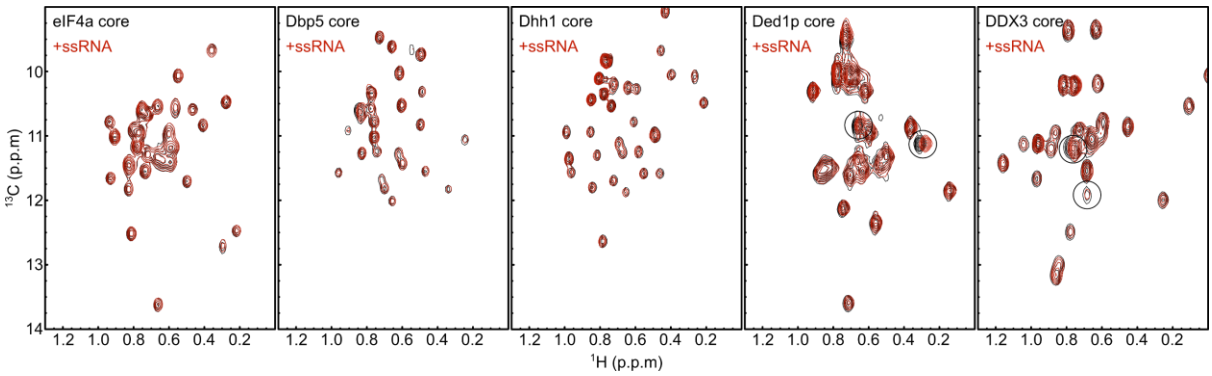


Figure 2.5: Single-stranded RNA is not bound by the helicase cores in the absence of nucleotides. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer ssRNA. As an example, circles highlight some of the affected signals of RecA1 for Ded1p and DDX3.

2.3.5 ssRNA and AMPPNP induce a partially closed conformation

To induce the closed conformation, both ssRNA and AMPPNP were subsequently added to the helicases and CSPs can be observed for eIF4a, Dbp5, and Ded1p (**Figure 2.6** and **Supplemental Figure 2.7**). CSPs can be mapped to residues of both RecA-like domains showing that binding of ssRNA in combination with AMPPNP is linked to a structural rearrangement of the helicase core. Signals for the open state are still detectable, which revealed that the closed conformation is not fully obtained. This implies that not all cores are in the closed conformation. In contrast, Dhh1 and DDX3 do not undergo any conformational rearrangements in the presence of ssRNA

and AMPPNP. Even though CSPs are observable, they are only visible for residues of the RecA1 domain and resemble the CSPs that were observed upon addition of AMPPNP (**Figure 2.4**). For both helicases, no RNA binding was observed immediately after the addition of the analogue AMPPNP.

However, the closed conformation of DDX3 was almost completely reached after 12 h of incubation, indicating that the binding of AMPPNP and the subsequent structural rearrangement is very slow (**Supplemental Figure 2.8**). In conclusion, our data demonstrates that binding of AMPPNP is necessary but not for all helicases sufficient in order to increase the affinity for ssRNA, resulting in the closed conformation of the helicase cores. However, in all cases open conformations were still observable implying that AMPPNP does not give rise to a stable closed conformation.

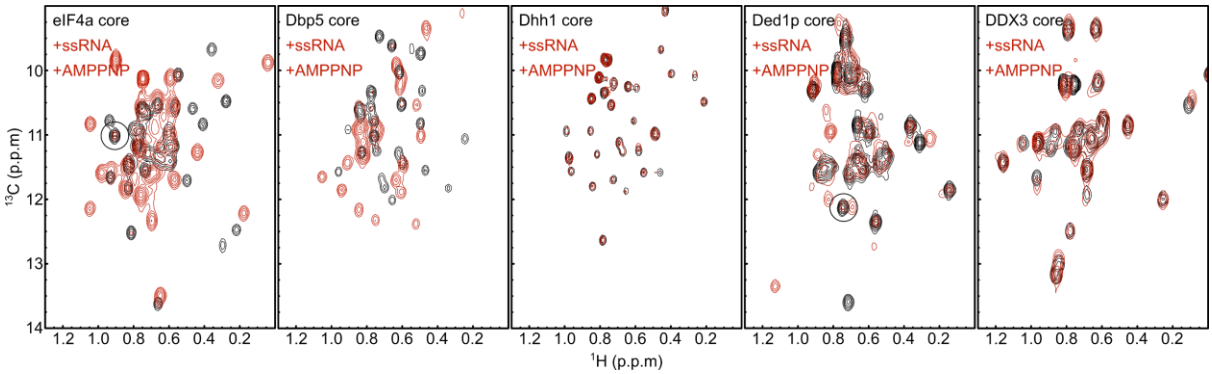


Figure 2.6: Single-stranded RNA and AMPPNP partially induce the closed conformation. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer ssRNA and AMPPNP. Signals for the open state that are still detectable are highlighted with circles.

2.3.6 Closed core conformation is induced with ADP-BeF_3 for all DEAD-box helicases

Next, we asked whether ADP-BeF_3 is capable of inducing the closed conformation in the DEAD-box RNA helicases. Following the addition of ssRNA, ADP, NaF, and BeF_2 were added to the NMR sample. All ^{1}H-^{13}C methyl TROSY spectra revealed strong CSPs for residues of both RecA-like domains demonstrates that all helicases underwent structural rearrangements (**Figure 2.7** and **Supplemental Figure 2.9**). Furthermore, the closed conformation is fully obtained as signals from the open conformation are no longer detectable. Taken together, these findings demonstrate that, in contrast to AMPPNP, ADP-BeF_3 is able to induce a complete transition from

the open to the closed conformation for all five helicase cores. In that light, we can conclude that ADP-BeF_3 is a potent ATP analogue for DEAD-box RNA helicases.

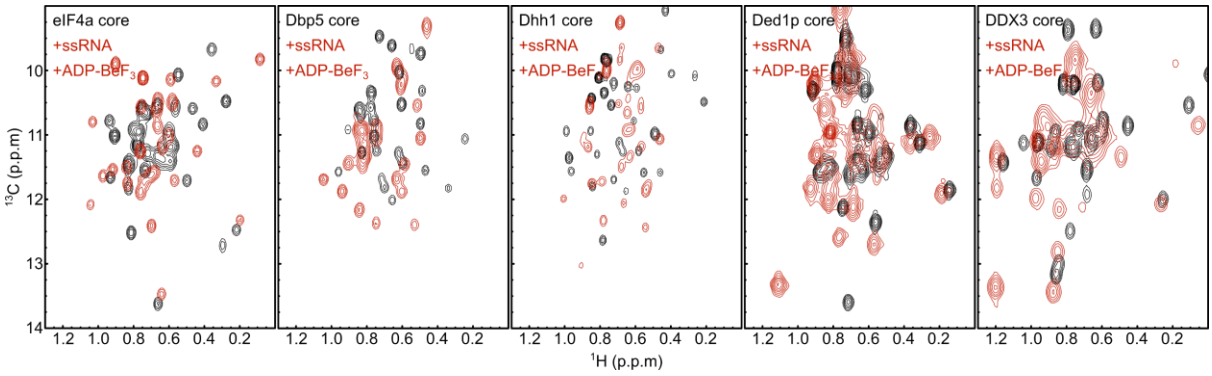


Figure 2.7: Single-stranded RNA and ADP-BeF_3 induce the transition from the open to the closed conformation. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer ssRNA and ADP-BeF_3.

2.3.7 Helicases show no affinity to double-stranded RNA in the absence of a nucleotide

As a follow up, we investigated the binding of dsRNA to the helicases. Like ssRNA, a 12mer dsRNA does not bind to any of the helicase cores in the absence of a nucleotide as no CSPs are observable (**Figure 2.8** and **Supplemental Figure 2.10**). While Ded1p and DDX3 revealed CSPs for a single resonance each upon addition of ssRNA, only a few weak CSPs are observable for Ded1p in the presence of dsRNA, but none are seen for DDX3. The few CSPs observed for Ded1p can be mapped to residues of the RecA1 domain, indicating very weak binding of dsRNA.

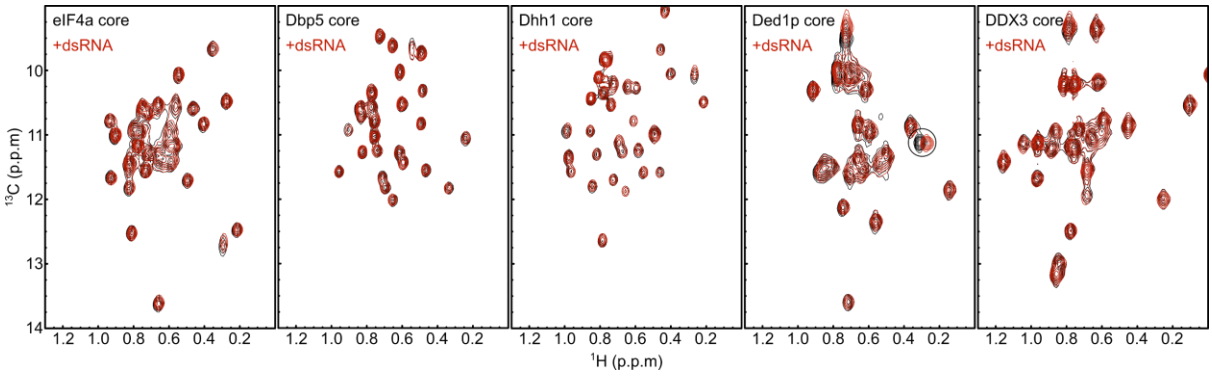


Figure 2.8: Double-stranded RNA is not bound by the helicase cores in the absence of an ATP analogue. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer dsRNA. As an example, a circle highlights one of the affected signals of RecA1 for Ded1p.

2.3.8 The closed conformation is not adopted with dsRNA and AMPPNP

Addition of AMPPNP together with dsRNA is not able to induce the transition from the open to the closed conformation (**Figure 2.9** and **Supplemental Figure 2.11**). Weak CSPs, detectable upon addition of dsRNA and AMPPNP, can be attributed to AMPPNP binding and are similar to spectra for which only AMPPNP was added (**Figure 2.4**). Thus, AMPPNP is not able to induce a conformational switch and does not increase the affinity for dsRNA.

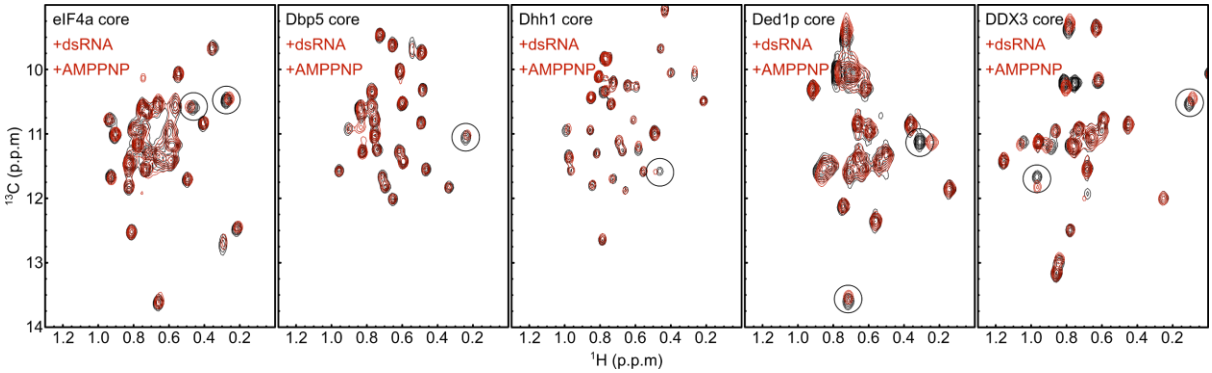


Figure 2.9: Double-stranded RNA and AMPPNP do not induce the transition from the open to the closed conformation. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of 12mer dsRNA and AMPPNP. As an example, circles highlight some of the affected signals of RecA1.

2.3.9 A partially closed conformation is obtained with ADP-BeF_3 and dsRNA for all DEAD-box helicases

Upon addition of 12mer dsRNA and ADP-BeF_3, CSPs are observed for residues of both domains, demonstrating that the closed state of the helicase cores is formed (**Figure 2.10** and **Supplemental Figure 2.12**). However, for eIF4a, Dhh1, and DDX3, the closed conformation was not fully reached as signals of the open conformation are still detectable even in excess of the ligands. This indicates that the transition from the open to the closed conformation is slow for these helicases.

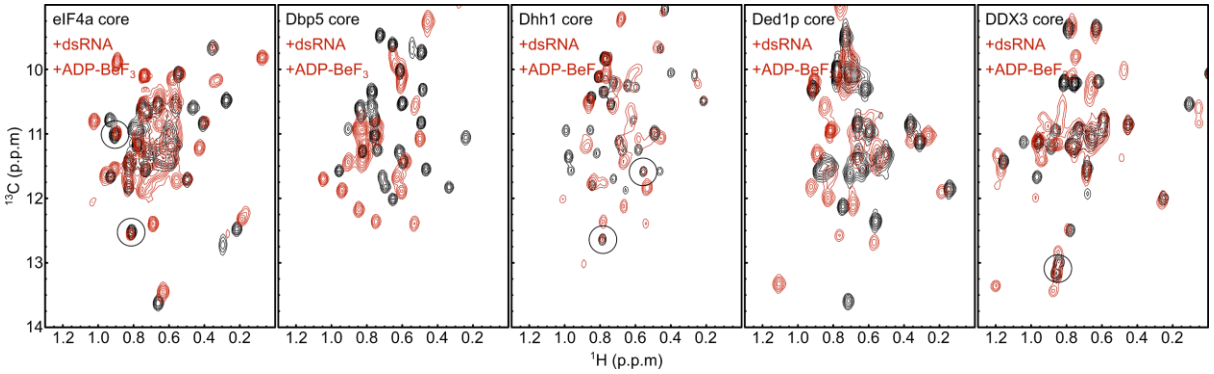


Figure 2.10: Double-stranded RNA and ADP-BeF$_3$ induce the transition from the open to the closed conformation. Overlay of the isoleucine region of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer dsRNA and ADP-BeF$_3$. Signals for the open state that are still detectable are highlighted with circles.

2.3.10 RNA binding affinity differs between helicases

Next, we sought to study RNA binding affinities of the helicases. For this we turned to fluorescence anisotropy binding experiments. We find that ssRNA is not bound in the absence of a nucleotide or in the presence of AMPPNP, in agreement with NMR titration experiments (**Figure 2.5**). Binding curves revealed that 12mer ssRNA is bound with high affinity only in the presence of ADP-BeF$_3$ (**Figure 2.11A-E**). Still, the RNA affinities differ between the helicases and the k_D of Dhh1 is significantly lower compared to the other helicase cores (**Figure 2.11F**).

However, we also noticed that RNA binding is not quantifiable in the presence of AMPPNP. This is in contrast to the NMR experiments, which revealed a partially closed conformation for eIF4A, Dbp5, and Ded1p in the presence of AMPPNP (**Figure 2.5**), for which a significant increase in RNA affinity would be expected. Very weak binding of ssRNA in the presence of AMPPNP is observed for eIF4A (**Figure 2.11A**), yet a k_D could not be extracted and binding may be too slow to be detected using fluorescence anisotropy.

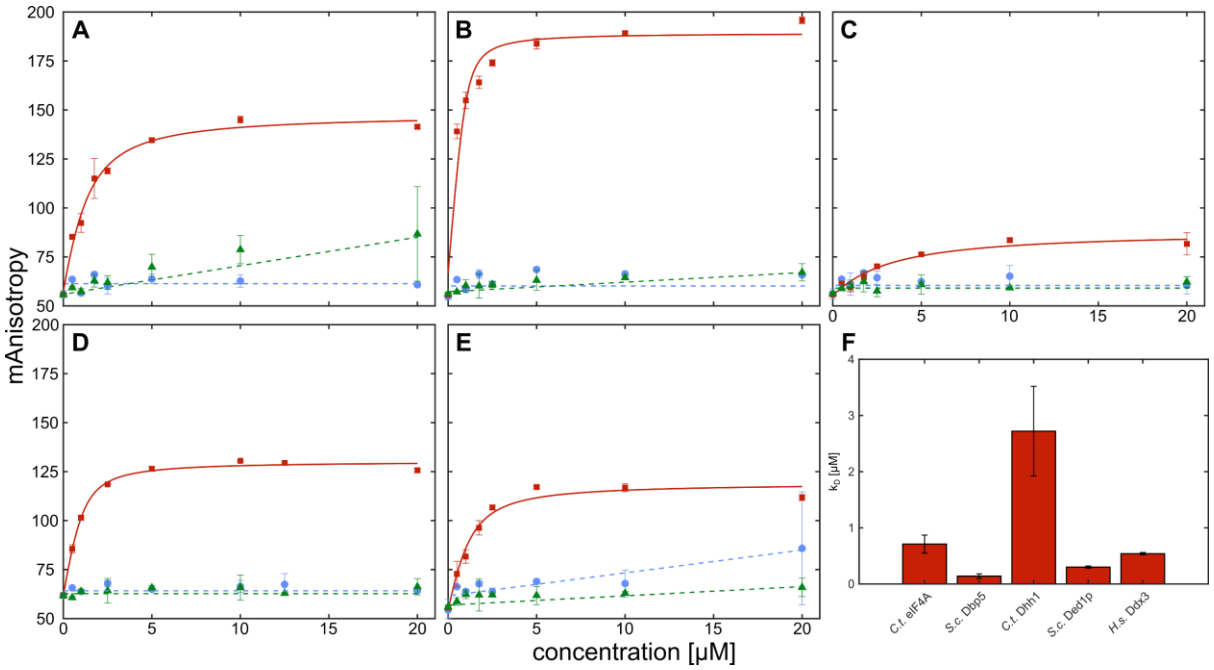


Figure 2.11: RNA Binding curves obtained from fluorescence anisotropy binding experiments. Binding of a 12mer ssRNA was tracked in the absence (blue dots) of a nucleotide, in the presence of AMPPNP (green triangles), and in the presence of ADP-BeF_3 (red squares) for (**A**) eIF4A, (**B**) Dbp5, (**C**) Dhh1, (**D**) Ded1p, and (**E**) DDX3. Each experiment was performed in technical triplicates at 25°C. Fit of RNA binding in the presence of ADP-BeF_3 is visualized by red curves. Blue and green dashed lines are linear fits. (**F**) Dissociation constants for each helicase in the presence of ADP-BeF_3 obtained from A-E.

2.3.11 ATPase activity strongly differs between the helicase cores

ATP hydrolysis requires the formation of the closed conformation. Therefore, the ATPase rates are an excellent reporter of the efficiency of the formation of the closed state. The rates reveal a strong difference for the formation of the closed state between the different helicase cores (**Figure 2.12** and **Supplemental Figure 2.13**). In the presence of a 12mer dsRNA, the ATPase turnover rate is substantially higher for Dbp5 (2 min^{-1}) compared to the other helicases, which implies that the closed state of Dbp5 is effectively formed. While the helicase orthologs Ded1p and DDX3 show almost similar rates around 0.5 min^{-1}, hardly any ATPase activity is detectable for eIF4A and Dhh1. Based on the ATPase activity assay, the transition from the open to the closed state is not very efficient for both helicases. The ATPase rates correlate with our NMR data because it was shown that the formation of the closed state with dsRNA and with AMPPNP or ADP-BeF_3 was either not possible or the closed state was not detectable (**Supplemental Figure 2.11** and **Supplemental Figure 2.12**). Especially for eIF4A

and Dhh1, the signal for the open conformation was still detectable in the presence of dsRNA and ADP-BeF_3 (**Figure 2.10**). Taken together, the efficiency of all five DEAD-box RNA helicases to form the closed conformation is very different.

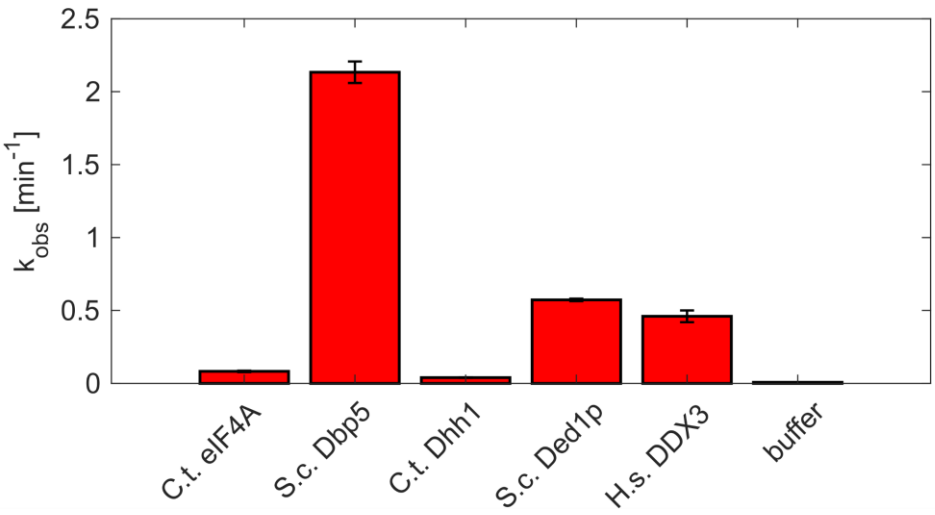


Figure 2.12: ATPase activity. ATPase turnover rates for each helicase core in the presence of a 12mer dsRNA. The average of three technical measurements at 25°C is shown.

2.4 Discussion

DEAD-box RNA helicases are found ubiquitously in all kingdoms of life and participate in all steps of the RNA metabolism. Therefore, it is of great interest to understand how these helicases function and how substrate RNA and nucleotide binding is mediated, ultimately leading to the formation of the closed helicase state. In this study, we have elucidated the differences in binding of RNA and the non-hydrolysable ATP analogues AMPPNP and ADP-BeF_3 between the core constructs of five different DEAD-box RNA helicases.

Similar to other DEAD-box RNA helicases, no specific interactions between both RecA-like domains are found in the absence of substrate RNA and nucleotides. We conclude that the domains tumble independently in solution and do not form specific contacts. This observation is in agreement with published data (221, 222).

For DEAD-box RNA helicases, the folded and highly conserved core harbors all functional parts of the helicase. Therefore, it is expected that similarities in substrate RNA binding for all these RNA helicases are observed. Such similarities were found

for the investigated DEAD-box RNA helicases. Nucleotide binding is detected for the RecA1 domains, whereas the RecA2 domains do not contribute to nucleotide binding. Similar residues in the RecA1 domains showed CSPs upon addition of ADP or AMPPNP, demonstrating that both ATP analogues interact at the same binding site (**Figure 2.3** and **Figure 2.4**). The observation that nucleotides bind only to the RecA1 domain is also in agreement with other studies of the core of DEAD-box proteins (209, 223).

Neither dsRNA nor ssRNA was found to interact significantly with the helicase cores in the absence of the ATP analogues AMPPNP and ADP-BeF_3. Small CSPs indicate that very weak binding occurs for ssRNA and dsRNA to Ded1p, while DDX3 showed only very weak binding of ssRNA.

In contrast to our observations, some DEAD-box RNA helicases have been shown to interact with RNA in the absence of nucleotides. A study of the RNA helicase Hera revealed that its helicase core, but also the isolated RecA1 domain, interacts with dsRNA with high affinity even in the absence of nucleotides (224). In such cases, studying the isolated RecA domains of DEAD-box proteins, which bind RNA in the absence of nucleotides, may provide more insight into how function is modulated by individual domains.

Contrary to our finding that apo DDX3 does not interact with dsRNA, Song and Ji solved a crystal structure of DDX3 with dsRNA in the absence of ATP or non-hydrolysable analogues (225). For this structure, two copies of DDX3 recognize a single dsRNA and each DDX3 molecule binds a single RNA strand. The authors demonstrate that the structure represents the pre-unwound state in the helicase cycle of a DEAD-box protein. Furthermore, they show that the structure is also found in solution based on small-angle X-ray scattering data. However, the observed mode of RNA binding could be due to the fact that the authors used a DDX3 construct that is 25 amino acids longer than the construct used in our study. This extension contains positively charged residues that may contribute to RNA binding and thus could be required for DDX3 RNA binding in the absence of nucleotides. Alternatively, their structure could be a crystallization artifact. A similar observation was made in a study using the bacterial DEAD-box RNA helicase DbpA (208). In this study, a 2:2 RNA:protein ratio was crystallized in the absence of an ATP analogue. Our NMR data

strongly indicate that the crystallized structure does not form in solution, as we could not detect any interactions between RNA and DDX3 in the absence of nucleotides (**Figure 2.8** and **Supplemental Figure 2.10**). Therefore, performing NMR binding experiments with the DDX3 construct used by Song and Ji would be necessary to prove their claim that they have solved the naturally occurring structure of a pre-unwound state of a DEAD-box helicase.

Our observations indicate that prior to RNA binding to the helicase cores, nucleotide binding is required in order to enhance the RNA affinity of a DEAD-box RNA helicase. For the helicases studied here, RNA binding and transition to the closed conformation are only observed in the presence of ATP analogues. In addition, our data revealed that the non-hydrolysable ATP analogue AMPPNP is capable of partially inducing the closed conformation for most helicases in combination with ssRNA (**Figure 2.6**). However, AMPPNP fails to induce a conformational transition in the cores for dsRNA (**Figure 2.9**). Consequently, AMPPNP was shown to be a non-ideal ATP analogue to induce a transition from the open to the closed helicase conformation. In contrast, the use of ADP-BeF_3 induced the partial formation of the closed state for both ss- and dsRNA in all cases. Therefore, our data suggest that ADP-BeF_3 is a suitable ATP analogue for the investigation of the closed state formation for the chosen DEAD-box RNA helicases. However, it should be noted that AMPPNP has been successfully used in most structural biology studies to obtain the closed conformation, and most published helicase structures have been solved in the presence of AMPPNP. Nevertheless, our study implies that the usefulness of different NTP analogues may differ for different DEAD-box RNA helicases.

Consistent with this finding, the fluorescence anisotropy data confirm the results of the NMR titration experiments (**Figure 2.11**). Using fluorescence anisotropy, ssRNA binding was only detectable in the presence of ADP-BeF_3. Again, no affinity for RNA could be determined in the absence of nucleotides or in the presence of AMPPNP. This further emphasizes that RNA is only bound by the closed conformation for the studied helicases.

Interestingly, the ATPase activity was significantly different for all helicases. This indicates that, despite structural similarity and the same conserved catalytic motifs, these helicases show different behavior with respect to the helical unwinding activity

of dsRNA. In the case of Dhh1, for which a lower activity was observed, N- and C-terminal auxiliary domains may be required to be fully functional. However, the reduced activity of Dhh1 could also be related to a lower binding affinity to RNA (**Figure 2.11**).

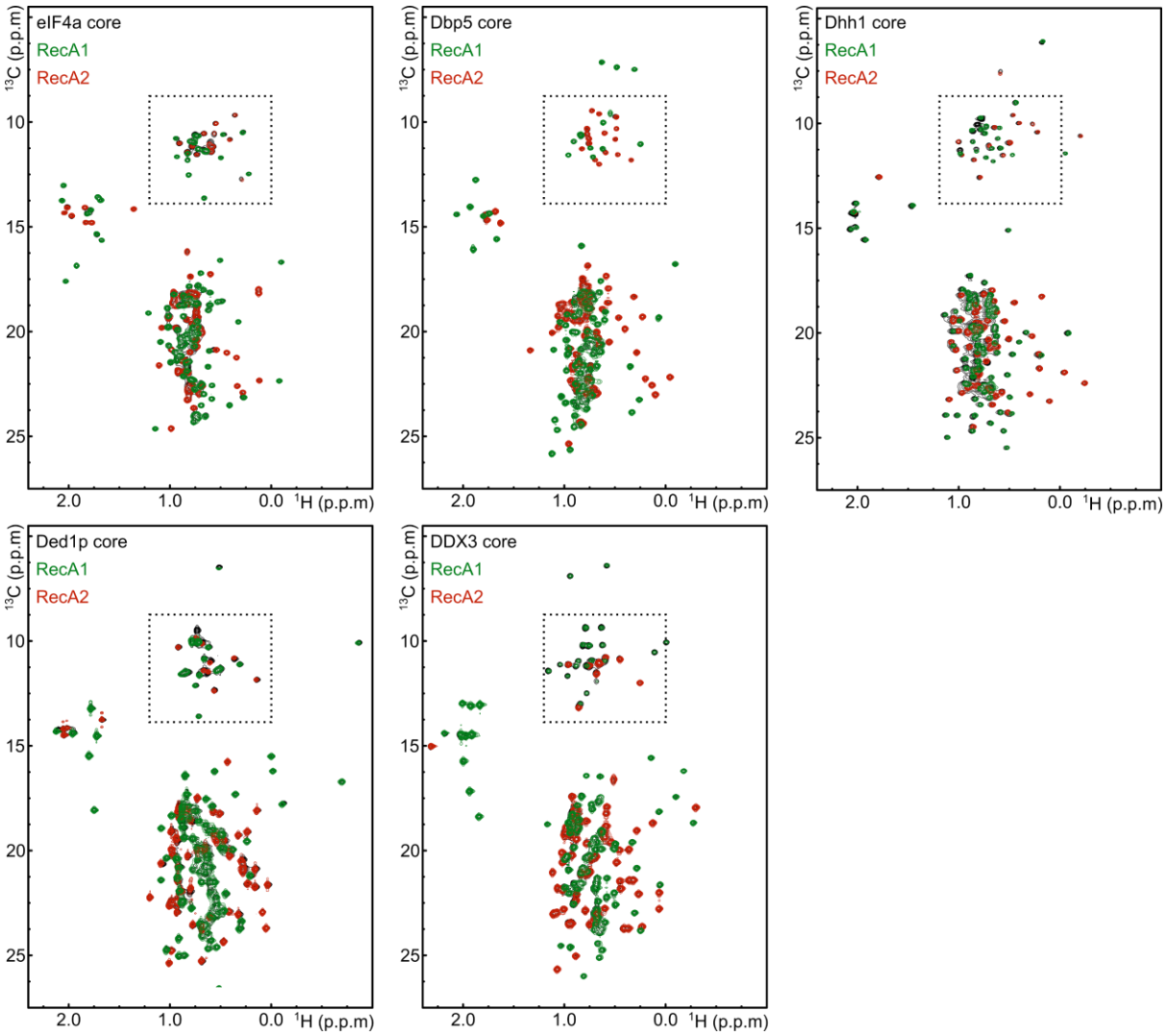
In summary, our results indicate that nucleotide binding is necessary to modulate RNA affinity to induce conformational changes in DEAD-box RNA helicases. Our results provide insight into the differences in the formation of the closed helicase core that are caused by two different ATP analogues using five different DEAD-box RNA helicases from different organisms. Our work highlights that even small differences in the structure of ATP analogues can have a significant influence on the conformational transition from the open to the closed conformation in DEAD-box RNA helicases.

In conclusion, the core structures of DEAD-box proteins differ in their RNA and nucleotide binding affinities, although they all contain conserved motifs for RNA- and nucleotide binding as well as ATP hydrolysis (**Supplemental Figure 2.1**). The mechanism by which different helicases recognize RNA substrates may be fine-tuned differently among the different DEAD-box RNA helicases. This needs to be further elucidated. For DEAD-box proteins that exhibit different substrate and nucleotide binding behaviors, the study of the isolated RecA-like domains will contribute to a better understanding of how this protein class show such differences.

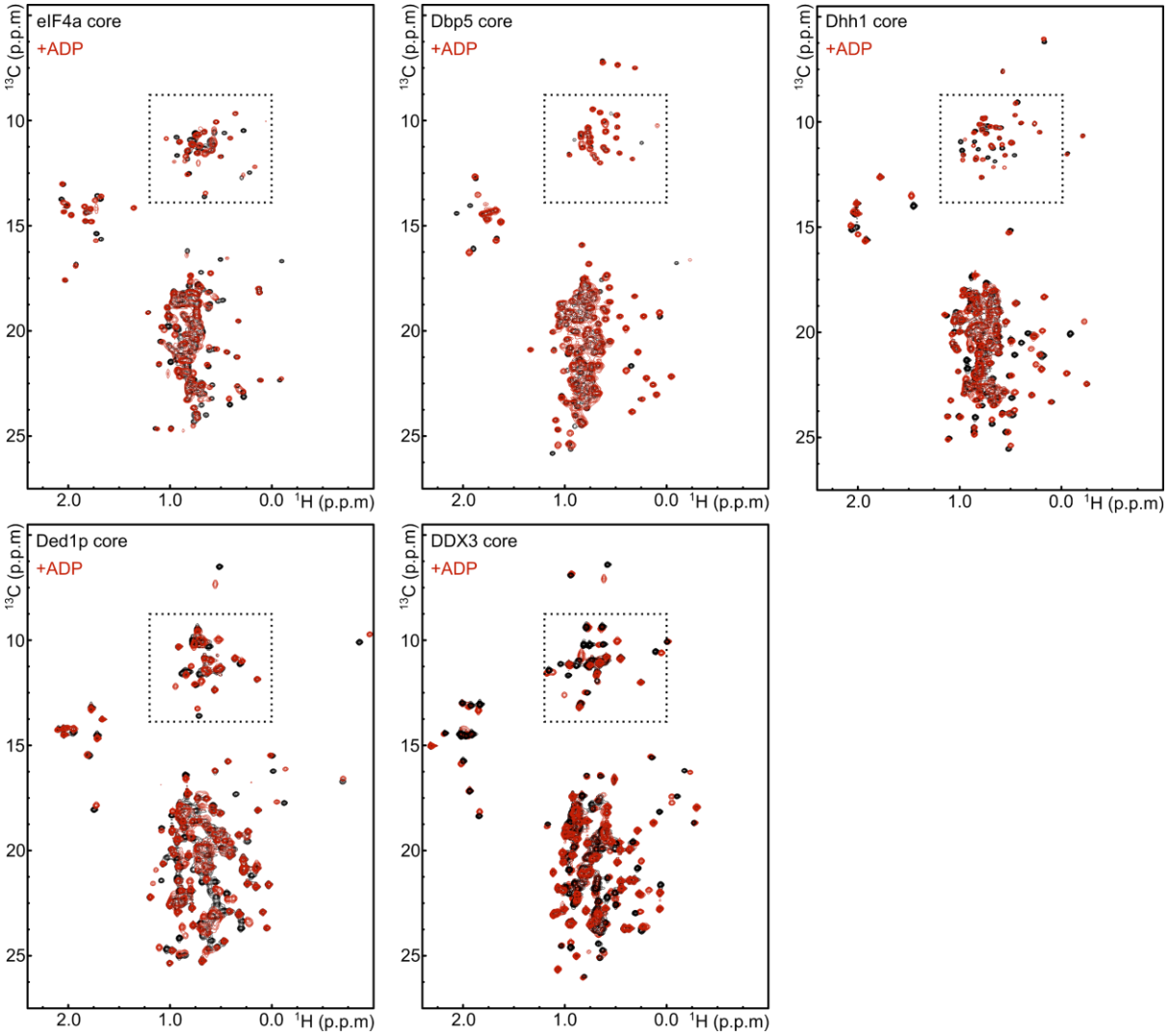
2.5 Supplement

```
                                                                  Q                   I
Ded1p      RSGGRWIDGKH-----VPAPRNEKAEIAIFGVPEDPNFQSSGINFDNYDDIPVDASGKDV 137
DDX3       --NSRWCDKSDEDDWSKPLPPSERLEQELFS------GGNTGINFEKYDDIPVEATGNNC 174
Dbp5       ------------SEYEVKVK-----------------------LA-----DIQADPNSP 19
eIF4A      ------------------------------------------------EGQIE-SNY 8
Dhh1       -------------DWKKGLN-----------------------IPAK-DTRTQTE-DVT 37

                                                                  Q                   I
Ded1p      PEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYSVPIVAN--GRDLMACAQTGSGKTGG 195
DDX3       PPHIESFSDVEMGEIIMGNIELTRYTRPTPVQKHAIPIIKE--KRDLMACAQTGSGKTAA 232
Dbp5       LYSAKSFDELGLAPELLKGIYAMKFQKPSKIQERALPLLLHNPPRNMIAQSQSGTGKTAA 148
eIF4A      DETVDSFDEMNLKPELLRGIYAYGFERPSAIQQRAIMPVIK--GHDVIAQAQSGTGKTAT 78
Dhh1       NTKGLEWEDFNLKRDLLKGIFEAGYEKPSPIQEESIPIALA--GRDILARAKNGTGKTAA 95

                                                                  Ia
Ded1p      FLFPVLSESFKTGPSPQP---ESQGSFYQRKAYPTAVIMAPTRELATQIFDEAKKFTYRS 252
DDX3       FLLPILSQIYSDGPGEALRAMKENGRYGRRKQYPISLVLAPTRELAVQIYEEARKFSYRS 292
Dbp5       FSLTMLTRVNPE------------------DASPQAICLAPSRELARQTLEVVQEMGKFT 190
eIF4A      FSISVLQKIDPN------------------LKQCQALILAPTRELAQQIQKVVVAIGDFM 120
Dhh1       FVIPALEKINPK------------------VSKIQCLILVPTRELAMQTSQVCKILGKHL 137

                          Ib                  Ic                          II
Ded1p      WVKACVVYGGSPIGNQLREIERGCDLLVATPGRLNDLLERGKISLANVKYLVLDEADRML 312
DDX3       RVRPCVVYGGADIGQQIRDLERGCHLLVATPGRLVDMMERGKIGLDFCKYLVLDEADRML 352
Dbp5       KITSQLIVPDSFE----KNKQINAQVIVGTPGTVLDLMRRKLMQLQKIKIFVLDEADNML 246
eIF4A      NVECHACIGGTSVREDMKALQEGPQIVVGTPGRVHDMIQRRFLKTDAMKMFVLDEADEML 180
Dhh1       GVNVMVTTGGTGLRDDIVRLQDAVHIVVGTPGRILDLASKQVADLSECPMFIMDEADKLL 197

                                      III
Ded1p      DM-GFEPQIRHIVEDCDMTPVGERQTLMFSATFPADIQHLARDFLSDYIFLSVGRVGSTS 371
DDX3       DM-GFEPQIRRIVEQDTMPPKGVRHTMMFSATFPKEIQMLARDFLDEYIFLAVGRVGSTS 411
Dbp5       DQQGLGDQCIRVKRFL----PKDTQLVLFSATFADAVRQYAKKIVPNANTLELQTNEVNV 302
eIF4A      SR-GFTEQIYDIFQLL----PQSTQVVLLSATMPQDVLEVTTKFMRDPVRILVKKDELTL 235
Dhh1       SP-EFTPVIEQLLQFH----PKDRQVMLFSATFPISVKAFSDNNMRDPYEINL-MDELTL 251

                                                IV
Ded1p      ENITQKVLYVENQDKKSALLDLLSAS-TDGLTLIFVETKRMADQLTDFLIMQNFRATAIH 430
DDX3       ENITQKVVWVEESDKRSFLLDLLNATGKDSLTLVFVETKKGADSLEDFLYHEGYACTSIH 471
Dbp5       DAIKQLYMDCKNEADKFDVLTELYGLMTIGSSIIFVATKKTANVLYGKLKSEGHEVSILH 362
eIF4A      EGIKQFYIAVEKEEWKLDTLSDLYETVTITQAVIFCNTRRKVDWLTEKLTQRDFTVSAMH 295
Dhh1       RGITQYYAYVEE-RQKVHCLNTLFSKLQINQSIIFCNSTNRVELLAKKITELGYSCFYSH 310

                    IVa                 V           Va/b
Ded1p      GDRTQSERERALAAFRSGAATLLVATAVAARGLDIPNVTHVINYDLPSD------VDDYV 484
DDX3       GDRSQRDREEALHQFRSGKSPILVATAVAARGLDISNVKHVINFDLPSD------IEEYV 525
Dbp5       GDLQTQERDRLIDDFREGRSKVLITTNVLARGIDIPTVSMVVNYDLPTLANGQADPATYI 422
eIF4A      GDMDQAQRDLIMKEFRSGSSRVLIATDLLARGIDVQQVSLVINYDLPTN------RENYI 349
Dhh1       AKMPQAARNRVFHDFRNGVCRNLVCSDLLTRGIDIQAVNVVINFDFPKN------AETYL 364

                VI
Ded1p      HRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILT----EANQEVPSFLKDAMMS----- 535
DDX3       HRIGRTGRVGNLGLATSFFNERNINITKDLLDLLV----EAKQEVPSWLENMAYEHHYKG 582
Dbp5       HRIGRTGRFGRKGVAISFVHDKNSFNILSAIQKYFGDIEMTRVPTDDW----DEVEKIVK 478
eIF4A      HRIGRGGRFGRKGVAINFVTAEDVRMMREIEQFYSTQIEEMPMNVADLI----------- 397
Dhh1       HRIGRSGRYGHLGLAINLINWDDRFNLYNIERELGTEIHPIPAEIPKNLYVYENPETIPR 424


Ded1p      ------ 535
DDX3       ------ 582
Dbp5       KVLKD- 482
eIF4A      ------ 397
Dhh1       PISTFK 430
```

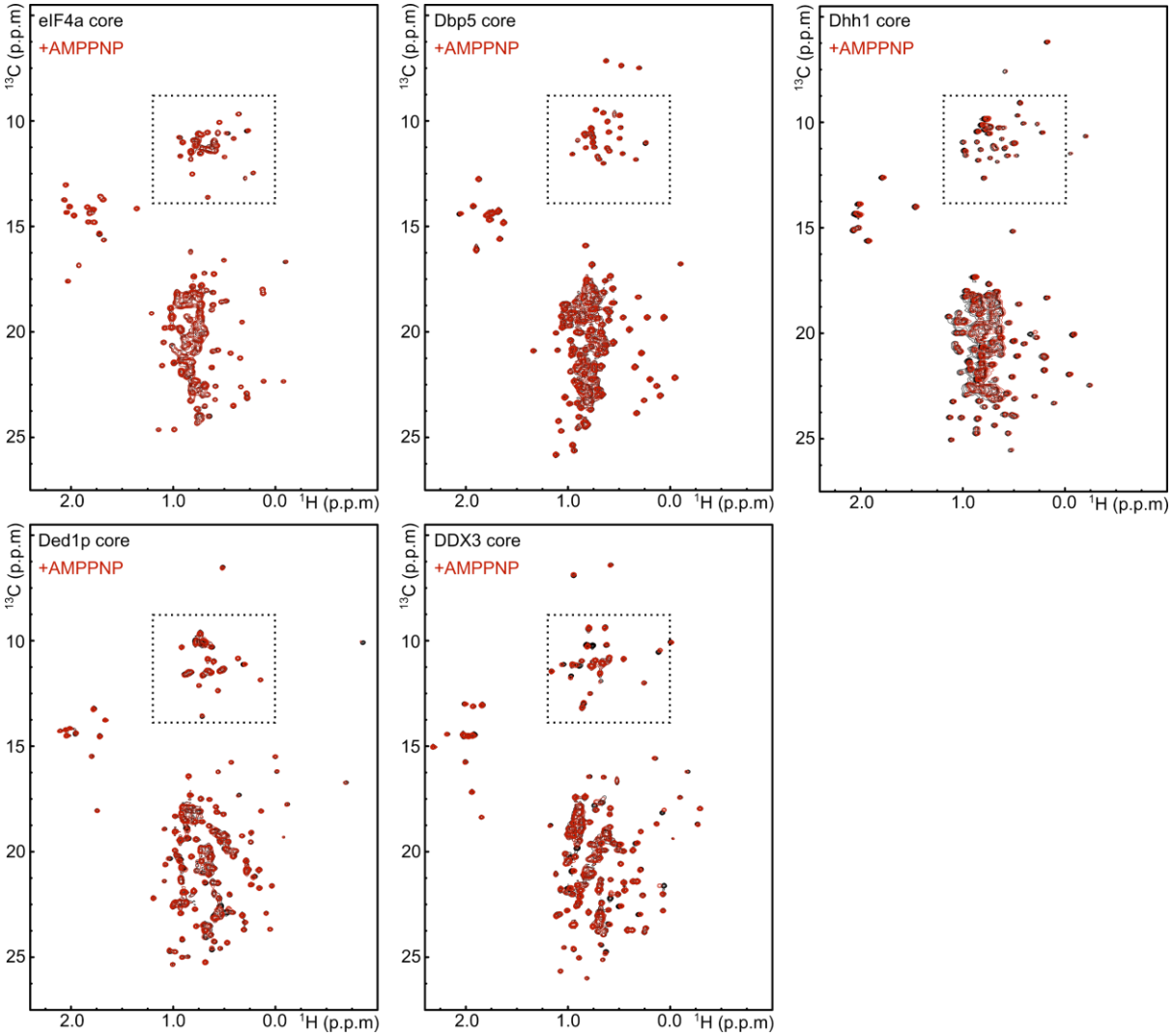
Supplemental Figure 2.1: Multiple sequence alignment of studied members of the DEAD-box RNA helicase family. The highly conserved sequence motifs are numbered and indicated by colored boxes. Red boxes: motifs are responsible for ATP binding and hydrolysis; green boxes: motifs coordinate nucleic acid and nucleotide binding; blue boxes: motifs are located in the nucleic acid binding site.



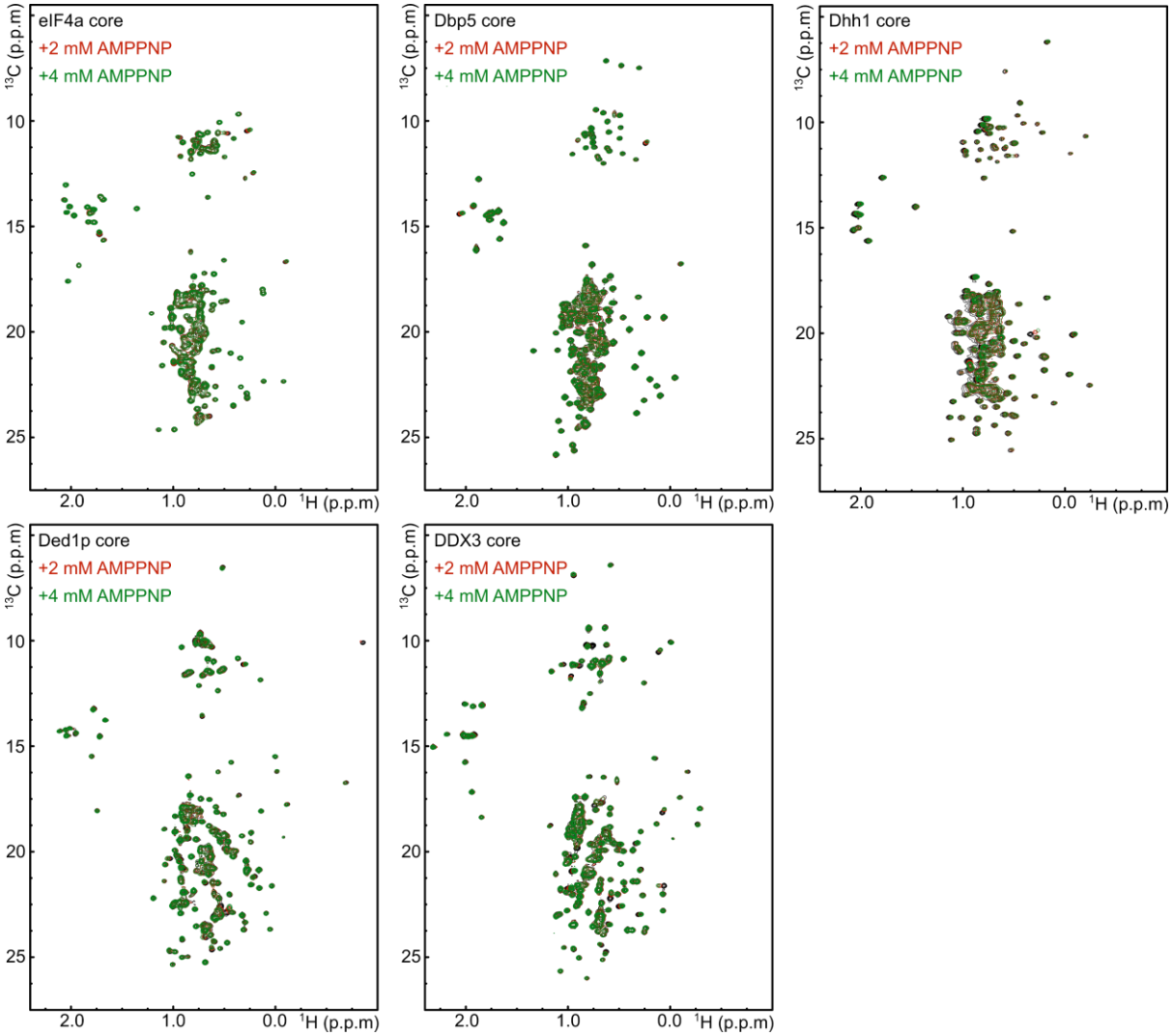
Supplemental Figure 2.2: RecA-like domains tumble independently in solution. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core (black), RecA1 (green), and RecA2 (red), respectively. The positions of the spectra shown in **Figure 2.2** are highlighted by dashed rectangles.



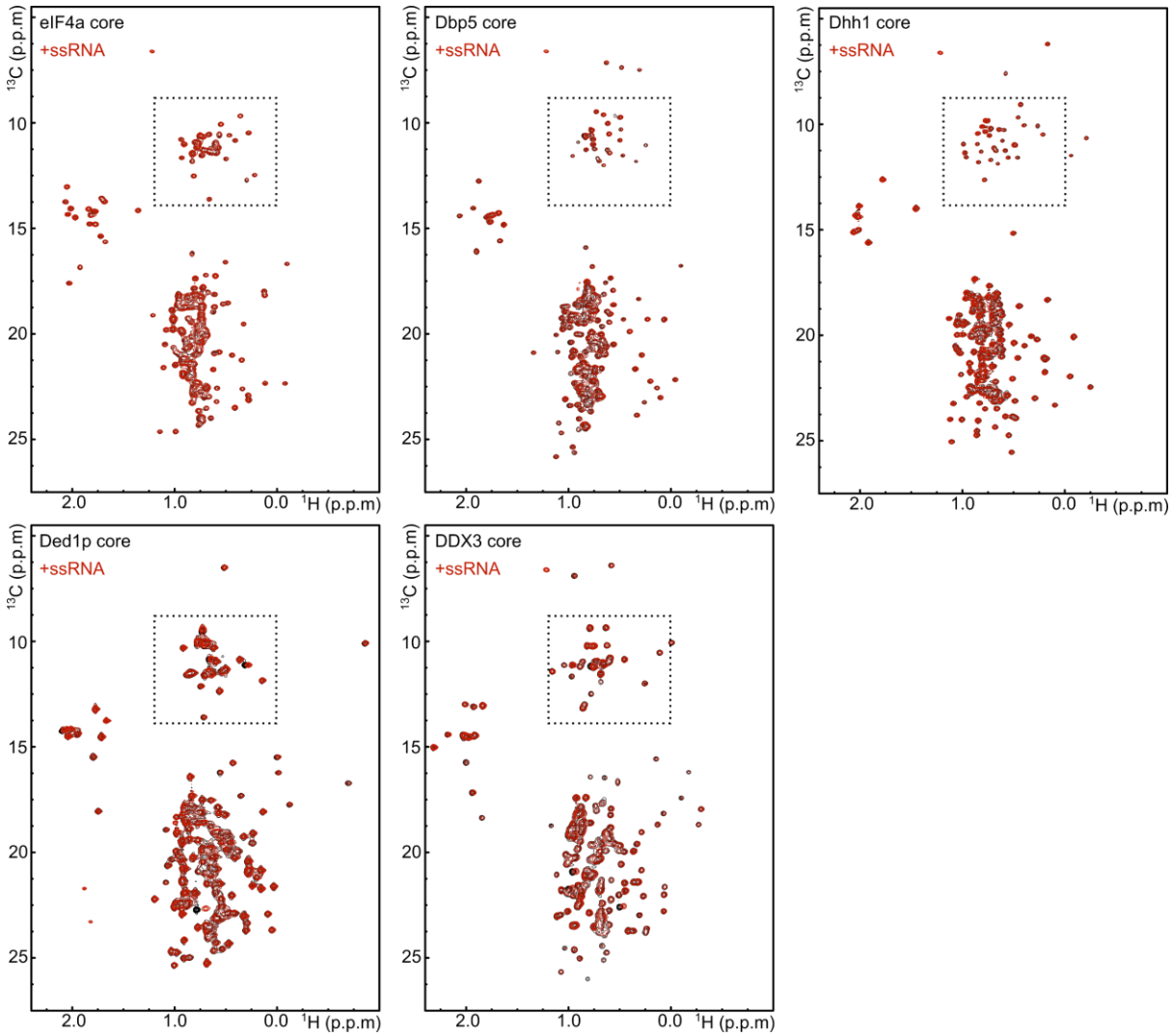
Supplemental Figure 2.3: ADP is bound by the RecA1 domain. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of ADP. The positions of the spectra shown in **Figure 2.3** are highlighted by dashed rectangles.



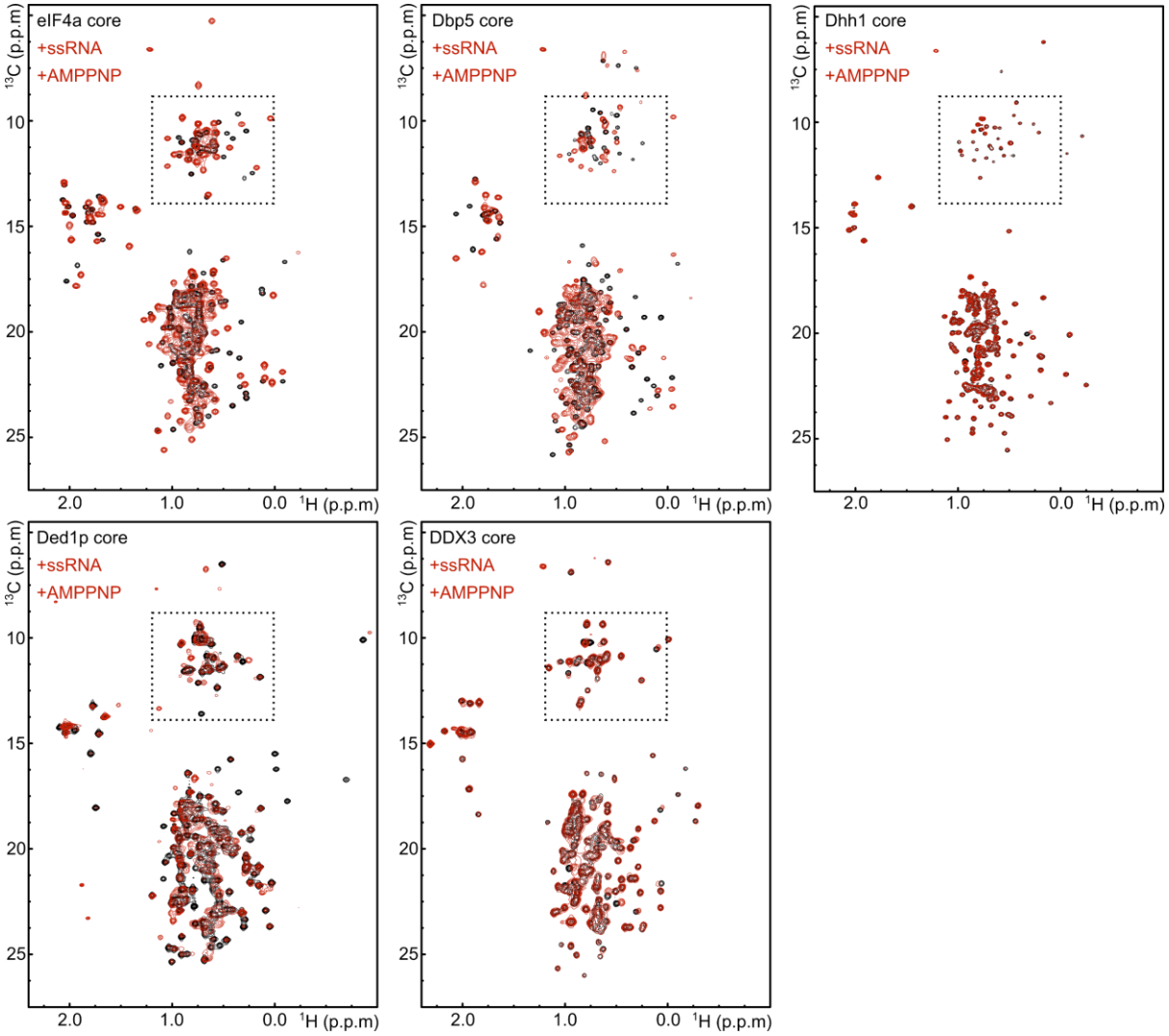
Supplemental Figure 2.4: AMPPNP is weakly bound by the RecA1 domain. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of AMPPNP. The positions of the spectra shown in **Figure 2.4** are highlighted by dashed rectangles.



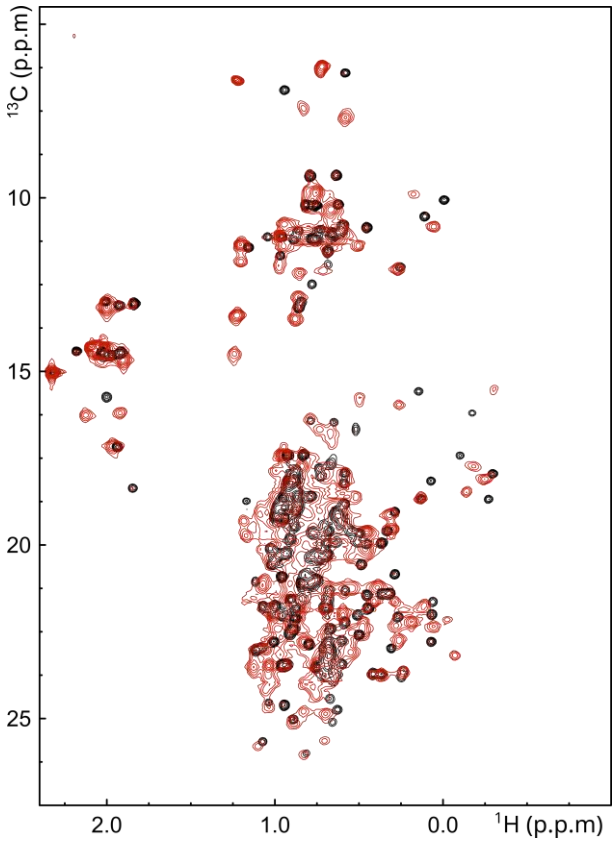
Supplemental Figure 2.5: Titration of AMPPNP. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after addition of 2 mM (red) and 4 mM (green) AMPPNP.



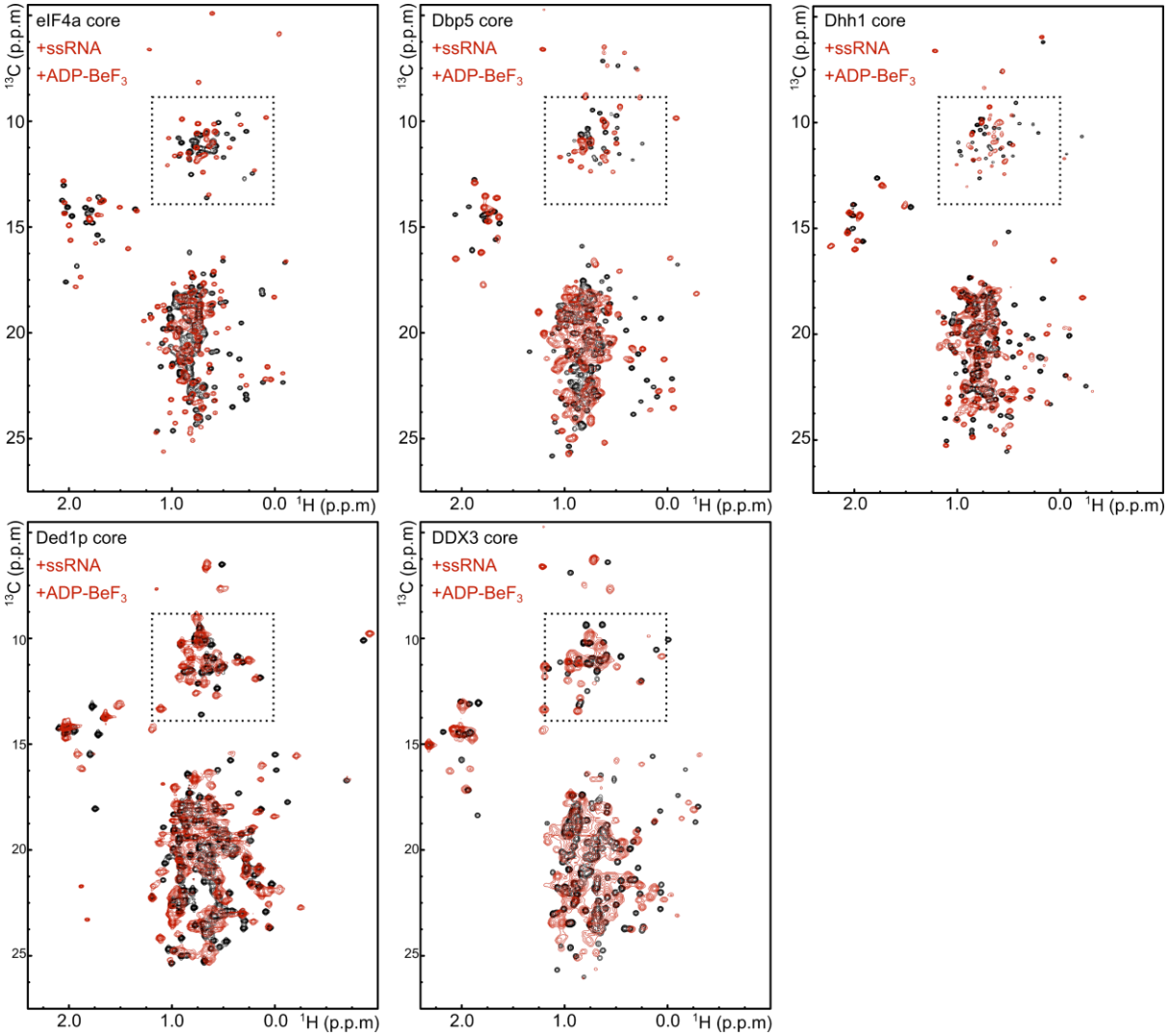
Supplemental Figure 2.6: Single-stranded RNA is not bound by the helicase cores in the absence of nucleotides. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer ssRNA. The positions of the spectra shown in **Figure 2.5** are highlighted by dashed rectangles.



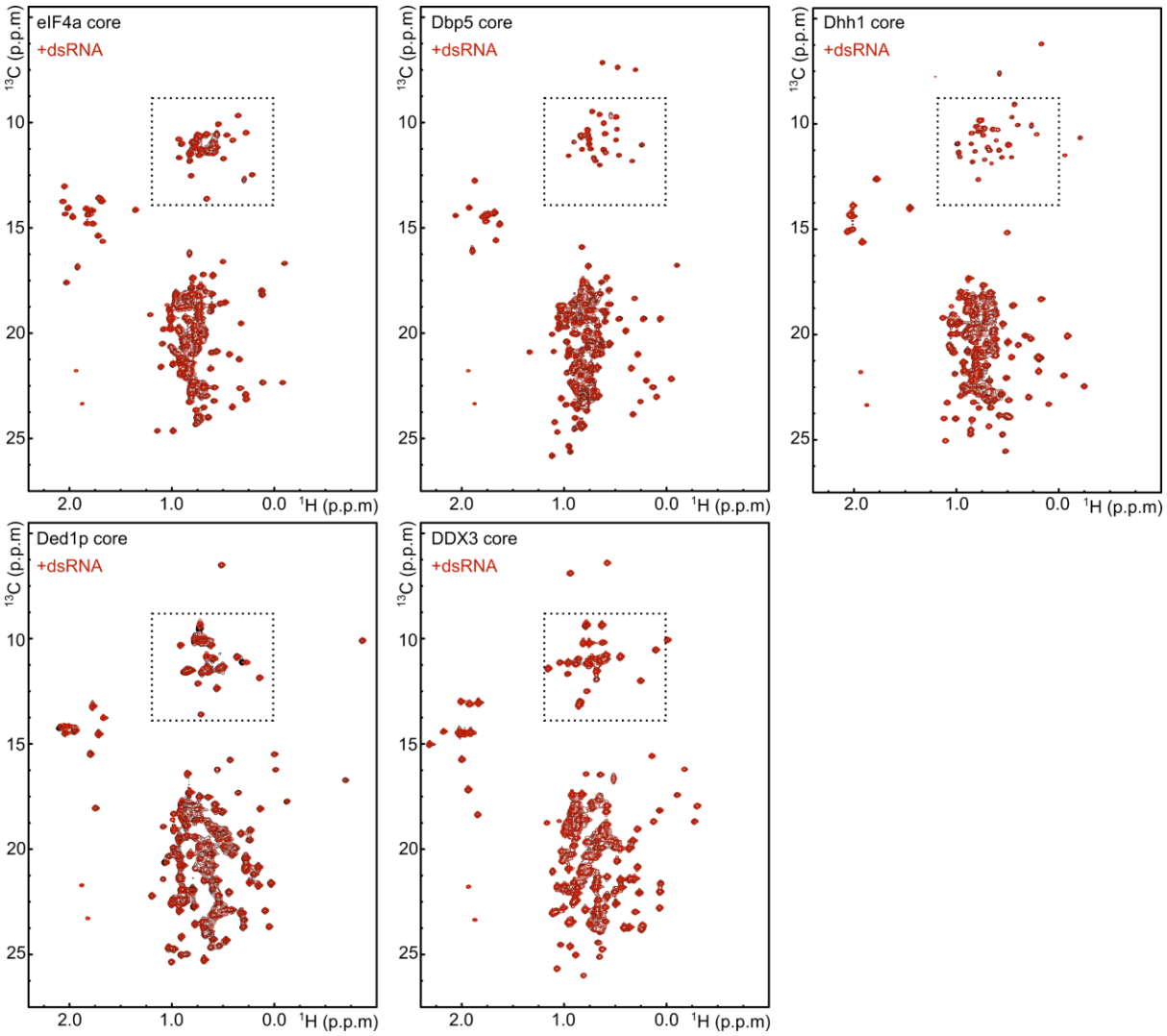
Supplemental Figure 2.7: Single-stranded RNA and AMPPNP induce a partial transition from the open to the closed conformation. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of AMPPNP followed by addition of a 12mer ssRNA. The positions of the spectra shown in **Figure 2.6** are highlighted by dashed rectangles.



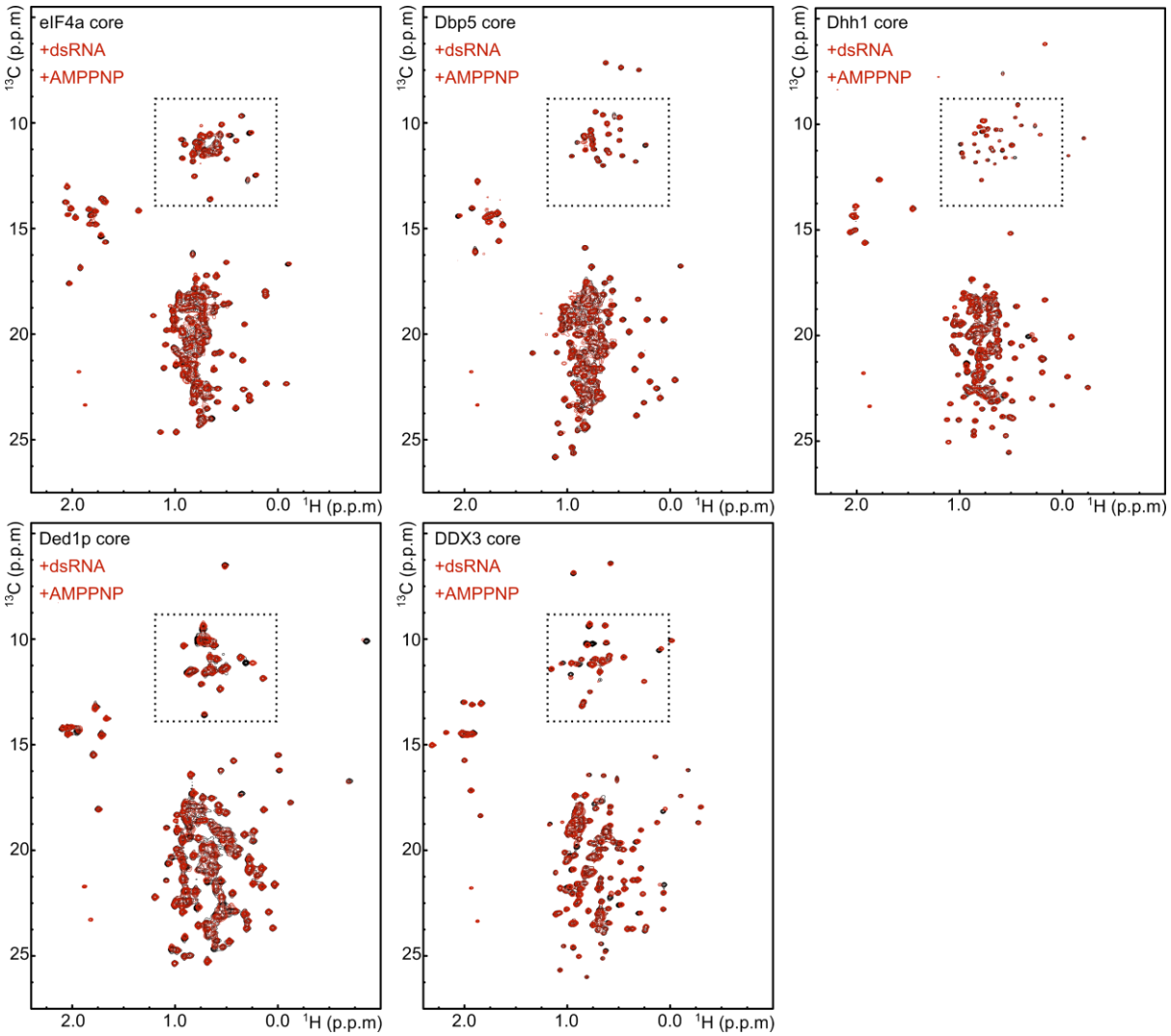
Supplemental Figure 2.8: DDX3 reaches the closed conformation after 12 h incubation. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled DDX3 before (black) and after (red) addition of a 12mer ssRNA followed by addition of AMPPNP.



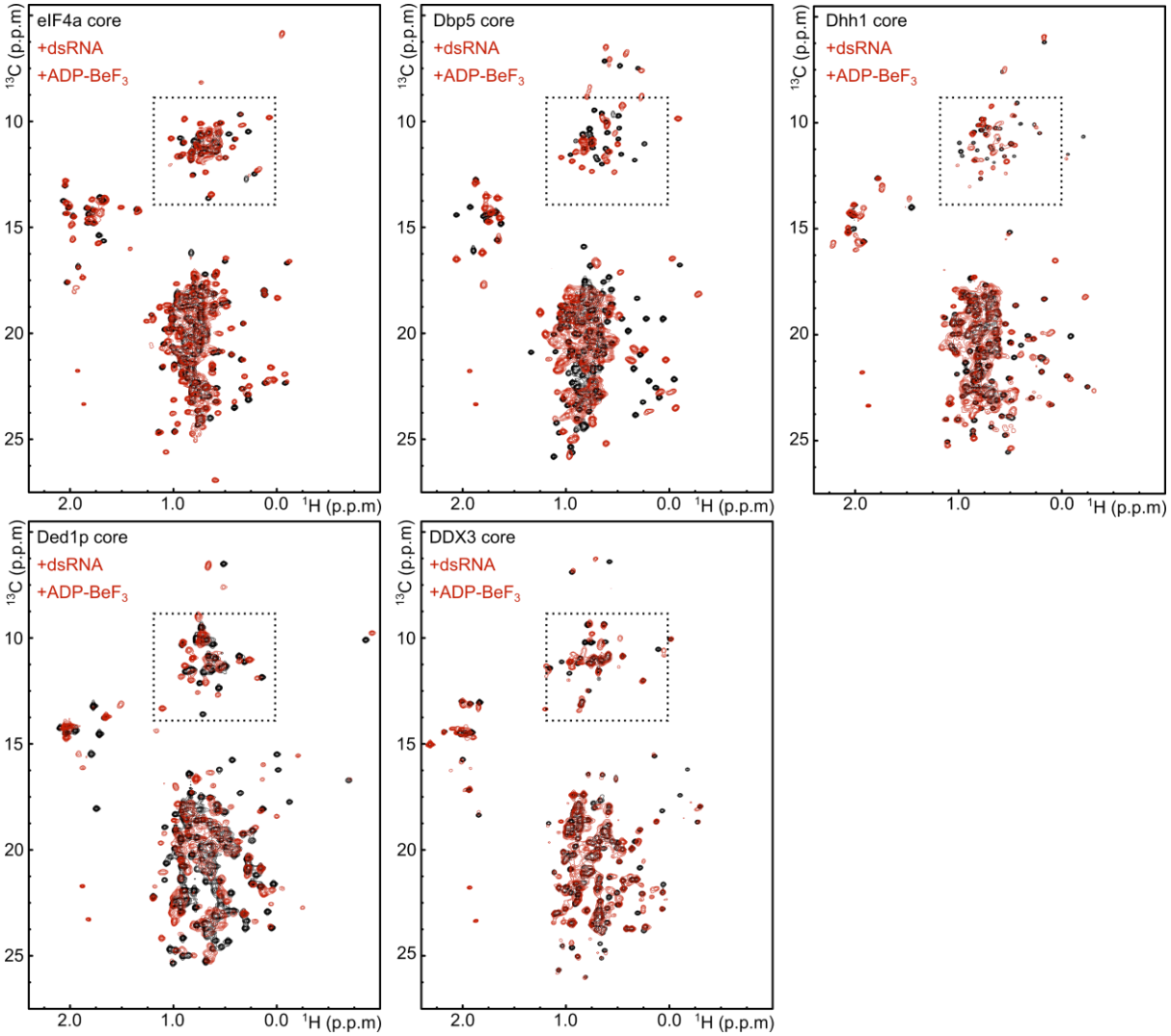
Supplemental Figure 2.9: Single-stranded RNA and ADP-BeF_3 induce the transition from the open to the closed conformation. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of ADP-BeF_3 followed by addition of a 12mer ssRNA. The positions of the spectra shown in **Figure 2.7** are highlighted by dashed rectangles.



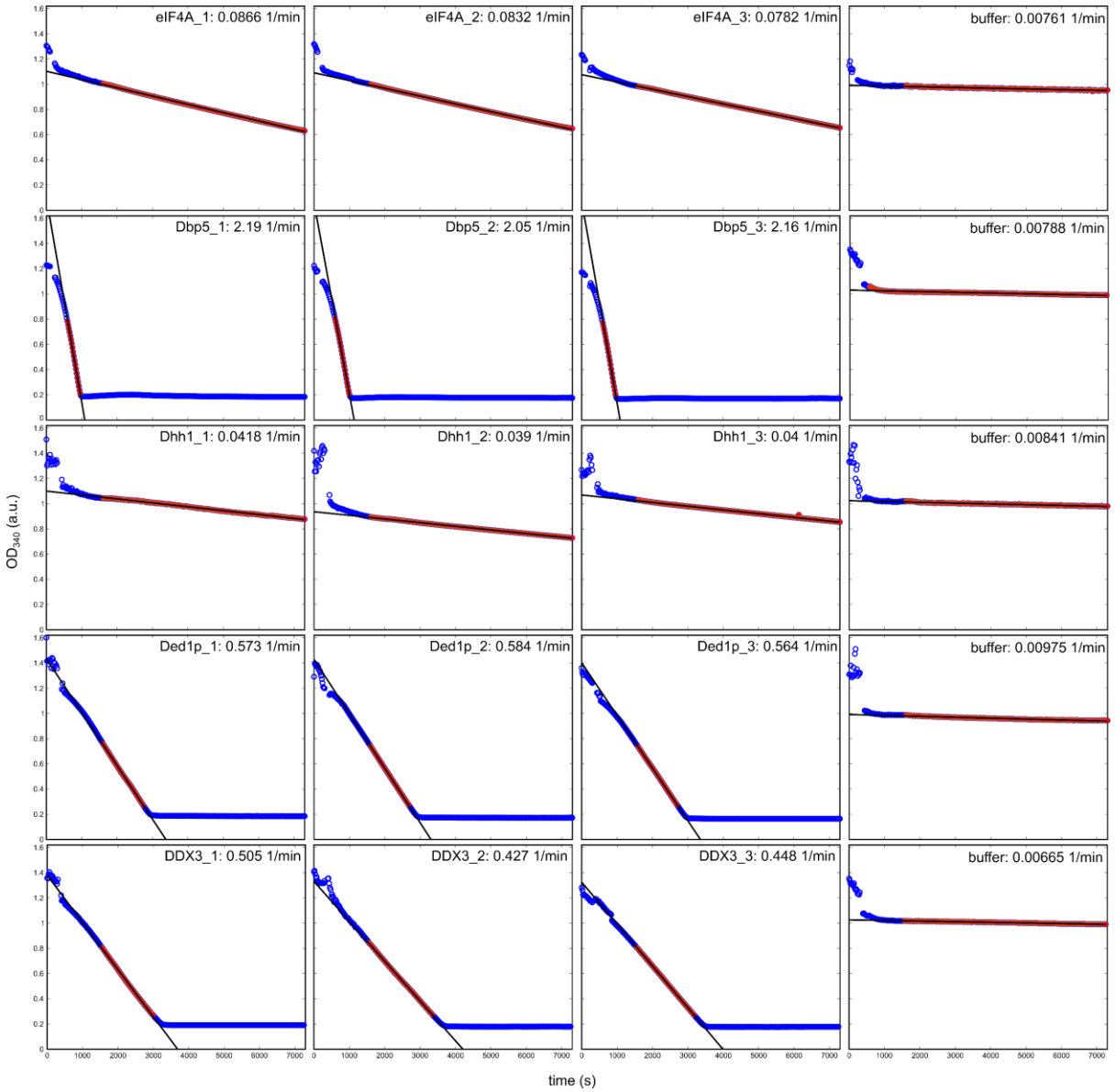
Supplemental Figure 2.10: Double-stranded RNA is not bound by the helicase cores in the absence of nucleotides. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of a 12mer dsRNA. The positions of the spectra shown in **Figure 2.8** are highlighted by dashed rectangles.



Supplemental Figure 2.11: Double-stranded RNA and AMPPNP do not induce the transition from the open to the closed conformation. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of AMPPNP followed by addition of a 12mer dsRNA. The positions of the spectra shown in **Figure 2.9** are highlighted by dashed rectangles.



Supplemental Figure 2.12: Double-stranded RNA and ADP-BeF_3 partially induce the transition from the open to the closed conformation. Overlay of ^{1}H-^{13}C methyl TROSY spectra of ILVM-labeled core before (black) and after (red) addition of ADP-BeF_3 followed by addition of a 12mer dsRNA. The positions of the spectra shown in **Figure 2.10** are highlighted by dashed rectangles.



Supplemental Figure 2.13: ATPase activity of the different DEAD-box RNA core constructs in the presence of a 12mer hybrid RNA. The linear regimes of the absorption time course (red circles) were fitted using a linear equation that is represented as a black line.

Supplemental Table 2.1: Protein constructs used in this study. All proteins start either with an N-terminal GG or GA that results from TEV cleavage.

Protein	**Sequence**	**Internal Ref.**
C. thermophilum Dhh1 core (17-430)	GGDWKKGLNIPAKDTRTQTEDVTNTKGLEWEDFNLKRDLLKGIFEAGYEKPSPIQEESI PIALAGRDILARAKNGTGKTAAFVIPALEKINPKVSKIQCLILVPTRELAMQTSQVCKI LGKHLGVNVMVTTGGTGLRDDIVRLQDAVHIVVGTPGRILDLASKQVADLSECPMFIMD EADKLLSPEFTPVIEQLLQFHPKDRQVMLFSATFPISVKAFSDNNMRDPYEINLMDELT LRGITQYYAYVEERQKVHCLNTLFSKLQINQSIIFCNSTNRVELLAKKITELGYSCFYS HAKMPQAARNRVFHDFRNGVCRNLVCSDLLTRGIDIQAVNVVINFDFPKNAETYLHRIG RSGRYGHLGLAINLINWDDRFNLYNIERELGTEIHPIPAEIPKNLYVYENPETIPRPIS TFK	#1707
C. thermophilum Dhh1 RecA1 (17-247)	GGDWKKGLNIPAKDTRTQTEDVTNTKGLEWEDFNLKRDLLKGIFEAGYEKPSPIQEESI PIALAGRDILARAKNGTGKTAAFVIPALEKINPKVSKIQCLILVPTRELAMQTSQVCKI LGKHLGVNVMVTTGGTGLRDDIVRLQDAVHIVVGTPGRILDLASKQVADLSECPMFIMD EADKLLSPEFTPVIEQLLQFHPKDRQVMLFSATFPISVKAFSDNNMRDPYEINLMD	#1814
C. thermophilum Dhh1 RecA2 (246-421)	GAMDELTLRGITQYYAYVEERQKVHCLNTLFSKLQINQSIIFCNSTNRVELLAKKITEL GYSCFYSHAKMPQAARNRVFHDFRNGVCRNLVCSDLLTRGIDIQAVNVVINFDFPKNAE TYLHRIGRSGRYGHLGLAINLINWDDRFNLYNIERELGTEIHPIPAEIPKNLYVYENPE T	#1819
C. thermophilum eIF4A core (12-397)	GGEGQIESNYDETVDSFDEMNLKPELLRGIYAYGFERPSAIQQRAIMPVIKGHDVIAQA QSGTGKTATFSISVLQKIDPNLKQCQALILAPTRELAQQIQKVVVAIGDFMNVECHACI GGTSVREDMKALQEGPQIVVGTPGRVHDMIQRRFLKTDAMKMFVLDEADEMLSRGFTEQ IYDIFQLLPQSTQVVLLSATMPQDVLEVTTKFMRDPVRILVKKDELTLEGIKQFYIAVE KEEWKLDTLSDLYETVTITQAVIFCNTRRKVDWLTEKLTQRDFTVSAMHGDMDQAQRDL IMKEFRSGSSRVLIATDLLARGIDVQQVSLVINYDLPTNRENYIHRIGRGGRFGRKGVA INFVTAEDVRMMREIEQFYSTQIEEMPMNVADLI	#1708
C. thermophilum eIF4A RecA1 (12-230)	GGEGQIESNYDETVDSFDEMNLKPELLRGIYAYGFERPSAIQQRAIMPVIKGHDVIAQA QSGTGKTATFSISVLQKIDPNLKQCQALILAPTRELAQQIQKVVVAIGDFMNVECHACI GGTSVREDMKALQEGPQIVVGTPGRVHDMIQRRFLKTDAMKMFVLDEADEMLSRGFTEQ IYDIFQLLPQSTQVVLLSATMPQDVLEVTTKFMRDPVRILVKKD	#1742
C. thermophilum eIF4A RecA2 (230-397)	GGDELTLEGIKQFYIAVEKEEWKLDTLSDLYETVTITQAVIFCNTRRKVDWLTEKLTQR DFTVSAMHGDMDQAQRDLIMKEFRSGSSRVLIATDLLARGIDVQQVSLVINYDLPTNRE NYIHRIGRGGRFGRKGVAINFVTAEDVRMMREIEQFYSTQIEEMPMNVADLI	#1743
S. cerevisiae Dbp5 core (69-482)	GGSEYEVKVKLADIQADPNSPLYSAKSFDELGLAPELLKGIYAMKFQKPSKIQERALPL LLHNPPRNMIAQSQSGTGKTAAFSLTMLTRVNPEDASPQAICLAPSRELARQTLEVVQE MGKFTKITSQLIVPDSFEKNKQINAQVIVGTPGTVLDLMRRKLMQLQKIKIFVLDEADN MLDQQGLGDQCIRVKRFLPKDTQLVLFSATFADAVRQYAKKIVPNANTLELQTNEVNVD AIKQLYMDCKNEADKFDVLTELYGLMTIGSSIIFVATKKTANVLYGKLKSEGHEVSILH GDLQTQERDRLIDDFREGRSKVLITTNVLARGIDIPTVSMVVNYDLPTLANGQADPATY IHRIGRTGRFGRKGVAISFVHDKNSFNILSAIQKYFGDIEMTRVPTDDWDEVEKIVKKV LKD	#1713
S. cerevisiae Dbp5 RecA1 (69-299)	GGSEYEVKVKLADIQADPNSPLYSAKSFDELGLAPELLKGIYAMKFQKPSKIQERALPL LLHNPPRNMIAQSQSGTGKTAAFSLTMLTRVNPEDASPQAICLAPSRELARQTLEVVQE MGKFTKITSQLIVPDSFEKNKQINAQVIVGTPGTVLDLMRRKLMQLQKIKIFVLDEADN MLDQQGLGDQCIRVKRFLPKDTQLVLFSATFADAVRQYAKKIVPNANTLELQTNE	#2270

S. cerevisiae Dbp5 RecA2 (295-482)	GGQTNEVNVDAIKQLYMDCKNEADKFDVLTELYGLMTIGSSIIFVATKKTANVLYGKLK SEGHEVSILHGDLQTQERDRLIDDFREGRSKVLITTNVLARGIDIPTVSMVVNYDLPTL ANGQADPATYIHRIGRTGRFGRKGVAISFVHDKNSFNILSAIQKYFGDIEMTRVPTDDW DEVEKIVKKVLKD	#2281
H. sapiens DDX3 core (122-582)	GGNSRWCDKSDEDDWSKPLPPSERLEQELFSGGNTGINFEKYDDIPVEATGNNCPPHIE SFSDVEMGEIIMGNIELTRYTRPTPVQKHAIPIIKEKRDLMACAQTGSGKTAAFLLPIL SQIYSDGPGEALRAMKENGRYGRRKQYPISLVLAPTRELAVQIYEEARKFSYRSRVRPC VVYGGADIGQQIRDLERGCHLLVATPGRLVDMMERGKIGLDFCKYLVLDEADRMLDMGF EPQIRRIVEQDTMPPKGVRHTMMFSATFPKEIQMLARDFLDEYIFLAVGRVGSTSENIT QKVVWVEESDKRSFLLDLLNATGKDSLTLVFVETKKGADSLEDFLYHEGYACTSIHGDR SQRDREEALHQFRSGKSPILVATAVAARGLDISNVKHVINFDLPSDIEEYVHRIGRTGR VGNLGLATSFFNERNINITKDLLDLLVEAKQEVPSWLENMAYEHHYKG	#1737
H. sapiens DDX3 RecA1 (122-410)	GGNSRWCDKSDEDDWSKPLPPSERLEQELFSGGNTGINFEKYDDIPVEATGNNCPPHIE SFSDVEMGEIIMGNIELTRYTRPTPVQKHAIPIIKEKRDLMACAQTGSGKTAAFLLPIL SQIYSDGPGEALRAMKENGRYGRRKQYPISLVLAPTRELAVQIYEEARKFSYRSRVRPC VVYGGADIGQQIRDLERGCHLLVATPGRLVDMMERGKIGLDFCKYLVLDEADRMLDMGF EPQIRRIVEQDTMPPKGVRHTMMFSATFPKEIQMLARDFLDEYIFLAVGRVGS	#1736
H. sapiens DDX3 RecA2 (410-582)	GGSTSENITQKVVWVEESDKRSFLLDLLNATGKDSLTLVFVETKKGADSLEDFLYHEGY ACTSIHGDRSQRDREEALHQFRSGKSPILVATAVAARGLDISNVKHVINFDLPSDIEEY VHRIGRTGRVGNLGLATSFFNERNINITKDLLDLLVEAKQEVPSWLENMAYEHHYKG	#1731
S. cerevisiae Ded1p core (82-535)	GGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNFQSSGINFDNYDDIPVDASGKDVPE PITEFTSPPLDGLLLENIKLARFTKPTPVQKYSVPIVANGRDLMACAQTGSGKTGGFLF PVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMAPTRELATQIFDEAKKFTYRSWVKAC VVYGGSPIGNQLREIERGCDLLVATPGRLNDLLERGKISLANVKYLVLDEADRMLDMGF EPQIRHIVEDCDMTPVGERQTLMFSATFPADIQHLARDFLSDYIFLSVGRVGSTSENIT QKVLYVENQDKKSALLDLLSASTDGLTLIFVETKRMADQLTDFLIMQNFRATAIHGDRT QSERERALAAFRSGAATLLVATAVAARGLDIPNVTHVINYDLPSDVDDYVHRIGRTGRA GNTGLATAFFNSENSNIVKGLHEILTEANQEVPSFLKDAMMS	#1740
S. cerevisiae Ded1p RecA1 (82-369)	GGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNFQSSGINFDNYDDIPVDASGKDVPE PITEFTSPPLDGLLLENIKLARFTKPTPVQKYSVPIVANGRDLMACAQTGSGKTGGFLF PVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMAPTRELATQIFDEAKKFTYRSWVKAC VVYGGSPIGNQLREIERGCDLLVATPGRLNDLLERGKISLANVKYLVLDEADRMLDMGF EPQIRHIVEDCDMTPVGERQTLMFSATFPADIQHLARDFLSDYIFLSVGRVGS	#1739
S. cerevisiae Ded1p RecA2 (369-535)	GGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLIFVETKRMADQLTDFLIMQNFR ATAIHGDRTQSERERALAAFRSGAATLLVATAVAARGLDIPNVTHVINYDLPSDVDDYV HRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTEANQEVPSFLKDAMMS	#1735

Chapter 3: Structural characterization and molecular interactions of the conserved DEAD-box RNA helicase Ded1p

Contributions: All experiments were performed by myself. I was assisted in the crystallographic analysis by Philip Wurm and Daniela Lazzaretti. The assignment of the N-terminal IDR of Ded1p as well as the NMR titration experiments of the N-terminal IDR and the RecA2 domain of Ded1p were performed by Jakob Schneider, a master student, under my supervision. Parts of this chapter were used for a manuscript in cooperation with the group of Prof. Dr. Simon Alberti (Technische Universität Dresden):

Ceciel Jegers, Titus M Franzmann, **Julian Hübner**, Jakob Schneider, Cedric Landerer, Sina Wittmann, Agnes Toth-Petroczy, Remco Sprangers, Anthony A Hyman, and Simon Alberti. A conserved and tunable mechanism for the temperature-controlled condensation of the translation factor Ded1p, *bioRxiv* (2022)

3.1 Introduction

The DEAD-box RNA helicase Ded1p belongs to the highly conserved family of ATP-dependent RNA helicases found in all eukaryotes, archaea, and also in some prokaryotes (10, 13, 14). Ded1p is a helicase of 604 amino acids and is involved in the translation initiation in *S. cerevisiae*, for which it has been shown that knock-out of Ded1p is lethal (173–176). Similar to the overall structure of DEAD-box RNA helicases, Ded1p consists of two central RecA-like domains that are connected via a linker (152, 153). The two RecA-like domains form the conserved helicase core, which is flanked by N- and C-terminal extensions whose sequence is not conserved among other DEAD-box proteins (157–159, 162, 163). Under normal growth conditions, Ded1p facilitates the translation of housekeeping proteins by the unwinding of secondary structures in the 5´UTR of mRNAs (160, 177). However, upon environmental stress, the condensation and recruitment of Ded1p to SGs leads to a switch in translation by repressing the translation of mRNAs with structured regions to the translation of unstructured mRNAs that usually encode for, e.g. heat shock proteins (46, 49). However, the molecular interactions that lead to the condensation of Ded1p have not been elucidated.

The full-length Ded1p protein has recently been shown to reversibly assemble into condensates upon exposure to heat, as evidenced by the formation of condensates in the cytoplasm of yeast cells (46, 226). As a consequence, the translation switch mentioned above takes place.

A structure of Ded1p is not yet available. However, several crystal structures of the helicase core of DEAD-box helicases exist, such as the structure of the human Ded1p ortholog DDX3 (186). Furthermore, an Alphafold2 structure model is predicted for Ded1p from *S. cerevisiae* (187, 188).

In this study, we aimed to solve the crystal structure of the Ded1p helicase core, which serves as a basis for NMR titration studies. We conclude that both the N- and C-terminal disordered extensions participate in intermolecular interactions that promote the formation of condensates. Furthermore, we investigated the temperature-dependence of these interactions and we observed that the interactions become stronger with decreasing temperature. Thus, we propose that the condensation of Ded1p is regulated by an interaction network involving the interplay of folded and IDRs of Ded1p.

3.2 Methods

3.2.1 Protein expression and purification

All genes for the folded Ded1p domains were cloned into pETM-11 plasmids encoding for an N-terminal TEV protease-cleavable His_6-tag. The genes encoding for the flanking disordered regions of Ded1p were cloned into pET-GB1 plasmids encoding an additional His_6-GB1 solubility tag followed by the TEV site. Mutants of the N-terminal IDR of Ded1p were generated by Gibson assembly (227). The plasmid was transformed into chemically competent *E. coli* BL21(DE3) codon plus cells (Stratgene).

To produce unlabeled protein, cells were grown at 37°C while shaking in LB medium until an OD_{600} of 0.6-0.8 was reached. Protein expression was then induced with the addition of 1 mM IPTG. Cells were shifted to 20°C and incubated for 12-18 h, after which they were harvested by centrifugation (6,000 g for 15 min at 20°C).

To obtain ILVM-labeled protein, 25 mL H_2O based M9 medium was inoculated with 0.5 mL of LB overnight culture and was incubated at 37°C until an OD_{600} of 0.8 was reached. Cells were harvested by centrifugation (4000 g for 10 min at 20°C) and were further used to inoculate 100 mL D_2O based M9 medium, containing 2 g/L ^{13}C-glucose and/or 0.5 g/L ^{15}N-NH_4Cl as the sole carbon or nitrogen source, to an OD_{600} of 0.15. The M9 medium was incubated overnight at 37°C, was diluted the next day with fresh M9/D_2O medium to a final volume of 800 mL, and the cells were further incubated at 37°C until an OD_{600} of 0.8 was reached. Next, 60 mg/L 2-ketobutyric acid-(4-^{13}C,3,3-d_2) and 100 mg/L 2-Keto-3-methyl-butyric acid-(dimethyl-$^{13}C_2$, 3-d) dissolved in 200 mL M9/D_2O medium, and 100 mg/L L-methionine (methyl-$^{13}CH_3$) were added one hour prior induction. Protein expression was induced by the addition of 1 mM IPTG, cells were shifted to 25°C, and harvested by centrifugation (6,000 g for 15 min at 25°C) after 12-18 h.

Folded Ded1p domains and the N-terminal extension were purified in a similar manner. Cell pellets were resuspended in 25 mL Buffer A (25 mM NaPi pH 7.4, 1 M NaCl, 1 mM DTT, 10 mM imidazole) supplemented with 1:1000 (v/v) Triton-X-100, 1 mg/mL lysozyme, and 0.1 mg/mL DNase I. Resuspended cells were lysed by sonication and cell debris was removed by centrifugation (39,191 g for 30 min at 4°C). The supernatant was filtered using a 0.45 µm syringe filter and was then applied to a gravity flow Ni-NTA column that was equilibrated with Buffer A. The column was then washed with five column volume of Buffer A and 10 mL high salt washing buffer (25 mM NaPi pH 7.4, 2 M NaCl, 1 mM DTT). Next, proteins were eluted with Buffer B (25 mM NaPi pH 7.4, 1 M NaCl, 1 mM DTT, 300 mM imidazole). The N-terminal His_6-tag was removed by TEV cleavage and the sample was dialyzed against dialysis buffer (25 mM HEPES pH 7.3, 500 mM NaCl, 1 mM DTT) overnight at room temperature. The cleaved N-terminal His_6-tag was removed by a second Ni-NTA column that was equilibrated with dialysis buffer. The flow-through was concentrated to a volume of ~2 mL and was further purified by SEC using a 16/600 Superdex S75 column that was equilibrated with SEC buffer (25 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT). Protein-containing fractions were pooled, concentrated, and the final protein concentration was determined based on the A_{280}.

For the purification of the C-terminal extension, Buffer A and Buffer B did not contain DTT but were supplemented with 1 M guanidinium hydrochloride (GdmCl).

Furthermore, SEC buffer contained 500 mM NaCl. In order to prevent LLPS, cell debris after sonication was removed by centrifugation at room temperature. For NMR experiments, SEC buffer was exchanged to NMR buffer (20 mM HEPES pH 7.3, 125 mM NaCl). All proteins were frozen in liquid nitrogen and stored at -80°C until they were used for further experiments. All protein constructs that were used in this study are listed in **Supplemental Table 3.1**.

3.2.2 Nuclear magnetic resonance spectroscopy

All NMR experiments were performed at 4°C or 25°C on Bruker 500, 600, and 800 MHz Avance Neo spectrometers equipped with either nitrogen-cooled (500 and 600 MHz) or helium-cooled (800 MHz) triple-resonance cryoprobes. For 2D H-N spectra, ^{1}H-^{15}N HSQC and ^{1}H-^{15}N TROSY spectra were recorded. The SOFAST-HMQC pulse sequence was used to record methyl-TROSY NMR spectra (217). For the backbone assignment, standard Bruker pulse sequences for HNCACB, HNcoCACB, HNCO, HNcaCO, HCccoNH, and HNcaNNH were used. All NMR spectra were processed using either Topspin 4.02 or NMRPipe 9.6 (218) and were visualized using NMRDraw or Cara 1.9.2 (228).

The final sample was in SEC buffer supplemented with 5% (v/v) D_2O for frequency locking. Concentrations of labeled protein ranged from 20-250 µM or 0.5-2 mM for backbone or methyl-group assignment experiments. For NMR titration experiments, unlabeled Ded1p regions were added in two times molar excess over the proteins. For the labeled C-terminal IDR, a final ratio of 1:0.25 was achieved upon the addition of the helicase core or the RecA1 domain. The unlabeled RecA2 domain was added in 10 times molar excess over the labeled C-terminal IDR. Chemical shift changes were calculated using the following equation.

Equation 9
$$CSP = \sqrt{\left(\frac{\Delta C/\Delta N}{4}\right)^2 + \Delta H^2}$$

where $\Delta C/\Delta N$ and ΔH are the shift differences in ppm in the $^{13}C/^{15}N$ and ^{1}H dimensions, respectively. CSPs were plotted on protein structures that were visualized with PyMol (189).

To illustrate the loss of signal intensity, the peak volumes before and after the addition of the unlabeled protein regions were integrated using NMRPipe. The relative loss of

signal intensity was calculated as the ratio of the peak volumes before reduction (I) divided by the peak volume after reduction (I_0).

For the automated methyl group assignment using MAGIC (199), a 3D HMQC-NOESY-HMQC (CCH-) spectrum with a mixing time of 250 ms was recorded. The corresponding coordinate list was generated using NMRFAM-SPARKY 1.414 (229). HNH- and CNH-NOESY spectra were recorded with a mixing time of 200 ms, respectively.

3.2.3 Crystallization

In sitting drops under vapor diffusion, 0.3 µL of the precipitant reservoir solutions were mixed with 0.3 µL of protein (50 mg/mL of Ded1p core, 70 mg/mL of Ded1p RecA1, and 60 mg/mL of Ded1p RecA2) in 20 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT. In order to crystallize the core in the presence of a non-hydrolysable nucleotide, screens were prepared with 25 mg/mL and 15 mg/mL Ded1p core in 20 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$, 2 mM ADP, 4 mM BeF_3, 10 mM NaF, and 14mer polyA ssRNA. After 6-8 weeks, crystals for the isolated RecA1 domain were obtained in 0.1 M HEPES pH 7.0, Jeffamine ED-2001 pH 7.0 at 18°C. Grown crystals were fished and incubated in the reservoir solutions that was supplemented with 20% weight per volume (w/v) PEG400 as cryoprotectant, and crystals were subsequently frozen. Diffraction data were obtained on the PXII beamline at the Swiss Light Source at 100 K with a wavelength of 1 Å. Data were processed using XDS (230) and molecular replacement, followed by several rounds of model building and refinement using Phenix (231) and Coot (232). All images of protein structures were made with PyMol.

3.3 Results

3.3.1 Crystallization of the DEAD-box RNA helicase Ded1p

Although structures of several DEAD-box RNA helicases are found in the protein data bank (PDB) database, a structure of the DEAD-box RNA helicase Ded1p is not yet available. A structure of the helicase core of the human Ded1p ortholog DDX3 has

been solved by the Doudna group (186), and Alphafold2 also predicts a structure for the full-length Ded1p. We sought to solve the structure as the structural information would be helpful for the interpretation of NMR data. To this end, we first attempted to crystallize the Ded1p core and also both isolated RecA1 and RecA2 domains in the absence of a substrate RNA and nucleotides. Different crystallization conditions were tested. While for the isolated RecA2 domain only small crystal needles were obtained, rod-shaped crystals could be grown for the RecA1 domain. No crystals could be obtained for the helicase core.

Crystals for the RecA1 domain were fished and one obtained crystal diffracted up to a resolution of 2.2 Å. To determine the structure, the molecular replacement approach was performed with a homology model. The homology model was prepared using the SWISS-MODEL server (233), where the structure of the DEAD-box RNA helicase DDX3 was used as a template (PDB id: 5E7I). The validation resulted in a space group of P61 and the structure is similar to previously published RecA1 domains of DEAD-box RNA helicases (**Figure 3.1A** and **Supplemental Table 3.2**). The structure of our solved structure matches the structure of Vasa and DDX3 (156, 186).

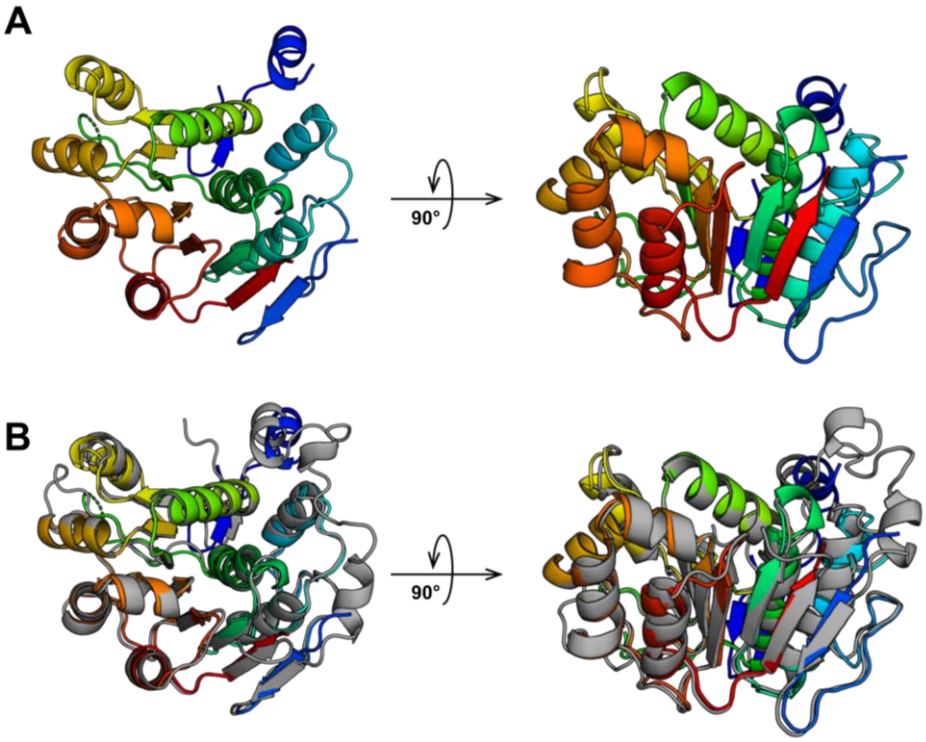


Figure 3.1: Crystallization of the Ded1p RecA1 domain. (**A**) Top view (right) and side view (left) of the *S. cerevisiae* Ded1p RecA1 domain. The structure is illustrated as a cartoon. The structure is rainbow colored from the N- to the C-terminus. (**B**) Superposition of the *S. cerevisiae* Ded1p RecA1 crystal structure (colored) and the predicted Alphafold2 structure (grey) of Ded1p RecA1. Both structures are shown as cartoons.

However, the electron density is not visible for all residues of the RecA1 domain. Residues are missing from two regions where loops are predicted. For the first region, residues from 108-127 are missing after the first alpha-helical element (**Figure 3.1A**, dark blue helix), a long loop with one alpha-helical element is predicted in the Alphafold2 structure (**Figure 3.1B**, grey cartoon). The second missing density belongs to another short loop stretching from residue 215-221 that is visualized by a dashed line between the green alpha-helix and the green beta-sheet. The electron density is most likely missing due to the high flexibility of these loop regions. Both the crystal and the Alphfold2 structures can be superimposed with a root mean square deviation (RMSD) of 0.931. The crystal structure has been deposited in the PDB database (PDB ID: 8PNK).

Although a high-resolution structure of the isolated RecA1 domain has been solved, a structure of the entire helicase core is still not available. Since both RecA domains tumble independently in the absence of RNA and nucleotides (Chapter 2.3.1), the open conformation of the helicase might be too flexible for crystallization. To make the protein more rigid, we attempted to crystallize the closed core conformation of the Ded1p helicase core. Attempts were made to crystallize the Ded1p core in the presence of an excess of ADP-BeF_3 and 14mer polyA ssRNA. Testing many crystallization conditions resulted in single rod-shaped crystals. However, no structural information could be obtained because the crystals did not diffract at all. In order to obtain a crystal structure of the entire helicase core, both in the open and the closed conformation, additional screening experiments for ideal crystallization conditions will need to be performed.

3.3.2 High quality NMR spectra can be recorded for all Ded1p domains

In order to investigate intra- and intermolecular interactions between different Ded1p domains that might be responsible for the LLPS behavior of Ded1p, we turned to NMR spectroscopy. We successfully established conditions under which high quality ^{1}H-^{15}N-correlation NMR spectra of each domain of Ded1p were recorded (**Figure 3.2**). Spectra of the IDP exhibit strong resonance broadening due to rapid exchange of the amide protons with the bulk solvent. Therefore, all subsequent NMR spectra of the N- and C-terminal IDRs were recorded at 4°C. In contrast, the quality of the spectrum of

the folded domains is reduced at low temperature due to the slow tumbling rate of the proteins, which leads to signal broadening. Therefore, ^{1}H-^{15}N-correlation NMR spectra of the folded domains were recorded at 25°C.

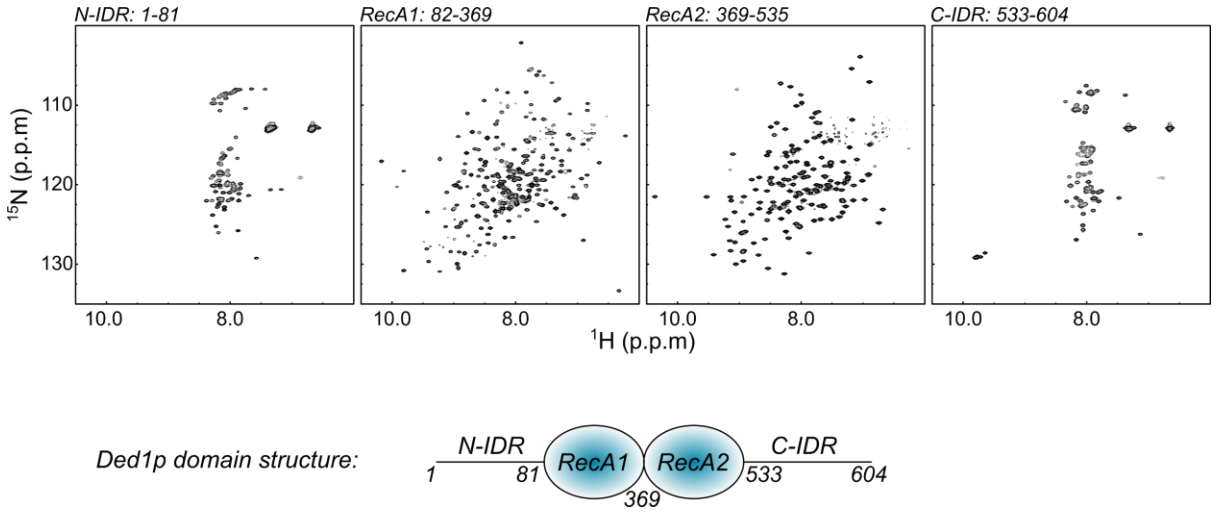


Figure 3.2: NMR spectra and domain architecture of Ded1p. Top: ^{1}H-^{15}N-correlation NMR spectra of the N-terminal IDR, the two central RecA domains, and the C-terminal IDR of Ded1p from *S. cerevisiae*. Spectra of the N- and C-terminal IDR were recorded at 4°C, while spectra of the isolated RecA domains were recorded at 25°C. Bottom: The domain boundaries of Ded1p are shown as cartoon.

For the NMR spectra of the N- and C-terminal IDR, poor chemical shift dispersions of their resonances can be seen, which are limited to a small region around 8 ppm (**Figure 3.2**, N-IDR and C-IDR). This indicates that neither of the two regions adopts any stable secondary structure. Apart from resonances arising from N-H groups in the side chains of certain amino acids such as asparagine, glutamine, and tryptophane, single resonances scattering around 8 ppm were found for both disordered extensions and could indicate that these residues might form partially populated structural elements in solution. For the isolated RecA domain, both were shown to be structured in solution, as visualized by a well dispersion of their resonances ranging from 7 to 10 ppm. The good spectral quality of all Ded1p domains allows to perform NMR titration experiments.

3.3.3 Backbone assignment of each isolated Ded1p domain

In order to extract further information from the ^{1}H-^{15}N correlation spectra and to use this information for interaction studies, the resonances of both IDRs and both RecA domains were assigned to their respective nuclei. This assignment will lay the foundation for further studies that will help to unravel the interactions between Ded1p that contribute to LLPS, but also to the RNA unwinding mechanism. As a result, 85% of the HN, N, Cα, Cβ, and C' backbone resonances were assigned to the Ded1p N-terminal disordered extension (**Figure 3.3**).

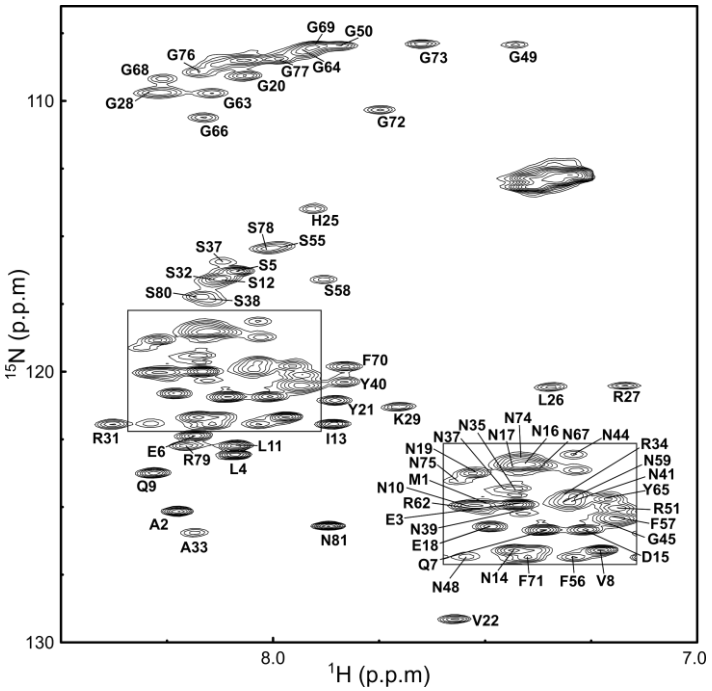


Figure 3.3: ^{1}H-^{15}N-correlation spectrum and residue-type specific backbone resonance assignment of the N-terminal IDR (residues 1-81) of Ded1p. 2D correlation and triple resonance spectra recorded at 500 MHz and 4°C were used for the backbone assignment. The assigned backbone amide resonances are labeled. Inset on the lower left side shows the boxed region in the center of the spectrum.

The N-terminal IDR of Ded1p is prion-like in its composition due to the high content of polar residues such as serine an asparagine (**Supplemental Table 3.1**). Furthermore, the sequence displays a high content of arginine and glycine residues. Interestingly, most of the resonances distributed around 8 ppm belong to the same stretch in the sequence of the N-terminal IDR of Ded1p (residues 21-27), a region that has been shown to harbor a motif for the binding of eIF4E (160).

Next, backbone assignment of the core domains was performed. Although 2D ^{1}H-^{15}N correlation spectra of the Ded1p helicase core can be recorded, many peaks overlap, and the spectral quality becomes too low for a 45 kDa protein to perform triple-resonance spectroscopy experiments, complicating an assignment. However, the isolated domains of Ded1p show well-dispersed ^{1}H-^{15}N correlation spectra (**Figure 3.2**), and both domains tumble independently in solution and show no specific inter domain interactions (Chapter 2.3.1). We therefore employed an “divide and conquer” strategy where a large complex, in our case the entire helicase core, is divided into smaller blocks, here the isolated RecA1 and RecA2 domains. This resulted in the assignment of 75% of the backbone resonances for the RecA1 domain of Ded1p (**Figure 3.4**).

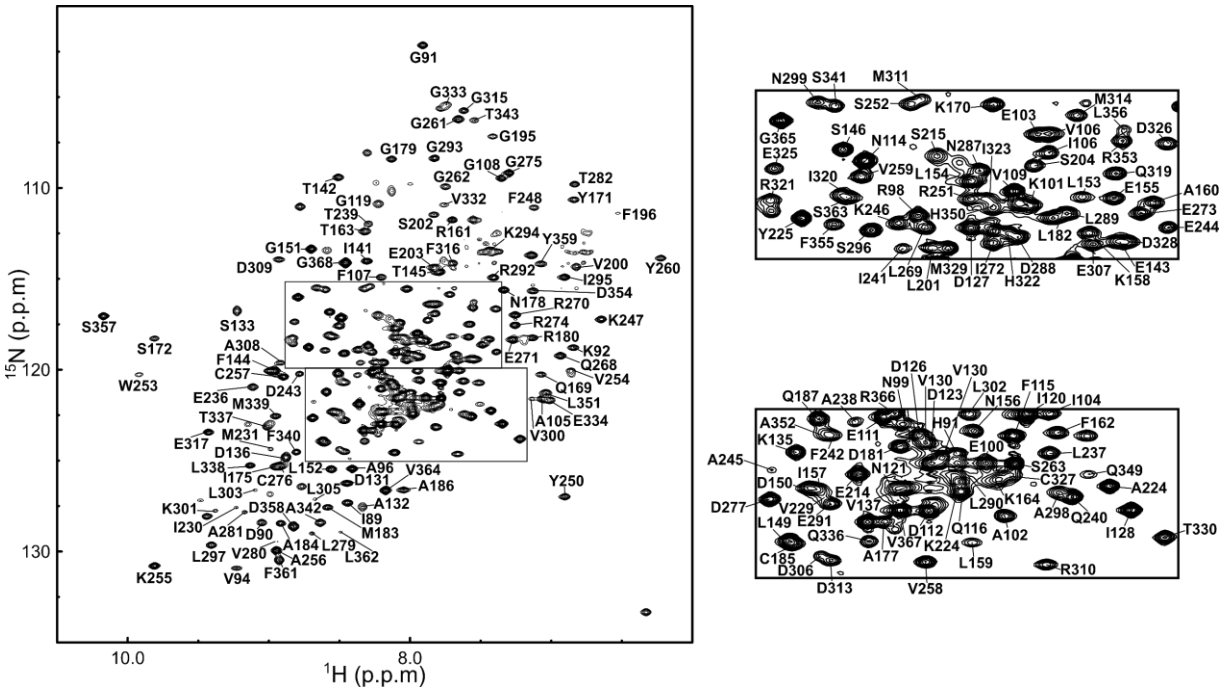


Figure 3.4: ^{1}H-^{15}N-correlation spectrum and residue-type specific backbone resonance assignment of the RecA1 domain (residues 82-369) of Ded1p. 2D correlation and triple resonance spectra recorded at 800 MHz and 25°C were used for the backbone assignment. The assigned backbone amide resonances are labeled. Insets on the left show the boxed regions in the center of the spectrum.

Like the RecA1 domain, the RecA2 domain gives a well-dispersed ^{1}H-^{15}N correlation spectrum. Here, 97% of all backbone resonances could be assigned to the RecA2 domain of Ded1p (**Figure 3.5**).

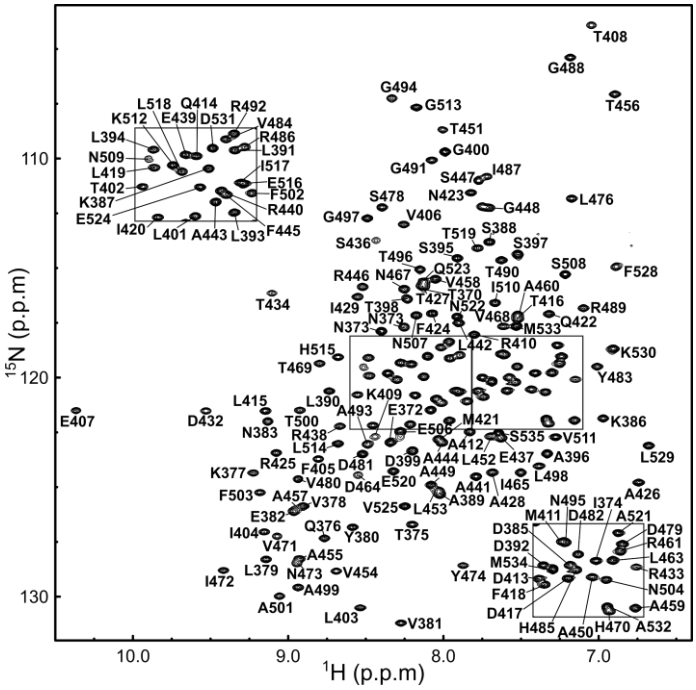


Figure 3.5: ^{1}H-^{15}N-correlation spectrum and residue-type specific backbone resonance assignment of the RecA2 domain (residues 369-535) of Ded1p. 2D correlation and triple resonance spectra recorded at 800 MHz and 25°C were used for the backbone assignment. The assigned backbone amide resonances are labeled. Insets on the left and right side show the boxed regions in the center of the spectrum.

Finally, the C-terminal disordered extension was also assigned. 96% of all resonances of the C-terminal extension of Ded1p could be assigned (**Figure 3.6**). Based on the backbone assignment, it was observed that the resonances of the C-terminal WW motif do not scatter around 8 ppm like the other resonances, but show a dispersion around 7 to 7.5 ppm, which could indicate the formation of a partially formed structure.

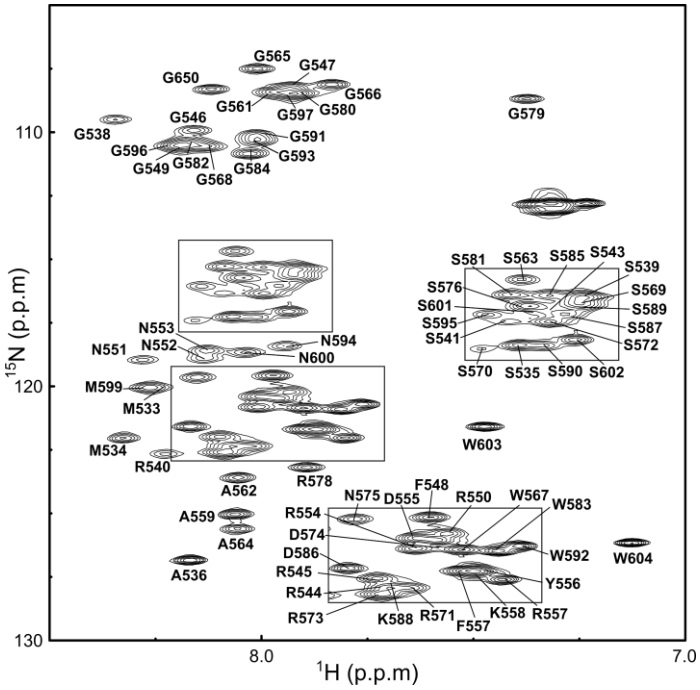


Figure 3.6: ^{1}H-^{15}N-correlation spectrum and residue-type specific backbone resonance assignment of the C-terminal IDR (residues 533-604) of Ded1p. 2D correlation and triple resonance spectra recorded at 500 MHz and 4°C were used for the backbone assignment. The assigned backbone amide resonances are labeled. Insets on the left and bottom side show the boxed regions in the center of the spectrum.

The assignments of the backbone amide resonances of the N-terminal IDR, the C-terminal IDR, and both isolated RecA domain have been deposited at the Biological Magnetic Resonance Bank under the accession numbers 51649 (RecA2 domain), 51650 (N-terminal IDR), 52030 (C-terminal IDR), and 52034 (RecA1 domain).

3.3.4 Methyl group assignment of the Ded1p helicase core

Classical heteronuclear NMR experiments with uniformly ^{13}C or ^{15}N labeled proteins are usually limited to proteins up to 25-30 kDa in size and more residues lead to more peak overlap in a 2D spectrum. Therefore, methyl groups are ideal reporters for studying the interactions of large proteins. There are less methyl groups than backbone amide resonances, so less resonance overlap occur. Methyl groups can be selectively labeled, and one can exploit favorable relaxation properties of the methyl groups in methyl-TROSY type of experiments.

Similar to the backbone assignment of the Ded1p core, the methyl group assignment was performed with the isolated RecA domains, and the assignments can then be transferred to the core, as it was previously shown that the sum of the methyl groups of both RecA domains looks identical to the methyl group spectrum of the full-length Ded1p helicase core (**Figure 2.2**, Chapter 2.3.1).

Therefore, all methyl groups of isoleucine, leucine, and valine of the RecA1 domain were ^{13}C labeled with specific precursors. An attempt was then made to assign the methyl groups for the RecA1 domain (residues 369-535) by recording 3D hCcccoNH and 3D HcccoNH spectra, which provide chemical shifts for carbons and hydrogens in the side chains of the residues. Due to low spectral quality, only a minimal number of resonances could be assigned. In addition, and to assign more methyl resonances, we made use of a software program. For proteins for which a structure is available, several software programs can be used to assign methyl groups. In our case we used the MAGIC algorithm (199). This algorithm uses an existing structure and combines it with experimental NOE data. A peak list is generated based on a recorded ^{13}C-^{13}C-^{1}H 3D SOFAST-HMQC-NOESY-HMQC spectrum of an ILVM-labeled protein. These NMR data, together with our crystal structure of the RecA1 domain (PDB id: 8PNK) of Ded1p, provided the basis for the assignment process, which is based on the MAGIC algorithm. The majority of the ILVM methyl groups could be assigned based on the MAGIC algorithm. The confidence score of the resulting methyl group assignment of the RecA1 domain is shown in **Figure 3.7**.

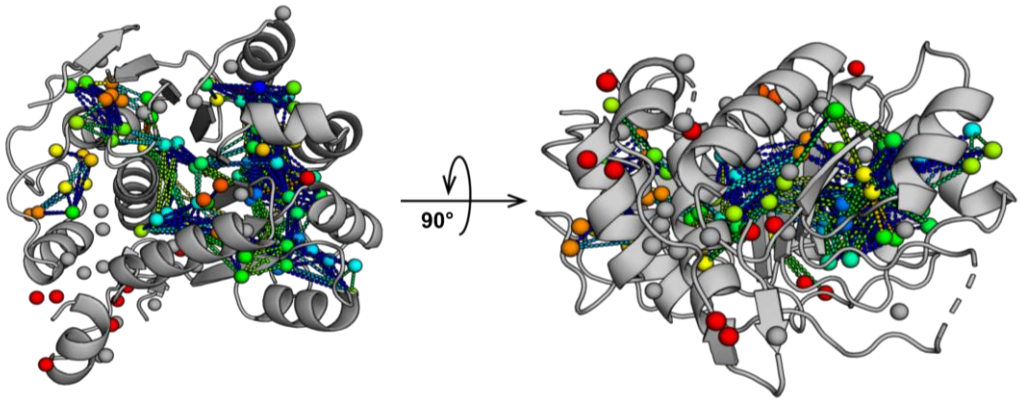


Figure 3.7: Methyl group assignment based on the MAGIC algorithm of the ILVM labeled Ded1p RecA1 domain. The confidence score of the MAGIC algorithm is shown. Each methyl group is represented as a sphere. The color of a sphere encodes the confidence score from red (low confidence) to blue (high confidence). Methyl groups that are isolated and therefore could not be assigned are shown as grey spheres. Connections between methyl groups are represented by dashed lines. Their color code is the same as that of the spheres.

The methyl group assignments of the RecA1 domain obtained by MAGIC were then used as a starting point to validate and continue the assignment based on 3D CCH-, HCH-, and CNH-NOESY spectra. Taken together, the use of the MAGIC algorithm in combination with the recorded NMR data resulted in the assignment of 74% of the methyl groups of an ILVM-labeled RecA1 domain (**Figure 3.8**).

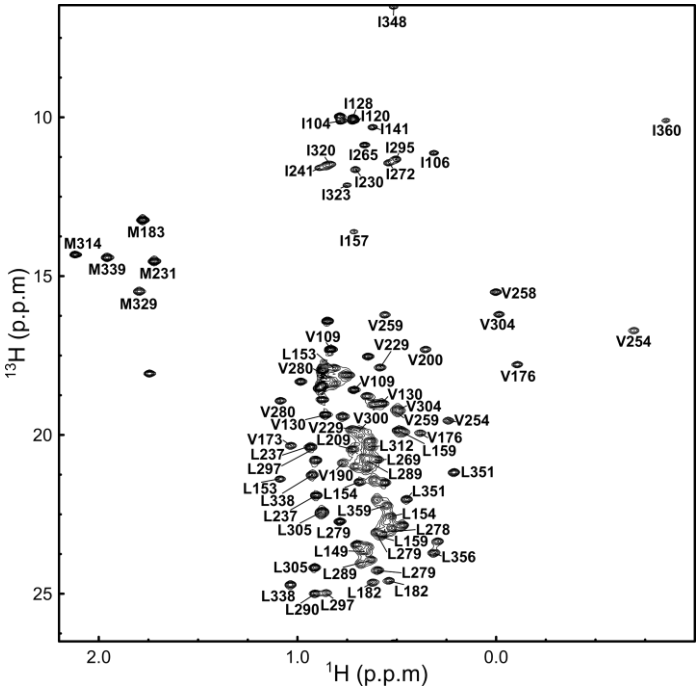


Figure 3.8: ^{13}C-methyl TROSY spectrum and residue-type specific resonance assignment of the ILVM-labeled methyl groups of the RecA1 domain (residues 82-369) of Ded1p. NMR spectra were recorded at 800 MHz and 25°C. The assigned methyl group resonances are indicated.

15 out of 17 Ile resonances, 19 out of 23 Leu resonances, 5 out of 6 Met resonances, and 9 out of 19 expected Val resonances could be assigned. The high number of missing Val residues is caused by a strong overlap of Val resonances around 0.5-0.7 ppm. In summary, NOE signals were assigned to the methyl groups of ILVM amino acids for the RecA1 domain using the MAGIC algorithm.

In order to assign more resonances, especially for valine residues, only one or two types of residues will be labeled. Individual residues will then be mutated to alternative residues and the resonances of the substituted residue can be easily identified as they are missing from the spectrum of the mutant protein. Such information will then be used for another round of refinement.

Resonances of the RecA2 domain were also assigned. Due to the high sample concentration (2 mM) and stability, a 4D HCccoNH was recorded, allowing the linkage and assignment of the ILV methyl groups. The assignment of the RecA2 resonances is shown in **Figure 3.9**.

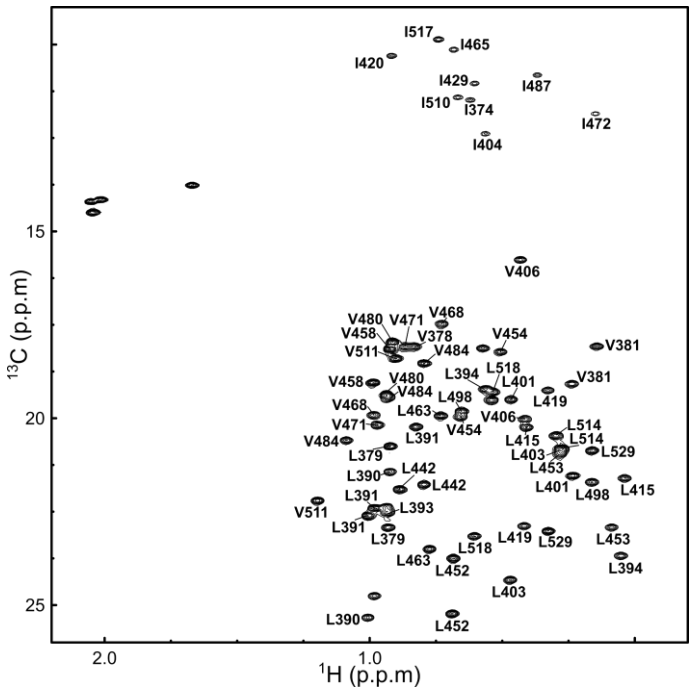


Figure 3.9: ^{13}C-methyl TROSY spectrum and residue-type specific resonance assignment of the ILVM-labeled methyl groups of the RecA2 domain (residues 369-535) of Ded1p. NMR spectra were recorded at 800 MHz and 25°C. The assigned methyl group resonances are indicated.

For the RecA2 domain, all 9 Ile resonances were assigned. Furthermore, 17 out of 18 Leu and 10 out of 11 Val resonances could be assigned. However, the assignment of the methionine resonances was not possible with this approach. To assign the methionine resonances, a 3D CCH-NOESY spectrum for an ILVM-labeled protein will be required. In addition, the mutation of individual methionine residues in an only M-labeled protein will be helpful in assigning these resonances.

By transferring the methyl group assignments from the isolated RecA domains to the helicase core, 79% of all ILVM methyl groups of the Ded1p core were assigned. Taken together, ^{1}H-^{15}N backbone and ^{13}C methyl group resonances could be mostly assigned for all domains of Ded1p. These assignments were used in subsequent NMR

titration experiments for mapping potential binding sites for the different Ded1p domains.

No methyl group labeling and assignment was performed for the N- and C-terminal IDRs due to the absence or low abundance of ILVM residues.

3.3.5 The intrinsically disordered regions of Ded1p interact with the folded helicase core

We experimentally tested potential interactions of Ded1p regions using NMR spectroscopy. As a first step, the interactions of the N-terminal IDR with the folded helicase domains were investigated. For this purpose, ^{1}H-^{15}N correlation spectra of the ^{15}N labeled Ded1p N-terminal IDR were recorded before and after the addition of the helicase core (RecA1-RecA2). **Figure 3.10** shows the spectra of the N-terminal IDR in the absence and presence of a two-molar excess of the unlabeled Ded1p helicase core. Several CSPs, but also line broadening, were observed for a number of resonances upon the addition of the helicase core. These observations can be explained by a direct interaction between both Ded1p regions. Based on the previous backbone assignment of the Ded1p N-terminal IDR (**Figure 3.3**), strong CSPs can be mapped to residues 21-33, while no or only weak CSPs were observed for other residues. Strikingly, the affected resonances are mostly the same ones that are distributed around 8 ppm and belong to the stretch of sequence that harbors the conserved motif for eIF4E binding.

As a follow up, NMR titration experiments were performed with a two-molar excess of the isolated and unlabeled RecA1 and RecA2 domains, respectively (**Figure 3.10**). Upon addition of the unlabeled RecA1 domain, no or very weak CSPs are observed and the spectrum resembles the spectrum of the N-terminal IDR alone. Therefore, no direct interaction of the N-terminal IDR with the isolated RecA1 domain is detectable. Remarkably, the titration of the unlabeled RecA2 domain again revealed a significant loss of resonance intensity and CSPs for the same residues as previously observed for the titration of the helicase core. Thus, our NMR titration experiments demonstrate that the specific binding site mediating the interaction with the N-terminal IDR can be localized to the RecA2 domain. Nevertheless, titration experiments with the RecA2 domain alone resulted in smaller CSPs than in titration experiments with the helicase

core (**Figure 3.10**, bottom). This suggests that additional interactions are mediated by the RecA1 domain when the core is added. A possible explanation is that the interaction with the RecA1 domain might be of low infinity and the excess of the RecA1 domain might not be enough to observe an interaction. Interactions between the N- and the C-terminal IDRs of Ded1p were also investigated and are described in **Figure 3.12**.

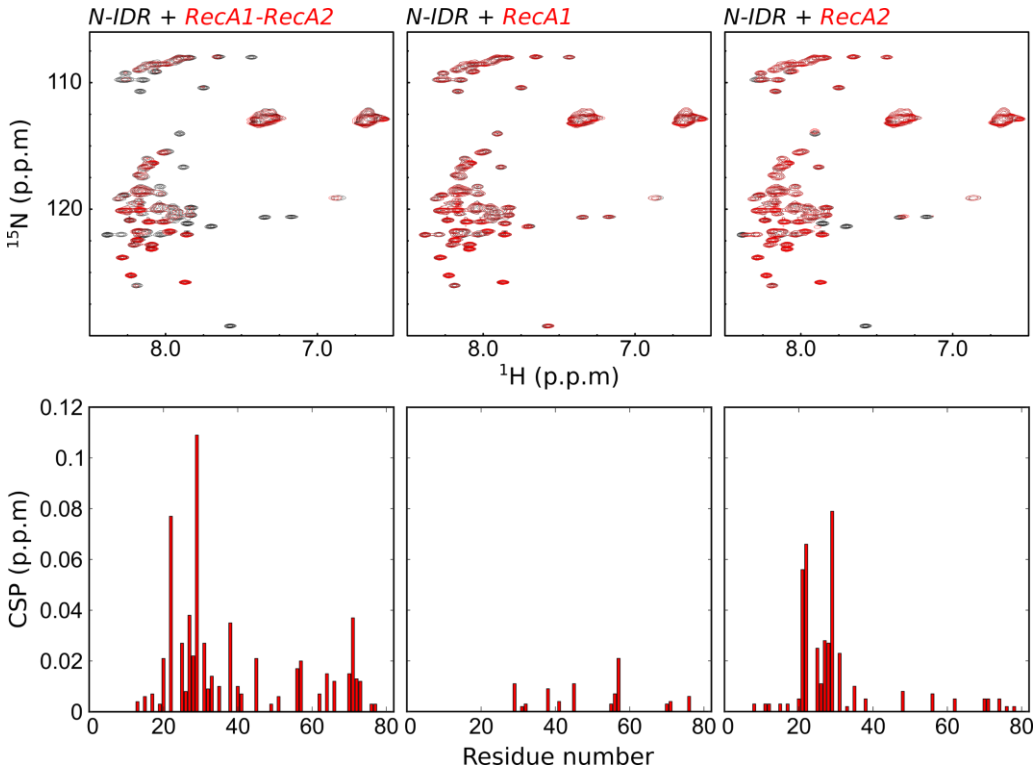


Figure 3.10: The N-terminal IDR interacts with parts of the folded helicase domains. Top: Shown are the ^{1}H-^{15}N NMR spectra of the labeled N-terminal IDR of Ded1p in the absence (black) and presence (red) of the Ded1p helicase core (RecA1-RecA2) (left), the isolated RecA1 domain (middle), and the isolated RecA2 domain (right). Individual resonances undergo CSPs and broaden upon the addition of the helicase core, but also for the isolated RecA2 domain, indicating a specific interaction. Upon the addition of the isolated RecA1 domain, no significant CSPs are observed, indicating that no interactions take place. Bottom: CSPs from the upper panel are plotted against the amino acid sequence of the N-terminal IDR of Ded1p. Large CSPs are detected upon the addition of the helicase core and the isolated RecA2 domain for residues 20 up to 30 of the IDR.

Upon the deletion of residues 21-35, the interaction of the Ded1p RecA2 domain with the N-terminal IDR is abolished as no CSPs are detected upon the addition of the RecA2 domain (**Figure 3.11**).

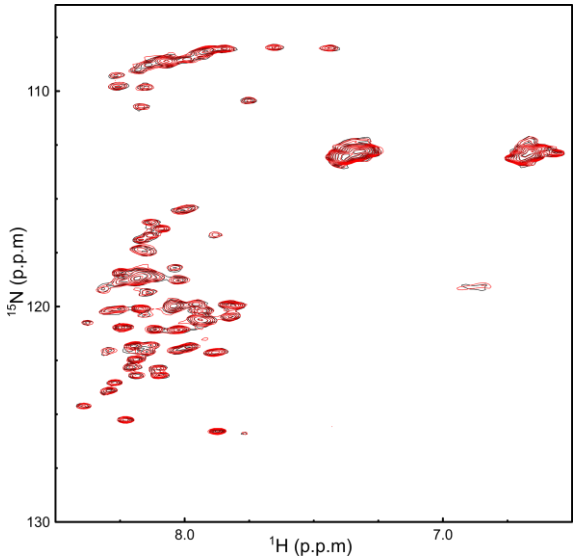


Figure 3.11: Deletion of residues 21-35 of the N-terminal IDR abolishes the interaction with the Ded1p RecA2 domain. Shown are the ^{1}H-^{15}N NMR spectra of the labeled N-terminal IDR with a deletion of residues 21-35 of Ded1p in the absence (black) and presence (red) of the Ded1p RecA2 domain.

In a similar manner, NMR titrations were performed to determine whether the C-terminal IDR interacts with the helicase core. The addition of the unlabeled helicase core to the ^{15}N labeled Ded1p C-terminal in a final ratio of 1:0.25 IDR:core immediately results in LLPS in the NMR tube, which could be seen by a strong turbidity of the solution. The resulting NMR spectrum after the addition showed a significant broadening of all resonances, while CSPs were not observed (**Figure 3.12A**, top). The relative difference in resonance intensity before and after the addition of the helicase core for all isolated resonances was plotted and showed that the loss of intensity is similar for all observed resonances (**Figure 3.12A**, bottom). All resonances lost up to 70% of their intensity due to strong peak broadening caused by a direct interaction with the helicase core resulting in LLPS. LLPS and line broadening are also observed upon the addition of the isolated RecA1 domain (**Figure 3.12**). This demonstrates that the interaction between the helicase core and the C-terminal IDR is mediated by the RecA1 domain. However, the loss of intensity is slightly reduced compared to the loss of resonance intensity with the helicase core. This also suggests that additional interactions are mediated by the RecA2 domain. Interestingly, the isolated RecA2 domain alone does not affect the resonances of the C-terminal IDR (**Figure 3.12B**). There is also no N-IDR:C-IDR interaction between the isolated N- and C-terminal extensions, as no resonances underwent CSPs or signal broadening (**Figure 3.12B**).

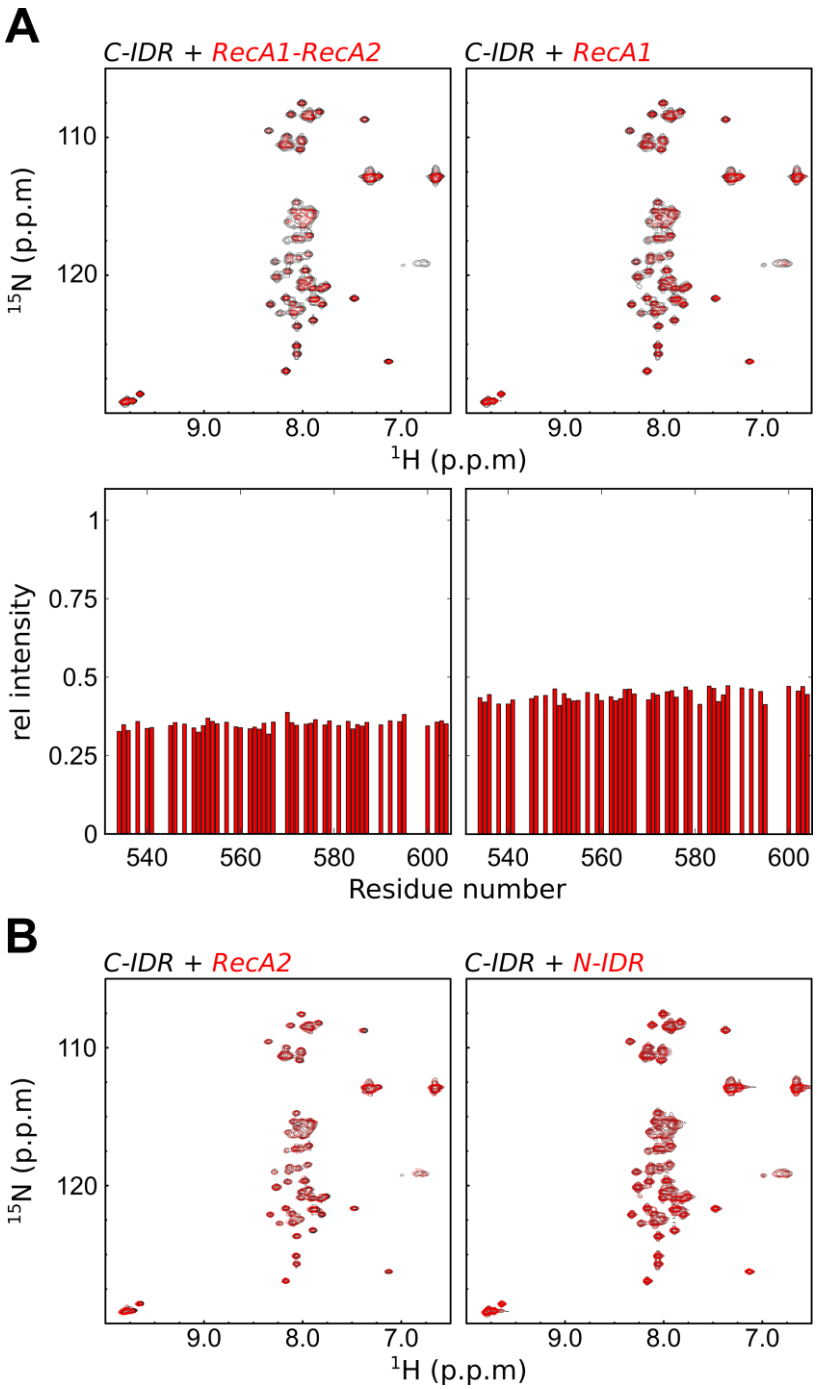


Figure 3.12: The C-terminal IDR interacts with parts of the folded helicase domains. (**A**) Top: Shown are the ^{1}H-^{15}N NMR spectra of the labeled C-terminal IDR of Ded1p in the absence (black) and presence (red) of the Ded1p helicase core (left) and the isolated RecA1 domain (right). Individual resonances broaden significantly upon the addition of the helicase core, but also for the isolated RecA1 domain, indicating a specific interaction. Bottom: The relative intensities of the upper panels are plotted against the amino acid sequence of the C-terminal IDR of Ded1p. All resonances are equally affected upon the addition of the helicase core or the isolated RecA1 domain. (**B**) Shown are the ^{1}H-^{15}N NMR spectra of the labeled C-terminal IDR of Ded1p in the absence (black) and presence (red) of the isolated RecA2 domain (left) and the N-terminal IDR (right) of Ded1p. The addition did not cause any CSPs or signal broadening, indicating that no interactions occur between these Ded1p regions.

In conclusion, both the N- and C-terminal extensions of Ded1p interact with the folded helicase core. For the N-terminal extension, the interaction is mainly mediated by the RecA2 domain, which interacts with a stretch of the sequence from residue 21 to 33, which has also been described as the binding site for eIF4E. In contrast, the C-terminal IDR interaction with the helicase core is mediated by the RecA1 domain, but no interaction site can be mapped because all resonances are equally affected. The broadening of the resonances is a consequence of complex formation, most likely related to condensation by LLPS at the protein concentrations required for NMR experiments.

3.3.6 Interaction between the N-terminal IDR and the RecA2 domain becomes stronger with decreasing temperature

To confirm the interaction between the N-terminal IDR and the RecA2 domain of Ded1p, the experiment was repeated with an inverted labeling. Spectra of the ^{15}N and ^{13}C-ILVM labeled RecA2 domain in the absence and presence of the unlabeled N-terminal IDR were recorded (**Figure 3.13**). CSPs are observed upon the titration of the IDR, further confirming an interaction between both Ded1p regions. For NMR experiments performed at 25°C, only weak CSPs are detectable for some resonances. Interestingly, the interaction becomes stronger as the temperature decreased, since the number and extent of CSPs increased as the temperature decreased. Based on the backbone and methyl group assignments of the RecA2 domain, a surface-exposed helix stretching from residues 380-397 was identified to interact directly with the N-terminal IDR (**Figure 3.13**, orange residues).

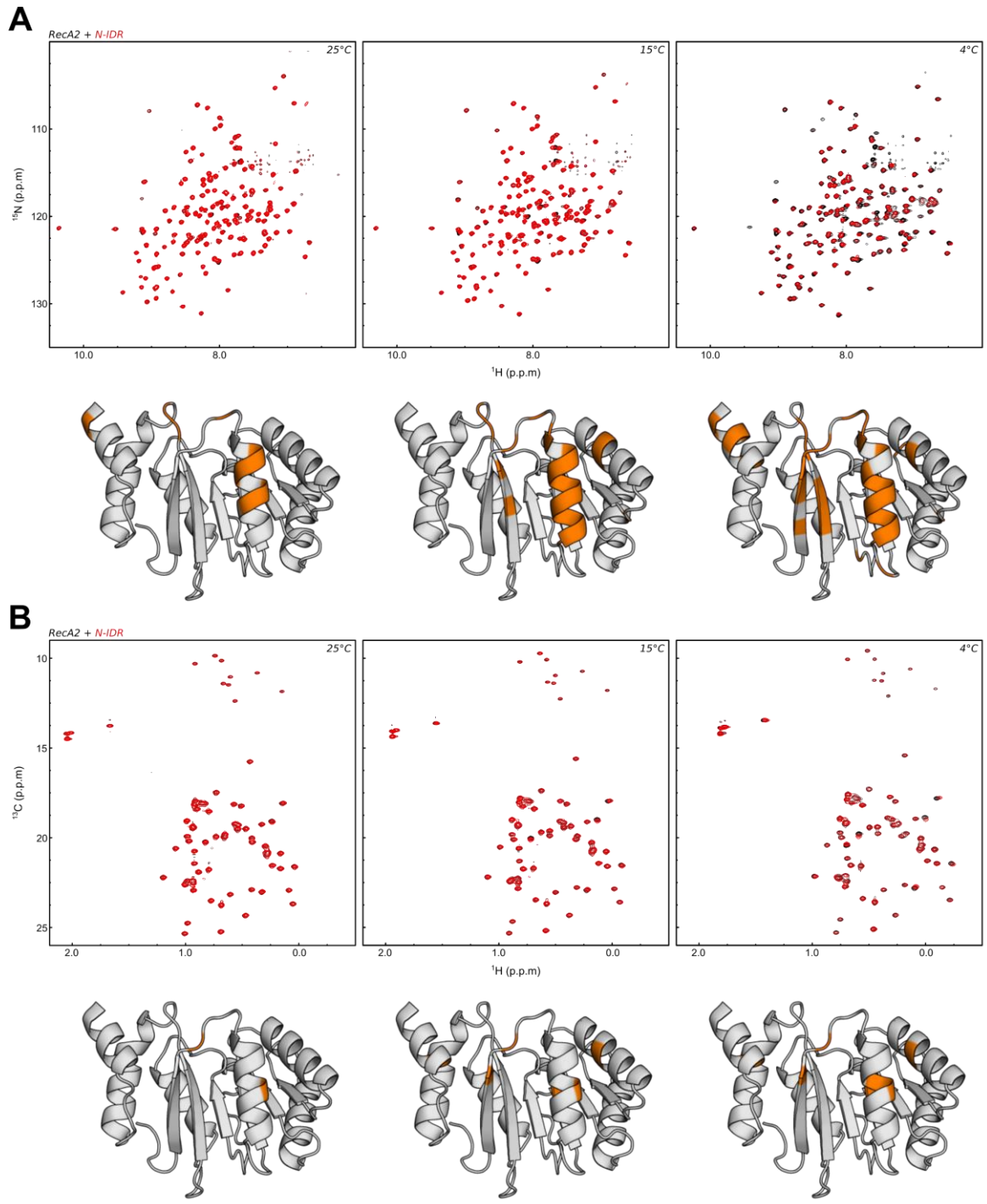


Figure 3.13: Mapping of the interactions between the RecA2 domain and the N-terminal IDR of Ded1p. (**A**) Top: ^{1}H-^{15}N spectra of the Ded1p RecA2 domain in the absence (black) and presence (red) of the N-terminal IDR at 25°C, 15°C, and 4°C. The extent of the CSPs increased at lower temperatures. Bottom: Observed CSPs from the upper panels are plotted on the Alphafold2 structure predition of the RecA2 domain of Ded1p. Resonances of residues that undergo CSPs are highlighted in orange. Residues 380-397, which belong to a surface-exposed helix, mediate the interaction with the N-terminal IDR. (**B**) Top: ^{13}C-methyl TRSOY spectra of the Ded1p RecA2 domain in the absence (black) and presence (red) of the N-terminal IDR at 25°C, 15°C, and 4°C. As previously described, the extent of CSP increased at lower temperatures. Bottom: Residues that undergo CSPs can be mapped to the same helix in the RecA2 domain.

As a follow-up experiment, the inverse NMR titration experiment were conducted between the RecA1 domain and the C-terminal IDR. NMR spectra of the ^{15}N and ^{13}C-ILVM methyl group labeled RecA1 domain were recorded before and after the addition of an excess of the C-terminal IDR (**Figure 3.14**).

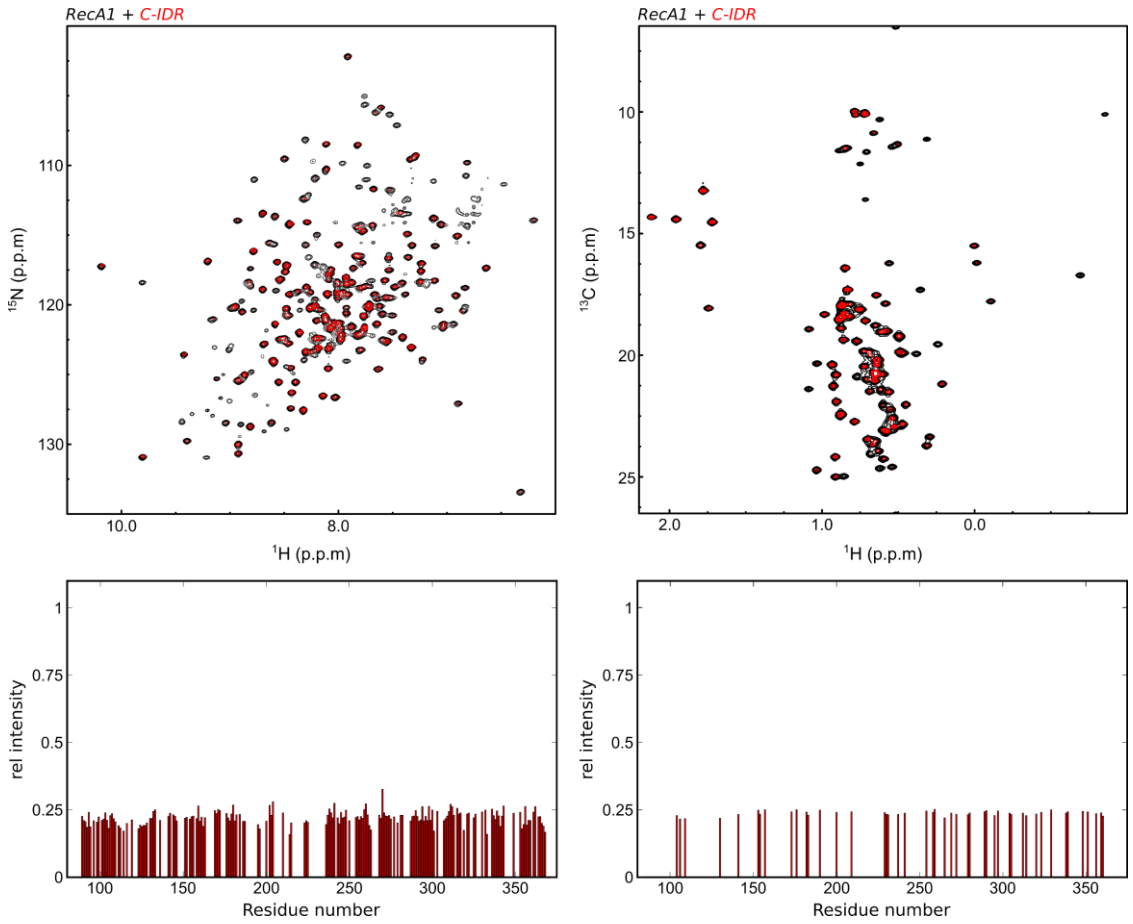


Figure 3.14: The interaction site between the RecA1 domain and the C-terminal IDR of Ded1p cannot be mapped. Top: ^{1}H-^{15}N and ^{13}C-methyl TROSY spectra of the Ded1p RecA1 domain in the absence (black) and presence (red) of the C-terminal IDR. A significant resonance broadening is observed upon addition of the C-terminal IDR. Bottom: The relative intensities of the upper panels are plotted against the amino acid sequence of the RecA1 of Ded1p. For both spectra, all resonances are equally affected by the addition of the isolated RecA1 domain.

As described previously, the addition of the unlabeled C-terminal IDR induced LLPS in the NMR tube. As it was already observed for the resonances of the C-terminal IDR (**Figure 3.12**), the resonances for the RecA1 domain broadened significantly after the addition of the C-terminal IDR, confirming an interaction between both regions of Ded1p. Also in this experimental setup, it is not possible to localize the interaction to a specific region of the RecA1 domain, as all resonances lost intensity equally,

irrespective of the ^{1}H-^{15}N TROSY or ^{13}C-methyl TROSY NMR spectra (**Figure 3.14**). Lowering the temperature caused the resonances to broaden to a point where detection was no longer possible (**Figure 3.15**).

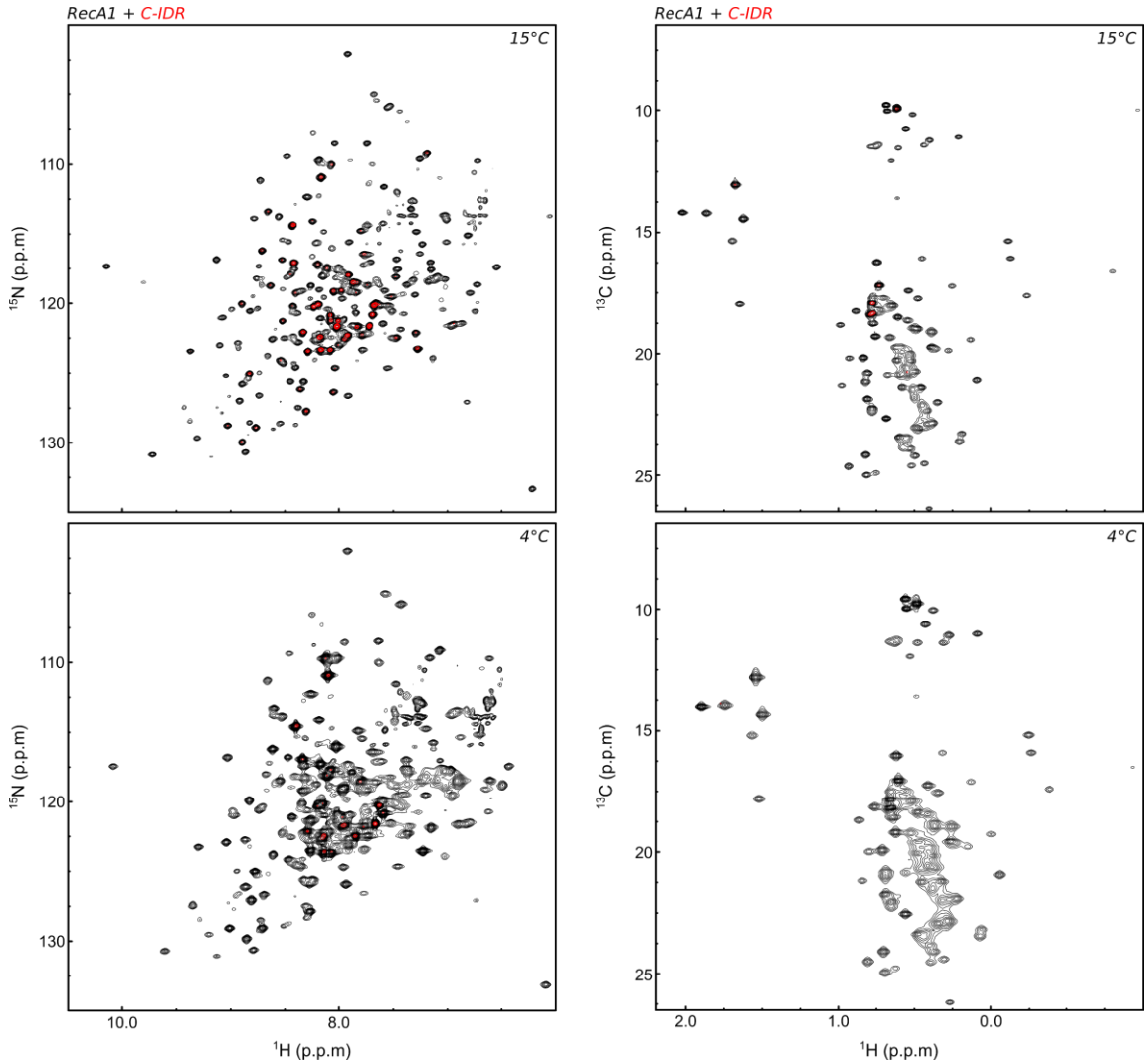


Figure 3.15: An interaction site between the RecA1 domain and the C-terminal IDR cannot be mapped. Left: ^{1}H-^{15}N spectra of the Ded1p RecA1 domain in the absence (black) and presence (red) of the C-terminal IDR at 15°C and 4°C. A significant resonance broadening is observed upon the addition of the C-terminal IDR. Right: ^{13}C-methyl TROSY spectra of the Ded1p RecA1 domain in the absence (black) and presence (red) of the C-terminal IDR at 15°C and 4°C. A significant resonance broadening is observed upon the addition of the C-terminal IDR.

Taken together, our NMR titration experiments demonstrate that the N-terminal IDR interacts with a surface-exposed helix of the RecA2 domain. This interaction is weakened upon an increase in temperature. In contrast, the C-terminal IDR was found to interact with the RecA1 domain. However, regardless of the labeling scheme used,

a binding site cannot be mapped due to LLPS, which causes a strong reduction in signal intensity as a result of significant resonance broadening.

3.4 Discussion

In this work, we made use of X-ray crystallography in combination with extensive NMR studies to elucidate the structure and interactions of the conserved yeast helicase Ded1p. Like most DEAD-box RNA helicases, Ded1p consists of a folded helicase core that is flanked by N- and C-terminal IDRs. Since only homology models or an Alphafold2 structure prediction of the Ded1p helicase were available, we sought to obtain the crystal structure of the helicase core and the isolated RecA domains. We solved the structure of the first RecA domain at a resolution of 2.2 Å. The structure resembles structures of other RecA-like domains found in the PDB database (156, 186, 208, 234). We also obtained crystals of the helicase core in the absence and presence of an ATP analog and RNA, and of the isolated RecA2 domain. However, these crystals did not display diffraction of sufficient quality for structure determination and will require further optimization.

To elucidate intra- or intermolecular interactions that occur between different Ded1p regions, we employed NMR spectroscopy. Both the disordered extensions and the RecA domains give high quality NMR spectra that allowed the assignment of most ^{1}H-^{15}N and ^{13}C-methyl group resonances. As a result, 85% of all backbone amide resonances of the DEAD-box RNA helicase Ded1p could be assigned. Furthermore, by assigning ^{13}C-labeled ILVM methyl groups of the isolated RecA domains and transferring them to the helicase core, a total assignment of 79% of the ILVM methyl resonances for the helicase core was achieved. For the RecA1 domain, we used the MAGIC algorithm, which combined the previously solved crystal structure of the RecA1 domain with NOE data and led to the initial assignment of most, though not all, resonances. In addition, other published algorithm such as methyl assignment using satisfiability (MAUS) will be tested, which has been reported to be more accurate and complete than the MAGIC algorithm (235). In addition, specific labeling of valine or leucine and mutation of specific residues will be used to complement the assignment of the RecA1 domain (196).

The assignments were used as a basis for NMR titration experiments to highlight any interactions between Ded1p regions. The NMR data revealed that the N- and C-terminal disordered extensions interact directly with the helicase core domain, with the N-terminal IDR interacting with the RecA2 domain while the C-terminal IDR contacts the RecA1 domain (**Figure 3.16A**).

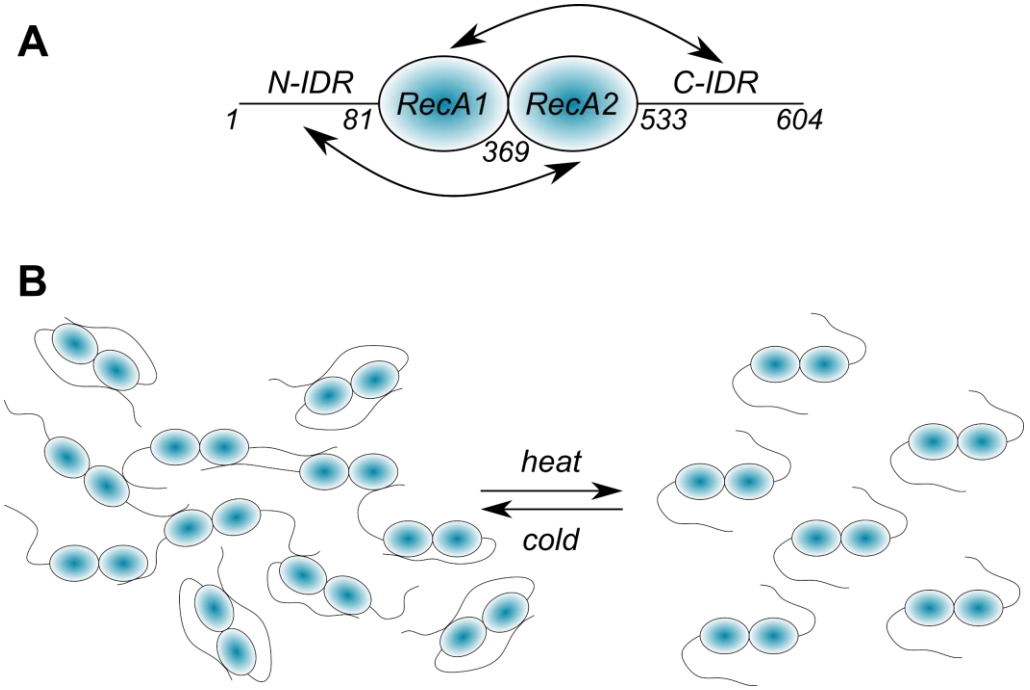


Figure 3.16: Interactions of the N- and C-terminal IDR with the helicase core domain of Ded1p. (**A**) Overview of the interactions between the N- and C-terminal IDR with the helicase core domain. (**B**) Proposed model for temperature-induced condensation of Ded1p. Under non-stress conditions, the N- and C-terminal extensions do not interact with the helicase core. Upon stress, in this case a decrease in temperature, the IDRs promote interactions between the helicases, allowing them to assemble and undergo LLPS.

While a sequence stretching from residue 21 to 30 was identified for the N-terminal IDR to be involved in the interaction, no binding site could be mapped for the C-terminal IDR due to the strong broadening of the resonances caused by LLPS. However, we observed that the N-terminal IDR:RecA2 interaction becomes weaker with increasing temperature. Based on our results, it can be speculated that the IDRs establish inter- and intramolecular interactions with decreasing temperature (**Figure 3.16B**).

LLPS was further demonstrated by the interaction between the C-terminal IDR and the RecA1 domain. Thus, our data suggest that the IDRs flanking the helicase core

provide protein-protein interactions that are one way to promote condensate formation. Taken together, our data further emphasize that a complex interplay between different regions of Ded1p influences condensation, mainly the interplay between the C-terminal IDR and the RecA1 domain, resulting in a further loss of signal intensity. However, maturation of LLPS-prone proteins cannot be excluded, and the loss of signal intensity is also a result of protein gelation.

Astonishingly, the Alphafold2 structure of the full-length Ded1p protein predicts an interaction between the RecA2 domain and the N-terminal IDR (**Figure 3.17A**). Our NMR data confirmed this interaction and also found that the same predicted residues are involved, both for the N-terminal IDR and for the RecA2 (**Figure 3.10** and **Figure 3.13**). Although we could not map the interaction between the C-terminal IDR and the RecA1 domain, Alphafold2 also predicts an interaction site for the C-terminal IDR at residues 555-559, which appear to associate with an α-helix of the RecA1 domain extending through residues 347-355 (**Figure 3.17B**). Region 555-559 in the C-terminal IDR contains a highly conserved RDYR motif that could be a candidate for interaction with the RecA1 domain. Due to LLPS, the interaction site could not be mapped in the NMR experiments. Given the surprisingly accurate Alphafold2 prediction for the N-terminal IDR, it will be interesting in the future to mutate the predicted site in the C-terminal IDR to verify the interaction.

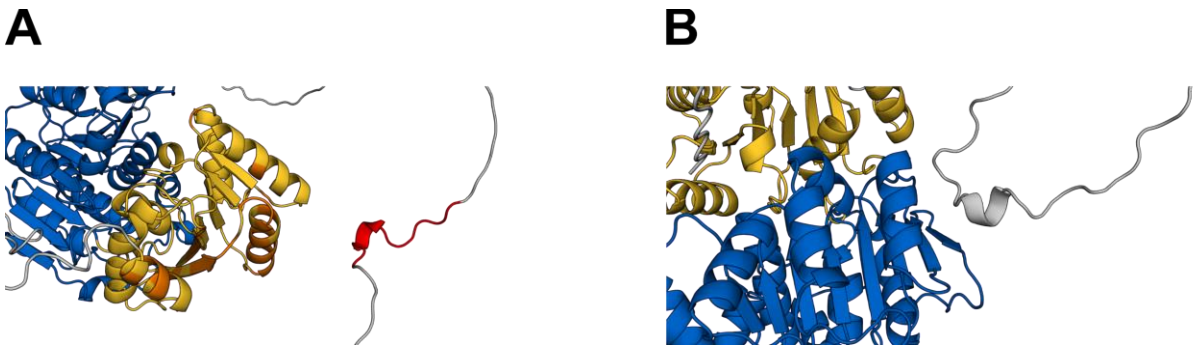


Figure 3.17: Alphafold2 model of *S. cerevisiae* Ded1p. The RecA1 domain is colored blue while the RecA2 domain is highlighted in yellow. Shown are the predicted interactions between the N-terminal IDR and the RecA2 domain (**A**) and the predicted interaction between the C-terminal IDR and the RecA1 domain (**B**) of Ded1p. In **A**, the N-terminal IDR:RecA2 interaction was confirmed by NMR titration experiments. Highlighted residues of the N-terminal IDR (red) show the sequence stretching from residue 21 to 30, which directly interacts with a surface-exposed helix stretching from residues 380-397 (orange residues) of the RecA2 domain.

Nevertheless, condensation of Ded1p with increasing temperature has been described *in vivo* (46, 226). In our *in vitro* studies, we visualized condensation with decreasing temperature. No condensation but only aggregation of Ded1p was observed when the temperature was increased. In addition, heat-induced assembly of Ded1p was shown to be reversible in yeast, but could not be reconstituted in *in vitro* studies (46). Such a finding suggests that additional cellular SG components, such as molecular chaperones and RNAs, are required for the disassembly of heat-induced Ded1p condensates, as it has already been shown for other heat-induced translation factor condensates (236, 237). Overall, the mechanism by which heat shock proteins assemble into condensates is not yet fully understood. This also implies further studies on the role of the different Ded1p regions, in particular the role of the N- and C-terminal IDR, and how they are affected by temperature changes.

In this work, NMR methods were used to elucidate the molecular mechanisms that drive the interaction between Ded1p helicases. We showed that the N- and C-terminal disordered extensions are not only required for Ded1p's role in translation initiation, but also participate in oligomerization by interacting with the folded helicase core of Ded1p. Many other proteins that condense upon temperature change contain folded and unfolded protein regions. Our proposed model of Ded1p condensation may therefore also be applicable to other translation initiation factors that condense upon temperature-induced stress.

3.5 Supplements

Supplemental Table 3.1: Protein constructs used in this study. All proteins start with an N-terminal GG that results from TEV cleavage.

Protein	Sequence	Internal Ref.
Ded1p RecA2 (369-535)	GGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLIFVETKRMADQLTDFLIMQNF RATAIHGDRTQSERERALAAFRSGAATLLVATAVAARGLDIPNVTHVINYDLPSDVDD YVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTEANQEVPSFLKDAMMS	#1735
Ded1p RecA1 (82-369)	GGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNFQSSGINFDNYDDIPVDASGKDVP EPITEFTSPPLDGLLLENIKLARFTKPTPVQKYSVPIVANGRDLMACAQTGSGKTGGF LFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMAPTRELATQIFDEAKKFTYRSWV KACVVYGGSPIGNQLREIERGCDLLVATPGRLNDLLERGKISLANVKYLVLDEADRML DMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPADIQHLARDFLSDYIFLSVGRVGS	#1739
Ded1p core (246-421)	GGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNFQSSGINFDNYDDIPVDASGKDVP EPITEFTSPPLDGLLLENIKLARFTKPTPVQKYSVPIVANGRDLMACAQTGSGKTGGF LFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMAPTRELATQIFDEAKKFTYRSWV KACVVYGGSPIGNQLREIERGCDLLVATPGRLNDLLERGKISLANVKYLVLDEADRML DMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPADIQHLARDFLSDYIFLSVGRVGST SENITQKVLYVENQDKKSALLDLLSASTDGLTLIFVETKRMADQLTDFLIMQNFRATA IHGDRTQSERERALAAFRSGAATLLVATAVAARGLDIPNVTHVINYDLPSDVDDYVHR IGRTGRAGNTGLATAFFNSENSNIVKGLHEILTEANQEVPSFLKDAMMS	#1740
Ded1p C-terminal extension (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSG	#2000
Ded1p N-terminal extension (1-81)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSN	#2269
Ded1p N-terminal extension (1-81), deletion of residues 21-35	GGMAELSEQVQNLSINDNNENGRNNSSNYNNNNGGYNGGRGGGSFFSNNRRGGYGNGG FFGGNNGGSRSN	#2317

Supplemental Table 3.2: Data collection and refinement statistics for *S. cerevisiae* Ded1p RecA1 82-369.

Ded1p RecA1 82-369 (PDB 8PNK)	
Data collection	
Space group	P 61
Cell dimension	
a, b, c (Å)	110, 110, 52.2
α, β, γ (°)	90, 90, 120
Resolution (Å)	45.78 - 2.2 (2.279 - 2.2)
R_{merge}	0.04857 (0.7115)
I/σ (I)	21.69 (2.48)
$CC_{1/2}$	1 (0.778)
Completeness (%)	99.66 (97.40)
Redundancy	6.6 (5.8)
Refinement	
Resolution (Å)	45.78 - 2.2 (2.279 - 2.2)
No. reflections	122286 (10246)
R_{work}/ R_{free}	0.1890/0.2258 (0.2367/0.2761)
No. atoms	
Protein	1981
Ligand	0
Water	83
B-factors ($Å^2$)	
Protein	56.09
Ligand	0
Water	51.36
R.m.s deviations	
Bond length (Å)	0.003
Bond angles (°)	0.54

Chapter 4: Tryptophan-Arginine sticker pairs determine Ded1p protein condensation

Contributions: The results of this chapter are in preparation for a publication. Therefore, this chapter is largely identical in text and figures to the manuscript:

Julian Hübner, Jan Philip Wurm, Valentin Thron, and Remco Sprangers. Tryptophan-Arginine sticker pairs determine Ded1p protein condensation, *in preparation*

All experiments were performed by myself. I was assisted by Philip Wurm in the setup and analysis of all NMR experiments. Some mutants for the LLPS assay were tested by Valentin Thron, an undergraduate student, under my supervision. Remco Sprangers, Philip Wurm, and I analyzed the data. The manuscript was written by Remco Sprangers and myself.

4.1 Introduction

Protein condensation, through a process that is often referred to as LLPS, is now recognized as a general strategy to dynamically and spatially organize cellular processes that are related to RNA metabolism (24, 238, 239). Mechanistically, the spontaneous de-mixing of biomolecules is the result of numerous redundant and intermolecular interactions between folded protein domains, disordered regions, and RNA, as well as direct contacts between IDRs and RNA (11, 52, 240). These transient interactions are of weak to moderate affinity, as a result, diffusion within the condensate and between condensates and the surrounding take place. However, the fluidity of condensates can be lost over time, and this maturation process has been linked to numerous diseases (75, 77, 78). Within condensates, many intermolecular interactions take place between folded protein domains, IDRs, and RNA. To describe the interactions between different IDRs in a simplified manner, the sticker-and-spacer model has been extensively used. In this model, direct interactions take place between several sticker residues that are separated by non-interacting spacer residues (95, 96, 109). Well-described stickers include phenylalanine, tyrosine, and arginine residues that interact through π-π, cation-π, and electrostatic contacts (102). Computationally (108) and based on NMR spectroscopic methods, the interactions between aromatic and arginine side chains have been observed for multiple IDRs (94, 95, 124). Experimentally, it has been well shown that the condensation propensity of an IDR

can also depend on the distribution of sticker amino acids and spacer regions within the highly dynamic protein chain (94).

SGs are cytoplasmic foci that form spontaneously when eukaryotic cells are exposed to stress conditions (49), and SGs contain mRNAs that are stalled in translation as well as numerous additional proteins involved in RNA metabolism. The dynamic nature of stress granules is reflected in the exchange of proteins between stress granules and other cytoplasmic foci, including P-bodies that are linked to mRNA storage and/or degradation. This notion has led to the suggestion that mRNAs are cycled between functionally different foci in a manner that targets transcripts for translational silencing, storage, or degradation. SGs and P-bodies contain multiple RNA helicases that are able to modulate intra- and intermolecular RNA-RNA interactions. For eIF4A, a DEAD-box RNA helicase that localizes to SGs, it has been shown that helicase activity reduces SG formation by reducing the number of intermolecular RNA-RNA interactions (241).

The *S. cerevisiae* Ded1p protein (DDX3 in humans) is another highly conserved DEAD-box RNA helicase and is found in both SGs and P-bodies (49, 184, 185). Ded1p is essential for translation initiation (173–175) and as a consequence Ded1p knockout cells are no longer viable (176). Ded1p interacts with the pre-translation initiation complex and scans the 5'UTRs of mRNAs where it can unwind secondary structure elements (160, 177). With this activity, Ded1p facilitates the translation of housekeeping proteins that contain structured RNA elements in the 5' UTR under growth conditions (46). Upon stress, these housekeeping mRNAs and Ded1p can be recruited into SGs, which down-regulates the translation of these mRNAs. Simultaneously, the translation of mRNAs that encode for stress-response proteins is up-regulated as these mRNA are not recruited to SGs and devoid of secondary structure elements in the 5' UTR, bypassing the requirement for Ded1p (46, 49).

On a structural level, DEAD-box RNA helicases consist of a highly conserved helicase core that comprises two folded RecA domains (152, 153). These RecA domains contain the sites that are important for RNA and ATP binding (206). During the RNA unwinding process, the relative conformation of the RecA domains changes from an open conformation, in which the RecA domains tumble independently, to a closed conformation in a manner that depends on the interactions with the substrate and

nucleotides (163, 208). The numerous DEAD-box helicases differ in the N- and C-terminal regions that flank the RecA core. These flanking regions are often disordered and are important for the interaction with substrates and other factors that regulate helicase activity. For Ded1p, the disordered C-terminal region also plays a role in protein oligomerization (242), which suggests a role of this region in protein condensation and potentially in SG formation.

Here we show that the C-terminal IDR of Ded1p is important for protein condensation, and we identified arginine and tryptophan residues as the central sticker residues.

4.2 Materials and Methods

4.2.1 Cloning, expression and purification

The gene encoding for the full-length Ded1p was acquired by amplification via polymerase chain reaction (PCR) from genomic DNA of *E. coli* BL21 DE3 cells. Furthermore, the gene was cloned into a pETM-11 plasmid which encodes for an N-terminal TEV protease cleavable His_6-tag. The gene encoding for the C-terminus of Ded1p was generated using Gibson assembly and was cloned into a pET-GB1 plasmid which additionally encodes for a N-terminal His_6-GB1 solubility tag followed by the TEV site. Point mutations were introduced by site-directed mutagenesis.

Plasmids encoding for full-length Ded1p constructs or mutants of the C-terminus of Ded1p were transformed into *E. coli* BL21 DE3 codon plus cells. Subsequently, cells were grown overnight at 37°C in 25 mL LB medium. This pre-culture was used the next day to inoculate the desired expression medium. Expression media were incubated at 37°C until an OD_{600} of 0.6-0.8 was reached. All growth media were supplemented with 50 mg/L kanamycin and 34 mg/L chloramphenicol. For the production of unlabeled protein, protein expression was induced by the addition of 1 mM IPTG in LB. Afterwards, cells were shifted to 20°C and harvested 12-18 h after induction. All cell pellets were stored at -20°C until further use. For the production of isotope-labeled protein, 25 mL LB culture was grown at 37°C. Cells were pelleted via centrifugation (6,000 g for 15 min at 20°C) and the cell pellet was used to inoculate M9 minimal medium containing 2 g/L ^{13}C-glucose and/or 0.5 g/L ^{15}N-NH_4Cl as only carbon or nitrogen source. Again, cells were incubated at 37°C until an OD_{600} of

0.6-0.8 was reached. 1 mM IPTG was added to induce protein expression, cells were shifted to 20°C, and cells were harvested 12-18 h after induction by centrifugation (6,000 g for 15 min at 20°C).

For the purification of full-length Ded1p constructs, cell pellets were thawed and resuspended in 25 mL Buffer A (25 mM NaPi pH 7.4, 1 M NaCl, 1 mM DTT, 10 mM imidazole) supplemented with 1:1000 (v/v) Triton-X-100, 1 mg/mL lysozyme, and 0.1 mg/mL DNase I. Resuspended cells were lysed by sonication and cell debris was removed by centrifugation (39,191 g for 30 min at 20°C). The supernatant was filtered using a 0.45 µm syringe filter and then applied to a gravity flow Ni-NTA column that was equilibrated with Buffer A. The column was then washed with five column volume of Buffer A and 10 mL of 25 mM NaPi pH 7.4, 2 M NaCl, 1 mM DTT buffer. Next, proteins were eluted with Buffer B (20 mM NaPi pH 7.4, 1 M NaCl, 1 mM DTT, 300 mM imidazole). The N-terminal His_6-tag was removed by TEV cleavage during dialysis against dialysis buffer (20 mM HEPES pH 7.3, 500 mM NaCl, 1 mM DTT) overnight at room temperature. The cleaved N-terminal His_6-tag was removed by a second Ni-NTA column that was equilibrated with dialysis buffer. Flow-through and wash fractions were combined, concentrated, and further purified by SEC using a 16/600 Superdex S75 column that was equilibrated with SEC buffer (20 mM HEPES pH 7.3, 500 mM NaCl, 1 mM DTT). Protein containing fractions were pooled, concentrated, and the final protein concentration was determined based on the A_{280}. Ded1p C-terminus constructs were expressed and purified in the same manner. In their case, Buffer A and Buffer B did not contain DTT but were supplemented with 1 M GdmCl. For NMR experiments, SEC buffer was exchanged to NMR buffer (20 mM HEPES pH 7.3, 125 mM NaCl). All proteins were frozen in liquid nitrogen and stored at -80°C until they were used for further experiments.

4.2.2 NMR experiments

All NMR experiments were performed on Bruker 500, 600, and 800 MHz Avance Neo spectrometers equipped with nitrogen-cooled (500 and 600 MHz) or helium-cooled (800 MHz) cryoprobes at 4°C in NMR buffer supplemented with 5% (v/v) D_2O and 200 µM trimethylsilylpropanoic acid (TSP). NMR samples had a concentration of 50-100 µM or 300-500 µM for backbone assignment experiments. For 2D spectra, ^{1}H-^{15}N

HSQC spectra were recorded and for the backbone assignment, standard Bruker pulse sequences for HNCACB, HNcoCACB, HNCO, HNcaCO, HCccoNH, HNcaNNH were used.

In order to determine the ^{15}N R_2 relaxation rates, NMR experiments were run with different delay times of 0 ms, 17 ms, 33.9 ms, 50.9 ms, 67.8 ms, 88 ms, and 101.8 ms. R_2 rates were recorded at a field strength of 800 MHz. Loss of intensities for each assigned resonance were fitted with a mono-exponential function using in house written Matlab scripts (Equation 10).

Equation 10 $$I = I_0 + \Delta I * e^{(-k*t)}$$

where I is the signal intensity, I_0 the initial signal intensity, ΔI is the difference in signal intensity, k is the rate constant, and t is the time.

The bell-shaped R_2 profiles were fit according to (Equation 11) (243).

Equation 11 $$R_2(i) = R_2^{intrin} \sum_{j=1}^{j=N} \exp\left(-\frac{|i-j|}{\lambda}\right)$$

where i is the residue number, λ is the chain persistence length, N is the chain length, and R_2^{intrin} is a scaling factor that depends in the intrinsic R_2 rate.

All ^{1}H-^{15}N CSA/DD cross-correlated transverse and longitudinal relaxation rates were recorded as it is described in (244). Experiments were conducted in 20 mM MES pH 6.3, 125 mM NaCl buffer at 800 MHz.

$^{1}H^{N}$-$R_{1,\rho}$ relaxation dispersion experiments were conducted in 20 mM MES pH 6.3, 125 mM NaCl buffer at 800 MHz on deuterated protein. The experiments were performed using a published pulse sequence (245). Here, a spin-lock length of 120 ms was applied.

All NMR spectra were processed using either Topspin 4.02 or NMRPipe 9.6 (218). For the analysis of the spectra, NMRDraw or Cara 1.9.2 (228) was used.

4.2.3 PRE measurements

For spin labeling, NMR-inactive cysteine mutants were incubated in the presence of a 5-fold excess of 4-maleimido-TEMPO in NMR buffer at room temperature for 2 h in the dark. Free 4-maleimido-TEMPO was removed by applying the solution onto a

PD10 Desalting Column. All PRE experiments were conducted in NMR buffer at 4°C. For this, 50 µM of NMR-active WT C-terminus was mixed with 100 µM of an NMR-inactive mutant carrying the spin label. ^{1}H-^{15}N HSQC spectra were recorded before and after reduction of the spin label by the addition of 2 mM sodium ascorbate. In both spectra, the volume of the peaks was integrated using NMRPipe and the PRE effect was calculated as the ratio of the peak volume before reduction (I) divided by the peak volume after reduction (I_0).

4.2.4 Phase separation assay

For sample preparation, the respective proteins were thawed at 30°C and subsequently centrifuged at 21,130 g for 5 min to remove aggregated protein. To induce the de-mixing of the protein into a protein-light and protein-dense phase, samples were placed in a centrifuge that was pre-cooled to 4°C. Samples were spun at 21,130 g for 2 h so protein droplets settle down. For the first 30 min, aliquots of the protein-light phase were taken every 5 min. For the remaining 90 min, aliquots of the light phase were taken every 15 min. For each aliquot, the protein concentration was determined three times and the concentrations were averaged. The concentrations at each time point were plotted and the data were fitted with a monoexponential function using in-house written Matlab scripts. The baseline resembles the c_{SAT}.

Equation 12 $$C = C_0 + \Delta C * e^{(-k * t)}$$

where C is the protein concentration, C_0 is the initial protein concentration, ΔC is the difference in protein concentration, k is the rate constant, and t is the time.

4.2.5 Helicase assay

A fluorescence-based single turnover helicase assay was performed in 96-well plates using a TECAN spark plate reader at 25°C. A 15mer hybrid RNA was used as a template for unwinding. Fluorescein labeled RNAs were obtained from either IDT or Eurofins Genomics and carried a fluorescein label at their 5' end. 1 µM labeled 15mer RNA (5'-GGAACAGUACGCUUG-3') and 1.5 µM of its counterpart (5'-GGCAAAUGAGGUCCCAAGCGUACUGUUCCGUC-3') were hybridized in NMR buffer supplemented with 5 mM $MgCl_2$ by heating up the mixture to 95°C for 1 min

followed by cooling down to room temperature. To prevent reannealing after the unwinding reaction, an excess of unlabeled 15mer ssRNA was added to the reaction mix. 100 µL of the reaction mixture contained 20 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$, 0.002% (v/v) Triton-X-100, 0.5 µM protein, 25 nM hybrid RNA, and 2 µM unlabeled 15mer ssRNA. All components were mixed in the well and incubated in the plate reader for 5 min. The helicase unwinding activity was finally started by the addition of 1 mM ATP. The decay of fluorescence intensity was recorded every 30 s for 1 h. Excitation and emission wavelengths were 485 nm and 535 nm, respectively. Experiments were recorded in technical triplicates for biological triplicates, and the data were finally fitted with a linear function using in house written Matlab scripts (Equation 13).

Equation 13 $$F = F_0 + \Delta F * t$$

where F is the fluorescence intensity, F_0 is the basal fluorescence intensity, ΔF is the difference in fluorescence intensity, and t is the time.

4.2.6 Fluorescence anisotropy

All fluorescence anisotropy measurements were performed in NMR buffer supplemented with 5 mM $MgCl_2$ and 0.002% (v/v) Triton-X-100. In all experiments, an RNA concentration of 25 nM was used. Solutions with increasing protein concentration were prepared independently. To determine affinities of the closed core conformation for ssRNA, a 5´-fluorescein labeled 15mer 5´-GGAACAGUACGCUUG-3´ RNA was used. The closed conformation was induced by the addition of 3 mM nucleotide, in this case ADP in addition to 3 mM BeF_2, and 10 mM NaF. For the RNA affinity measurements of the open conformation, no nucleotide was added. Experiments were recorded in triplicates of biological duplicates using a TECAN spark plate reader at 25°C. Excitation and emission wavelengths were 485 nm and 535 nm, respectively. Data were analyzed and fitted to a one-site binding model using in house written Matlab scripts (Equation 14).

Equation 14 $$A = A_0 + \Delta A * \left(\frac{[R] + [D] + K_D}{2} - \sqrt{\left(\frac{[R] + [D] + K_D}{2}\right)^2 - [D] * [R]} \right)$$

where A is the anisotropy, A_0 is the anisotropy of the free RNA, ΔA is the increase in anisotropy due to binding, [R] is the concentration of fluorescein-labeled RNA, [D] is the Ded1p concentration, and K_D is the dissociation constant of the protein/RNA complex.

4.2.7 Fluorescence microscopy

Solutions for fluorescence microscopy analysis were prepared and performed in 384-well plates at 20 µL reaction mix. The concentration of the 15mer hybrid RNA was lowered to 5 nM in order to reduce background noise. 20 µL of the reaction mixture contained 20 mM HEPES pH 7.3, 125 mM NaCl, 1 mM DTT, 5 mM $MgCl_2$, 0.002% (v/v) Triton-X-100, 0.5-2 µM protein, 5 nM hybrid RNA, and 2 µM unlabeled ssRNA. The closed protein state was induced by the addition of 3 mM ADP in addition to 3 mM BeF_2, and 10 mM NaF. RNA and protein were added last in order to prevent maturation of droplets that could have formed during preparation. Microscopy was conducted using a Keyence BZ-X810 fluorescence microscope at room temperature.

4.3 Results

4.3.1 The isolated C-terminal IDR of Ded1p is capable of condensation

To quantify the propensity of Ded1p to condensate into a dense phase, the protein solution is concentrated to a level at which a protein dense phase coexists with a protein light phase. The dense protein condensates are subsequently removed by rapid centrifugation, after which the concentrations of the protein light phase are at the c_{SAT}. A low c_{SAT} implies that a protein has a high propensity to form condensates, whereas a high c_{SAT} is representative of a protein with a low propensity to form condensates. For the full-length Ded1p protein, we determined a c_{SAT} of 14 µM (**Supplemental Figure 4.1**), indicating that the protein readily condenses even in the absence of RNA or other factors. The C-terminal IDR of Ded1p is highly important for Ded1p condensation as a version of the helicase that lacks this region is impaired in self-aggregation (**Supplemental Figure 4.1**). The central role of the C-terminal IDR

(**Figure 4.1**) is further corroborated by the observation that the isolated C-terminal IDR condenses at a concentration that exceed 55 µM (**Supplemental Figure 4.1**).

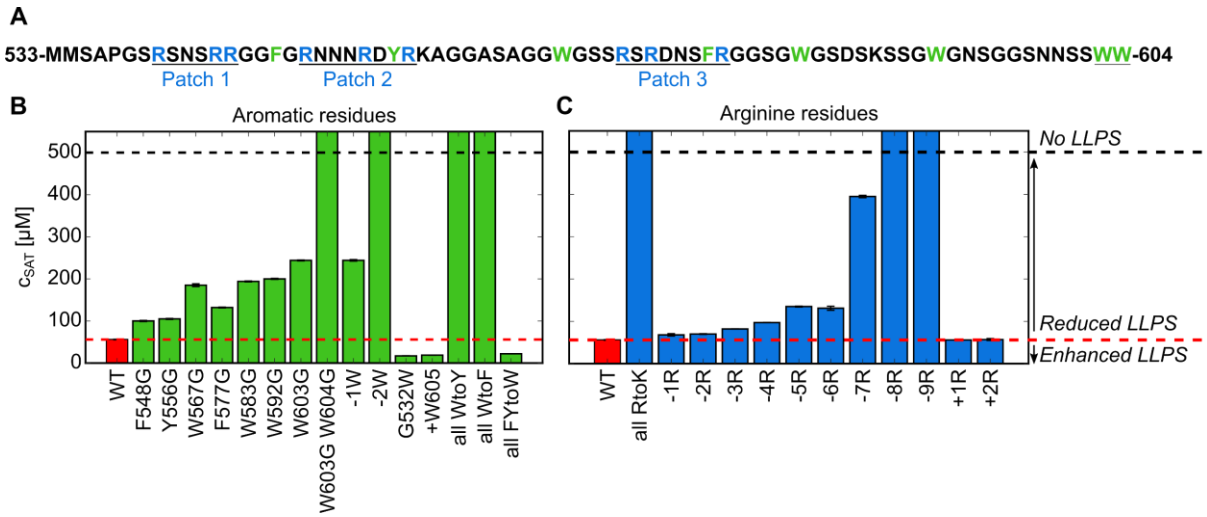


Figure 4.1: Saturation concentration of the C-terminal IDR. (**A**) Amino acid sequence of the Ded1p C-terminal IDR. Aromatic residues (Phe, Tyr, and Trp) are highlighted in green, while Arg residues are highlighted in blue. (**B**) The c_{SAT} for the WT C-terminal IDR (red) and for versions of the C-terminal IDR in which aromatic residues were mutated (green). The red dashed line represents the c_{SAT} of the WT C-terminal IDR. Bars that are lower (higher) than this red line indicate a higher (lower) tendency to form a condensed phase. No condensation was observed for protein versions for which the bar is higher than the black dashed line (500 µM). (**C**) The c_{SAT} of protein versions for which arginine residues were mutated (blue).

Next, we aimed to identify residues in the Ded1p C-terminal IDR that are responsible for condensation. Initially, we focused on the tyrosine and two phenylalanine residues, as these amino acids have been previously identified as sticker residues in other IDRs. Substitution of a single tyrosine (Y556G) or phenylalanine (F548G or F557G) residue slightly increased the c_{SAT} (**Figure 4.1B**), indicating a minor role for these residues in Ded1p condensation. Next, we focused on the five tryptophan residues, amino acids that have not been generally assigned as strong sticker residues. Interestingly, we found that substitution of any single tryptophan residue (W567G, W583G, W592G, W603G, or W604G; where residue 604 is the C-terminal amino acid in Ded1p) had a stronger impact on protein condensation than removal of the other aromatic residues (**Figure 4.1B**). This finding implies that tryptophan is a stronger sticker residue than tyrosine or phenylalanine. Interestingly, mutation or removal of two tryptophan residues (W603G+W604G or -2W) impaired protein condensation to such a degree that condensation could no longer be observed up to a protein concentration of 0.5 mM (**Figure 4.1B**). The importance of tryptophan residues for the condensation of Ded1p

is further corroborated by three observations. First, the addition of additional tryptophan residues either at the N-terminal (G532W) or at the C-terminal end of the IDR (+W605) significantly reduced the c_{SAT} below 17 µM. Second, mutating all tyrosine and phenylalanine residues to tryptophan residues (all FYtoW) also reduce the c_{SAT} significantly to below 22 µM, and third, mutation of all tryptophan residues to either tyrosine (all WtoY) or phenylalanine (all WtoF) fully abolished protein condensation in our assay (**Figure 4.1B**). Based on these findings, we conclude that tryptophan functions as a highly potent sticker residue that drives the condensation of the C-terminal region of Ded1p.

Furthermore, we focused on the arginine residues that have previously been identified as sticker residues. To that end, we mutated all nine arginine residues to lysine residues (all RtoK), which abolished protein condensation of the Ded1p C-terminal IDR (**Figure 4.1C**). This finding confirms that arginine is also a sticker residue in Ded1p. Next, we set out to explore if all arginine residues are active stickers and mutated an increasing number of arginine residues to glycine residues. Interestingly, removal of up to six arginine residues (-1R, -2R, -3R, -4R, -5R, and -6R) only slightly increased the c_{SAT} (**Figure 4.1C**). Only upon removal of seven (-7R), eight (-8R), or all nine (-9R) arginine residues, a strong reduction in protein condensation was observed. This finding implies that the five tryptophan residues in the C-terminal IDR of Ded1p interact with three arginine residues to induce protein condensation. The importance of this 5W:3R ratio is underscored by the observation that the addition of additional arginine residues (+1R, +2R) does not enhance protein condensation as there is an excess of arginine residues in the WT sequence (**Figure 4.1C**). The addition of tryptophan residues does promote condensation as the WT sequence has a shortage of these residues (**Figure 4.1B**).

To unravel whether the order or sequence distribution of tryptophan or arginine residues in the protein is important for condensation, we moved these amino acids in the amino acid sequence. In a first experiment, we reversed the sequence of the entire IDR, which did not influence the c_{SAT} (**Supplemental Figure 4.2**). Protein variants for which the aromatic residues were equally spaced throughout the protein sequence have a slightly lower tendency to condensate. The spacing of the arginine residues plays no role for protein condensation, since neither the clustering of these positive charges at the N-terminal part of the sequence nor the equal distribution of these

residues in the sequence influenced the c_{SAT} (**Supplemental Figure 4.2**). Based on these experiments, we conclude that the order and distribution of sticker residues throughout the sequence is not important for protein condensation. This result contradicts with findings for the human RNA helicase DDX4 and for the Nephrin intracellular domain, where clusters of positively and negatively charged residues were a prerequisite for protein condensation (67, 246). In the latter two cases, condensation is largely based on electrostatic interactions. The shuffling of charged residues then results in a disappearance of locally charged regions and subsequently intermolecular interactions are lost. In the case of the Ded1p IDR, the clustering or distribution of tryptophan and arginine residues does not influence local properties to a degree that protein condensation is affected.

4.3.2 The intermolecular contacts are fuzzy

We have established above that tryptophan and arginine residues are responsible for the condensation of the C-terminal IDR of Ded1p. Next, we made use of NMR spectroscopy to observe the contacts between these residues. To that end, we recorded ^{1}H-^{15}N HSQC spectra, which provide residue-specific information (**Figure 4.2A**). It should be noted that these spectra are recorded on samples for which the concentrations are below c_{SAT}. Recording NMR spectra above c_{SAT} is not possible as this would result in rapid and irreversible maturation of the protein into a gel-like state. However, since the same intermolecular contacts take place below and above c_{SAT}, as we recently showed for the P-body protein Edc3 (74), this approach allows us to identify the initial and central contacts that result in protein condensation. In our NMR experiments, we followed the resonance frequencies of the two C-terminal tryptophan residues (W603, W604) upon a step-wise removal of the all nine arginine residues and observed that this resulted in a stepwise change of the tryptophan resonance frequencies (**Figure 4.2A**). Since the tryptophan and arginine residues are far apart in the primary sequence (**Figure 4.1A**), the CSPs of the tryptophan residues must result from a change in the arginine-tryptophan interactions and not from direct effects due to the mutations. We conclude that the arginine and tryptophan residues are in direct contact, likely through cation-π stacking interactions. Interestingly, the removal of a single arginine residue is already sensed by the tryptophan residues,

albeit this has no effect on protein condensation. This observation implies that arginine-tryptophan interactions are highly dynamic and the single arginine-tryptophan interactions are very short-lived.

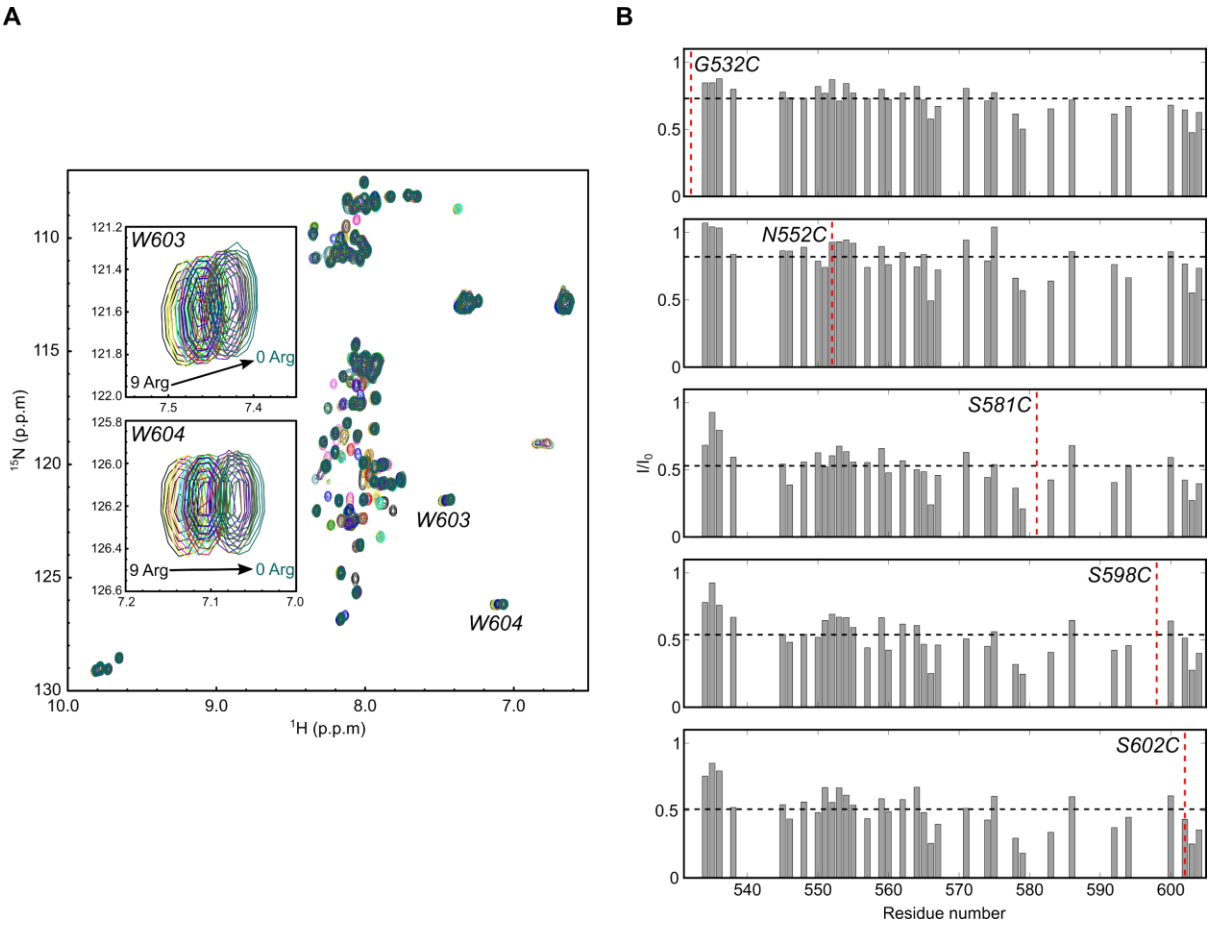


Figure 4.2: Observation of contacts between Ded1p C-terminal IDRs. (**A**) Overlay of ^{1}H-^{15}N HSQC NMR spectra of the wild-type Ded1p C-terminal IDR and versions in which an increasing number of Arg are removed (-1R to -9R), which results in a step-wise CSP of W603 and W604. (**B**) Intermolecular PRE measurements were performed by attaching a spin-label to an NMR-inactive IDR at five different positions. The relative resonance intensities in the NMR-invisible IDR (I/I_0) were plotted against the residue number. The average intensity is indicated by black dashed lines. The positions of the spin-label in the NMR-inactive IDR are indicated by red dashed lines. Since PREs are intermolecular, the site where a spin-label is attached is not necessarily fully bleached. All experiments were performed at identical concentrations that are below the c_{SAT}.

Based on the observed tryptophan CSPs upon removal of arginine residues, it is not possible to deduce whether the tryptophan-arginine contacts are intra- or intermolecular. To directly observe intermolecular contacts between Ded1p C-termini, we made use of NMR PRE measurements (**Figure 4.2B**). To that end, we attached a paramagnetic spin-label to an NMR-invisible (unlabeled) protomer and mixed it with

an NMR visible (^{15}N labeled) protomer. Upon interaction of the NMR-invisible spin-labeled IDR with the NMR-visible IDR, significant line-broadening of resonances are expected for those residues that come close to the spin label in the NMR-inactive protomer. In total, we attached the spin-label to five different positions in the Ded1p IDR (one at a time). When we attached the spin-label at the N-terminal part of the IDR (532 or 552), we observed moderate line-broadening for all resonances (I/I_0 ~ 0.7, where I and I_0 are the resonance intensities of the presence and absence of the spin-label, respectively). In the case where we have attached the spin-label towards the C-terminal part (581, 598, or 602) we observe stronger PREs (I/I_0 ~ 0.5) for all resonances, apart from the far N-terminal residues. Based on this observation we conclude that the N-terminal part of the IDR is less involved in intermolecular contacts than the C-terminal part. This observation is in agreement with the distribution of tryptophan residues that are located in the C-terminal half of the sequence. In addition, we conclude that there are no stable specific contacts between the protomers, as all resonances are similarly affected by the spin-label, underscoring the transient nature of the intermolecular interactions.

4.3.3 Chain mobility is linked to the degree of LLPS

The PRE effects that we measured can reach distances of 20 Å. As a result, the extracted information is of intermediate resolution, and it is not possible to obtain residue-specific contact information. To complement the PRE data, we here make use of ^{15}N R_2 relaxation experiments, which can probe the mobility of the IDR on a per-residue basis. For an ideal IDR that lacks intra- and intermolecular interactions, one expects an R_2 vs. residue profile that follows a bell-shaped profile, where the mobility at both ends of the chain is enhanced (R_2 rates are reduced) and where the mobility in the middle of the chain is more or less homogeneous (243). Interestingly, the relaxation rates of the WT C-terminal IDR of Ded1p deviate significantly from such a bell-shaped profile (**Figure 4.3A**).

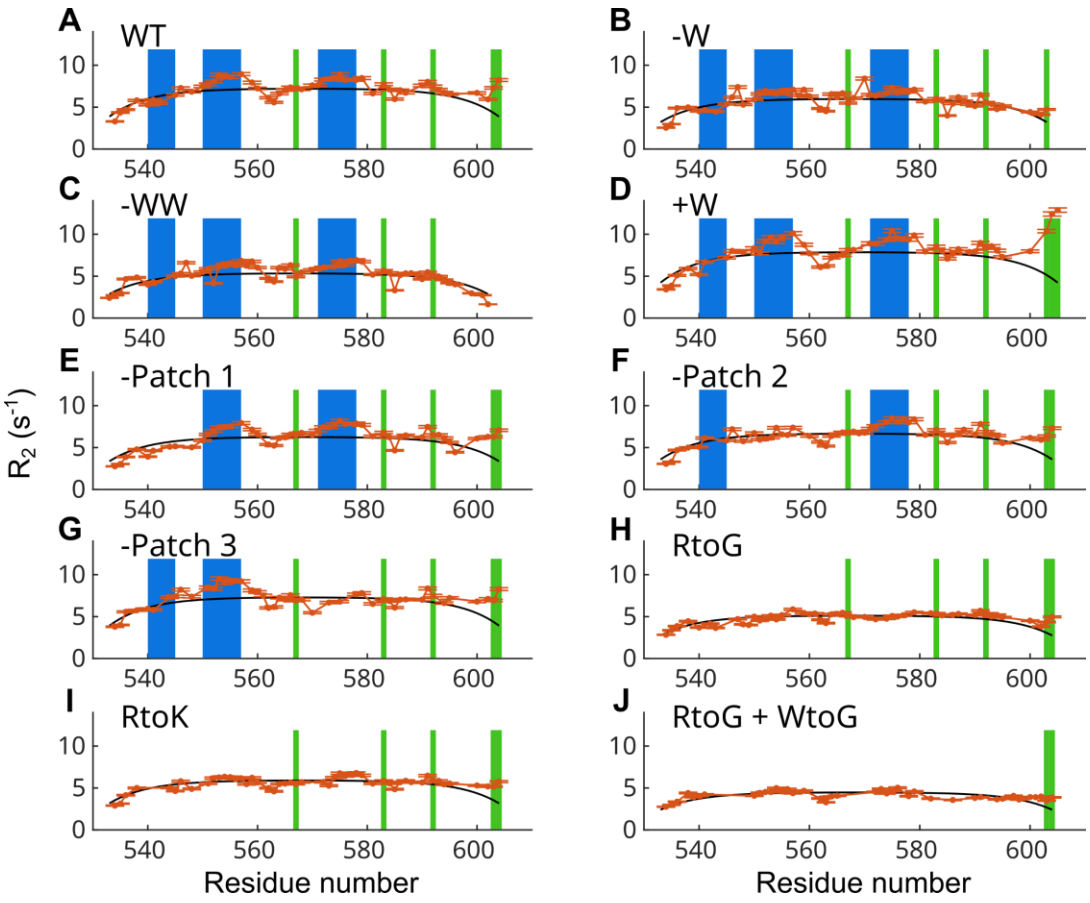


Figure 4.3: ^{15}N R_2 relaxation rates of the backbone amide groups of the WT and mutant versions of the Ded1p C-terminal IDR. (**A**) The WT C-terminal IDR experiences elevated relaxation rates at the far C-terminal region where the WW motif is located, as well as at two of the three arginine patches (highlighted in blue). The tryptophan locations are highlighted in green. (**B**) as in **A**, but for a construct of the protein that lacks the C-terminal tryptophan (-W), (**C**) the two C-terminal tryptophan residues (-WW), (**D**) contains an additional C-terminal tryptophan residue (+W), (**E**) lacks the first three arginine residues (-patch 1), (**F**) lacks the second set of three arginine residues (-patch 2), (**G**) lacks the last three arginine residues (-patch 3), (**H**) lacks all arginine residues due to mutations to glycine (RtoG), (**I**) lacks all arginine due to mutations to lysine (RtoK), (**J**) or that contains none of the arginine or aromatic residues apart from the two C-terminal tryptophan residues.

In particular, the R_2 relaxation rates are enhanced in the far C-terminal region of the IDR that harbors a pair of tryptophan residues, and in two of the three arginine rich patches. These enhanced R_2 relaxation rates can result from locally reduced mobility in the backbone on the ps-ns timescale or from a chemical exchange process on the µs-ms timescale. To assess whether the R_2 profiles contain contributions from a chemical exchange process, we determined the transverse cross-correlation rates ($\eta_{xy,z}$). These rates are well suited to report on the degree of order in IDRs (247) and, as opposed to R_2 rates, are insensitive to chemical exchange processes (248, 249).

The cross-correlated relaxation rates show the same residue-specific variations as the R_2 rates do, indicating that the variations in the R_2 rates are not due to µs-ms chemical exchange processes (**Supplemental Figure 4.3**). The lack of exchange broadening is further confirmed by amide proton relaxation dispersion experiments (**Supplemental Figure 4.4**), which show flat profiles for all residues. Taken together, we conclude that the variations in the R_2 profiles result from differences in local chain mobility that is caused by short lived interactions between side chains.

To obtain direct insights into the interactions between arginine and tryptophan side chains, we mutated residues that are important for protein condensation (**Figure 4.1**) and assessed whether the mobility in the protein chain is enhanced accordingly as interactions are lost. Removal of the C-terminal tryptophan residue (-W) indeed resulted in an enhanced mobility of the C-terminal region (**Figure 4.3B**). Removal of both C-terminal tryptophan residues (-WW) enhanced the mobility of the C-terminal region further (**Figure 4.3C**). At the same time, the addition of a tryptophan residue to the C-terminus (+W) significantly enhanced the R_2 rates in this region of the protein (**Figure 4.3D**). These data are fully consistent with a model in which tryptophan residues participate in interactions that are responsible for protein condensation. Next, we focused on the three arginine patches in the sequence, of which patch 2 and patch 3 clearly display reduced mobility. In agreement with that, the mutation of patch 1 (-patch 1; arginine to glycine mutations) did not affect local mobility significantly (**Figure 4.3E**). The mutation of patch 2 (-patch 2) or patch 3 (-patch 3) on the other hand resulted in the clear increase of the mobility in these regions (**Figure 4.3E, F**). The removal of three arginine residues in these patch 1, patch 2, and patch 3 mutants does not significantly affect protein condensation as the chain contains sufficient other arginine residues (six) to interact with the tryptophan residues. This is also reflected in the mobility of the C-terminal tryptophan residues, which remained reduced when one of the three arginine patches were mutated. Next, we removed all arginine residues by mutation to either glycine (RtoG) or lysine (RtoK). In these mutant IDRs, the mobility of the C-terminal tryptophan residues is indeed enhanced (**Figure 4.3H, I**) as the arginine interacting partners are lacking. We explain the observation that the mobility of the two C-terminal tryptophan residues is still not as high as for the N-terminal glycine residues by the bulky nature of these aromatic residues (243). As a final test, we also designed a protein that lacks all arginine,

tyrosine, phenylalanine, and tryptophan residues apart from the far C-terminal tryptophan pair (RtoK + YtoG + FtoG + 3WtoG). This did not further enhance the mobility of the C-terminal tryptophan residues and indicates that tryptophan residues do not directly contact other tryptophan residues, tyrosine, or phenylalanine residues. Based on the above, we conclude that tryptophan and arginine residues are involved in direct contacts that result in protein condensation.

4.3.4 Tryptophan-Arginine interactions are also important for Ded1p condensation

Next, we verified that our findings on the isolated Ded1p C-terminal region can be transferred to the full-length protein. To that end, we determined the influence that tryptophan and arginine mutations in the C-terminal IDR have on the c_{SAT} of the full-length protein. Removal of the entire C-terminal IDR significantly increased the c_{SAT} of the full-length protein from 14 µM to more than 230 µM (**Figure 4.1** and **Supplemental Figure 4.1**), and in agreement with our findings above, removal of one (-1W) or two (-2W), as well as the mutation of all C-terminal arginine residues to either lysine (RtoK) or arginine (RtoG) reduced the potency of full-length Ded1p to condensate (**Figure 4.4**).

Furthermore, the addition of a single tryptophan residue to the C-terminal IDR (+W) enhances the propensity for condensation (**Figure 4.4**). In that light, our findings on the isolated IDR can be directly transferred to the full-length protein. Nevertheless, it is worth mentioning that the c_{SAT} of the full-length protein (14 µM) is lower than that of the isolated C-terminal IDR (56 µM), which implies that there are additional weak intermolecular contacts that involve residues in the N-terminal IDR and/or the folded RecA domains that contribute to protein condensation.

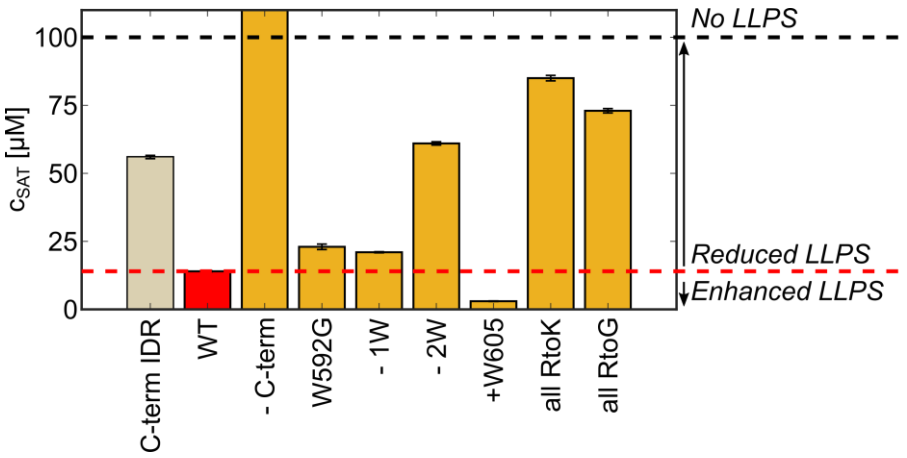


Figure 4.4: Saturation concentrations of the Ded1p full-length constructs. The c_{SAT} of full-length Ded1p (WT, red bar) is lower than the c_{SAT} of the isolated C-terminal IDR (grey bar), whereas the c_{SAT} of the Ded1p in the absence of the C-terminal IDR (-C-term) is above the concentration that we could achieve. Substitution (W592G) or removal of one (-1W), two (-2W) C-terminal tryptophan residues, or the mutation of the arginine residues in the C-terminal IDR to either lysine (all RtoK) or glycine (all RtoG) residues, significantly impairs the condensation propensity of the full-length protein. Addition of a tryptophan (+W605) to the C-terminal IDR of full-length Ded1p lowers the c_{SAT}, as was observed in the isolated C-terminal IDR. The red dashed line represents the c_{SAT} of the full-length WT protein. Bars that are lower (higher) than this red line indicate a higher (lower) tendency to form a condensed phase. No condensation was observed up to a protein concentration of 0.1 mM for protein versions for which the bar is higher than the black dashed line.

4.3.5 Ded1p condensation is independent of helicase activity

Ded1p destabilizes or unwinds dsRNA regions in an ATP-dependent and non-processive manner by docking onto the substrate at a region that is locally single-stranded (169). Interestingly, Ded1p activity depends on the enzyme concentration (**Supplemental Figure 4.5**), which led to the notion that the protein acts as a trimer and that oligomerization of Ded1p provides a means to regulate helicase activity (242), although it remains unclear how oligomerization and activity are structurally linked. Ded1p versions that lack the C-terminal IDR are less active, indicating that the C-terminal region is important for oligomerization (242). To test the relationship between protein condensation (that is an extreme of oligomerization) and enzyme activity, we made use of three minimal mutations, first the removal of the two C-terminal tryptophan residues (-2W), second the mutant in which all arginine residues are replaced by lysine residues (RtoK) and third the mutant that contains an additional tryptophan residue (+W). The first two of those mutants reduce protein condensation, whereas the latter one enhances protein condensation (**Figure 4.4**). As a control, we

made use of the WT protein and the protein that lacks the entire C-terminal IDR (-C-term).

Addition of RNA to the Ded1p protein enhances the propensity for the formation of condensates significantly and we next set out to identify protein concentrations at which the WT protein and the C-terminal mutants display a difference in phase separation behavior in the presence of both RNA and the ATP analogue ADP-BeF_3, which can mimic turnover conditions as closely as possible. Using a fluorescent wide-field microscope, we determined that the Ded1p mutant that contains an additional tryptophan formed visible droplets at protein concentrations of 0.5, 1.0, and 2.0 µM (**Figure 4.5**). The WT protein displays visible droplets only at and above a concentration of 1.0 µM, whereas the mutant that lacks the two C-terminal tryptophan residues does not condensate up to a concentration of 2.0 µM. These results confirm the importance of the tryptophan residues for protein condensation, even in the presence of RNA and nucleotides.

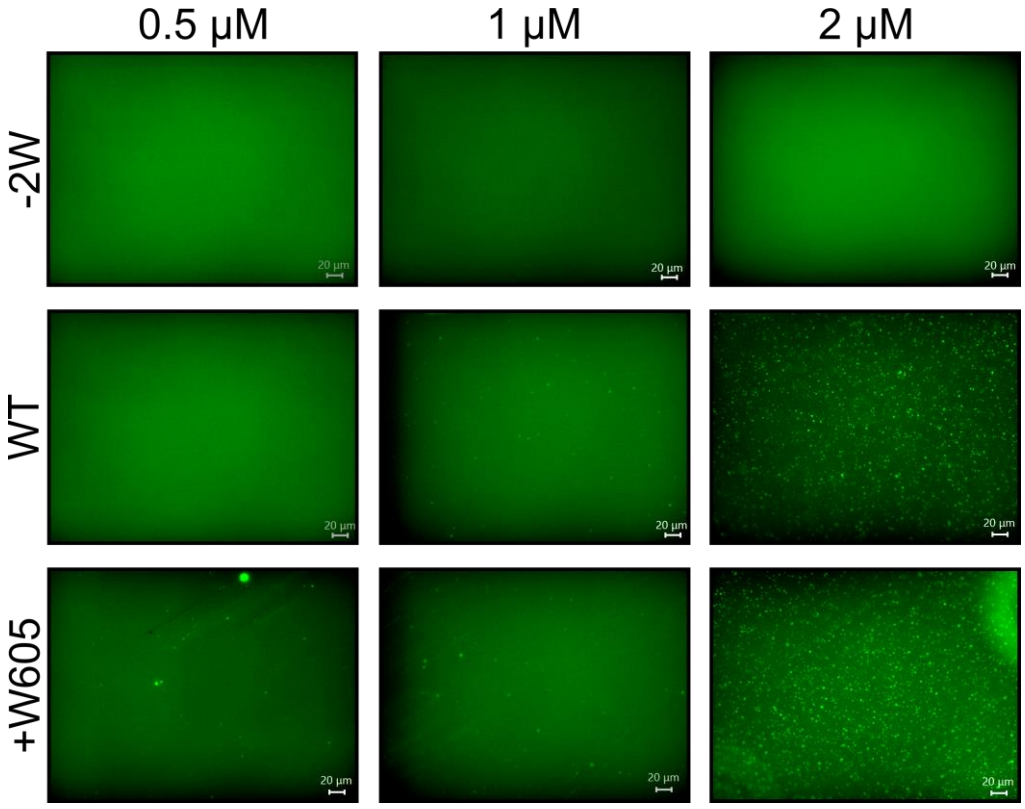


Figure 4.5: Droplet formation of Ded1p full-length variants. Ded1p forms droplets in the presence of a 15mer hybrid RNA. At 1 µM protein concentration, condensates are observable for the WT while already 0.5 µM protein are enough for the +W605 mutant to form droplets. The protein variant lacking both C-terminal Trp showed no droplet formation.

To ensure that the mutations in the enzyme do not influence the interactions with RNA, we determined the binding affinities of the WT and mutant Ded1p enzymes for a ssRNA (**Figure 4.6A**). First, we performed experiments in the absence of nucleotides and found that the WT protein interacts with ssRNA with intermediate affinity (k_D of 1.6 ± 0.20 µM). This affinity is reduced (k_D of 5.2 ± 0.22 µM) upon removal of the C-terminal region, indicating that this IDR plays a role in the interaction with the substrate. Importantly, the conservative mutations that replace tryptophan with glycine or arginine with lysine had negligible effects on substrate interactions. Second, we performed experiments in the presence of ADP-BeF_3 such that a closed helicase:substrate:nucleotide complex can be formed (**Figure 4.6A**). In this state, the enzyme interacts with ssRNA through conserved sequence motifs in both RecA domains (156, 163). As expected, the affinity of the WT protein for RNA is significantly enhanced (k_D of 44 ± 13 nM) by the presence of the ATP analogue. In this state, the binding affinity that the core contributes to the enzyme:RNA interaction is thus large and deletion or modifications of the C-terminal IDR no longer play a significant role in the overall binding process. Taken together, we conclude that mutations in the C-terminal tail of Ded1p have no effect on the protein:RNA interactions.

Next, we performed RNA unwinding assays (**Figure 4.6B**) at an enzyme concentration of 0.5 µM, which is far below the concentration at which the protein that lacks the C-terminal tryptophan residues phase separates, and just below the concentration where the WT Ded1p displays visible droplets, but sufficiently high such that the variant with the extra tryptophan (+W605) displays visible condensates. We argued that if protein condensation would influence helicase activity, then this should be best visible under these conditions. Based on nine measurements (three biological replicates that were each performed with three technical replicates), it appears that the activity of Ded1p is highest for the protein variants that are most prone to undergo condensation. However, the observed differences are small and not statistically highly relevant ($p > 0.01$). Based on that we conclude that Ded1p activity depends on the C-terminal IDR, but that this is not directly correlated with protein condensation.

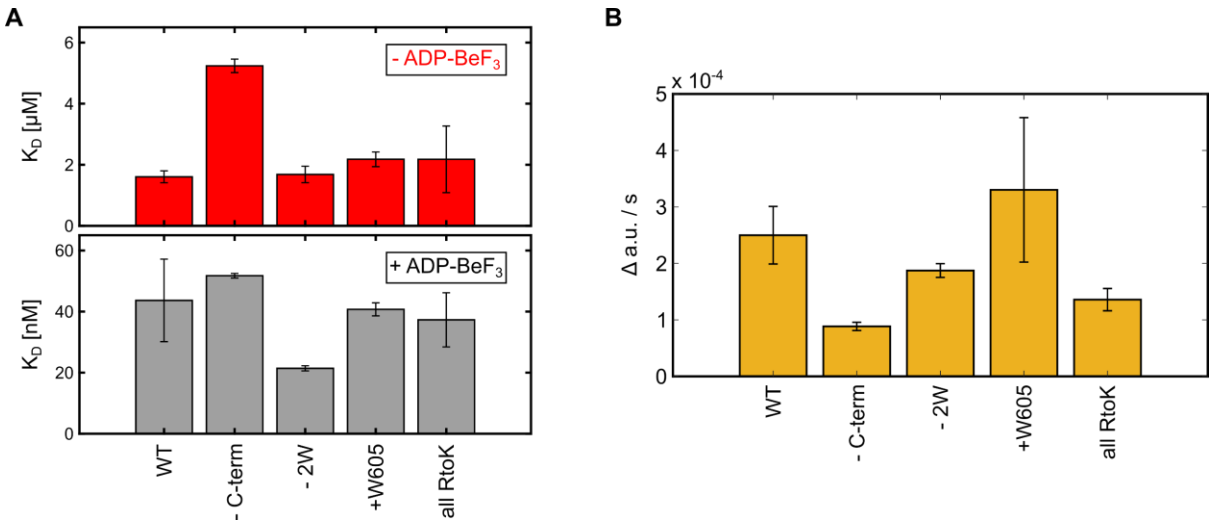


Figure 4.6: Droplet formation of Ded1p full-length variants is not linked to its helicase activity and RNA binding. (**A**) In addition to this, fluorescence anisotropy binding experiments show that the C-terminal IDR is required for RNA binding in the absence of nucleotides. ADP-BeF_3 strongly enhances RNA binding affinity. (**B**) Helicase unwinding activity of Ded1p variants was followed by the unwinding of a fluorescein-labeled 15mer hybrid RNA over time.

4.4 Discussion

Protein condensation is now recognized as a common way to organize the complex cellular environment. Understanding the mechanisms that result in the spontaneous clustering of proteins into cellular macroscopic foci is thus an important aspect of biology. It is well established that multivalent and weak interactions between folded protein domains, IDRs and RNA form the basis for the network of intermolecular interactions (240). When the components reach a specific threshold concentration, these interactions result in the formation of microscopically observable protein-RNA condensates.

Here we have elucidated the molecular details that result in the condensation of the RNA helicase Ded1p and shown that the C-terminal IDR is the main contributor to condensation. Interactions between IDRs are often described using a stick-and-spacer model, where interactions between arginine/lysine and phenylalanine/tyrosine residues are commonly observed in natural (109, 130, 202, 250) and artificially disordered proteins (251). Here we have shown that arginine and tryptophan residues form highly efficient sticker-pairs.

The importance of arginine residues for protein condensation has been well established before, e.g. for the proteins FUS, the N-terminal IDR of DDX4, LAF-1, and nuclear speckles (108, 130, 131, 252), as well as for model peptides (253). The central role of arginine for protein condensation in biology can be explained by two features. First, these residues are involved in protein-RNA interactions and thus are especially potent in promoting cluster formation of RNPs granules (91, 240, 254, 255), second, arginine is capable of forming electrostatic, cation-π, as well as π-π interactions (70, 100), where the π-π interactions between two arginine residues can be stronger than the electrostatic repulsion (133). Lysine residues, on the other hand, are unable to form favorable stacking interactions. The loss of protein condensation upon arginine to lysine mutations thus underscores the importance of intermolecular π-related interactions (108, 131). In that light, the tryptophan arginine interactions that we observe here are of a strong π-π nature, as the replacement of the arginine residues with lysine residues impairs condensation. In agreement with that, the addition of 1,6-hexanediol to the C-terminal IDR of Ded1p (**Supplemental Figure 4.6**) also interferes with protein condensation (256).

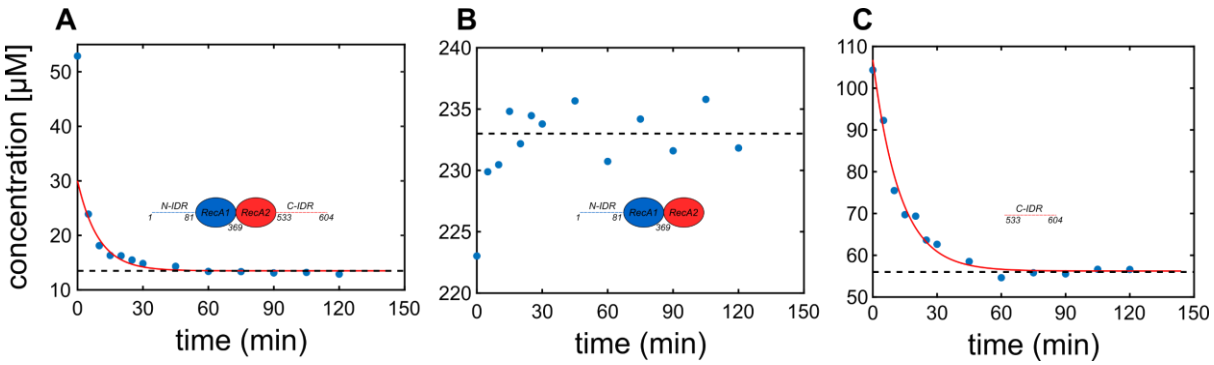
So far, tryptophan residues have not been extensively described as general stickers that shape protein condensation. The few studies that address tryptophan residues in the context of protein condensation include the C-terminal IDR of TDP-43 (125, 126). In that case, protein condensation is reduced, but not abolished, upon removal of the three tryptophan residues. Mechanistically, condensation of TDP-43 cannot be considered as a simple sticker-and-spacer model as one of the central tryptophan residues is located in a region with a strong helical propensity and self-association occurs through the formation of interacting secondary structure elements (125, 126). In addition, tryptophan residues have been shown to aid in the extracellular condensation of gelactin-3 (257). It was shown for only a limited number of artificial systems that tryptophan can make contacts with arginine residues to facilitate condensation (128, 129).

Structural information on the atomic details behind the arginine-tryptophan interactions can be deduced from known folded proteins where these interactions increase protein stability (258). A search in the PDB using the IMAAAGINE webserver (259) for structures where one arginine residue is within 3 Å (hydrogen-bond distance) from two tryptophan residues resulted in 81 hits. In most cases, these tryptophan-arginine

interactions display an orientation where the guanidino group of the arginine side chain is stacked between the indole rings of two tryptophan residues such that planar π-π interactions are formed (**Supplemental Figure 4.7**). These energetically favorable interactions could be exploited in an intermolecular manner to promote condensation of IDRs that contain multiple arginine and tryptophan residues. Interestingly, the 5 tryptophan :3 arginine ratio that appears to be optimal for the condensation of the Ded1p C-terminal IDR indeed suggests the sandwiching of one arginine residue by two tryptophan residues.

The reason that potent interactions between arginine and tryptophan residues have not been generally recognized in the light of protein condensation can be due to the low abundance of tryptophan residues in IDRs (260). Thus, the presence of tryptophan and arginine residues in IDRs may be an indication for protein condensation. Interestingly, the presence of at least two tryptophan residues in the far C-terminal part of Ded1p is a conserved feature (159). In that light, our findings aid the development of computational methods that aim at predicting the condensation potential of IDRs. These methods go hand-in-hand with the *in vitro* reconstitution approaches of cellular foci from purified components (52, 240, 261, 262) such that the cellular mechanisms that underlay the formation of dynamic cellular condensate can be unraveled.

4.5 Supplement



Supplemental Figure 4.1: Condensation of Ded1p is mainly influenced by the C-terminal IDR. The degree of phase separation is linked to the critical c_{SAT} that is required for a protein to undergo condensation. Here, the saturation c_{SAT} was determined by removing the dense phase through rapid centrifugation. Over time, the decay in concentration of Ded1p in the light phase is followed. The concentration of the solution approaches a plateau value during the removal of the dense protein phase, which takes 30 to 60 minutes (see Methods). (**A**) The Ded1p full-length protein de-mixes into a protein dense phase above a concentration of 14 µM. The copy number of Ded1p in the cell is around 32,000 (263); based on a cellular volume of 42 μm^3, this translates into a cellular concentration of 1.3 µM. The effective concentration of Ded1p is higher as Ded1p localizes solely to the cytoplasm. *In vitro* protein condensation of the isolated Ded1p protein thus takes place at a concentration that is around 10-fold higher than the cellular protein concentration. (**B**) The removal of the C-terminal IDR of Ded1p impairs self-aggregation (c_{SAT} > 230 µM), highlighting the importance of this region for condensation. (**C**) The isolated disordered C-terminal IDR alone undergoes condensation above a concentration of 55 µM.

WT
533-MMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSKSSGWGNSGGSNNSSWW-604

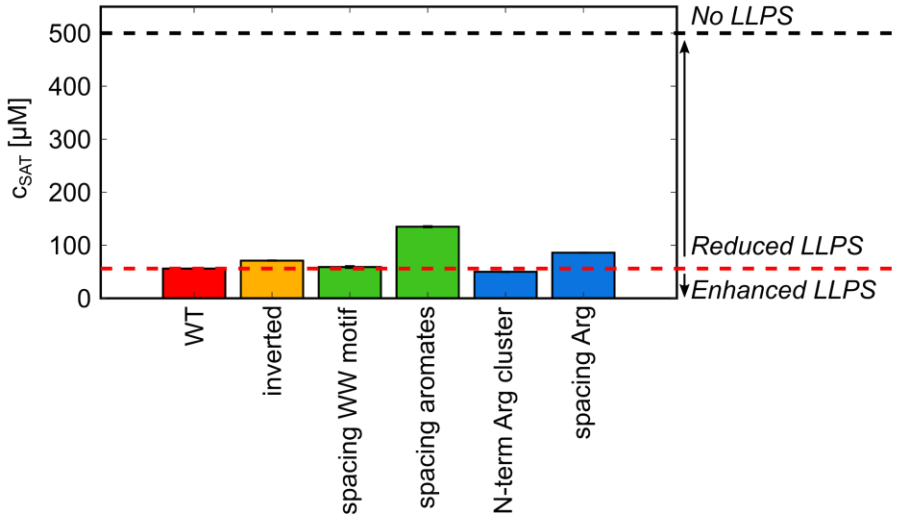
inverted
533-WWSSNNSGGSNGWGSSKSDSGWGSGGRFSNDRSRSSGWGGASAGGAKRYDRNNNRGFGGRRSNSRSGPASMM-604

spacing WW motif
533-MMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSKSSGWGNSGGSNNSSWGGW-606

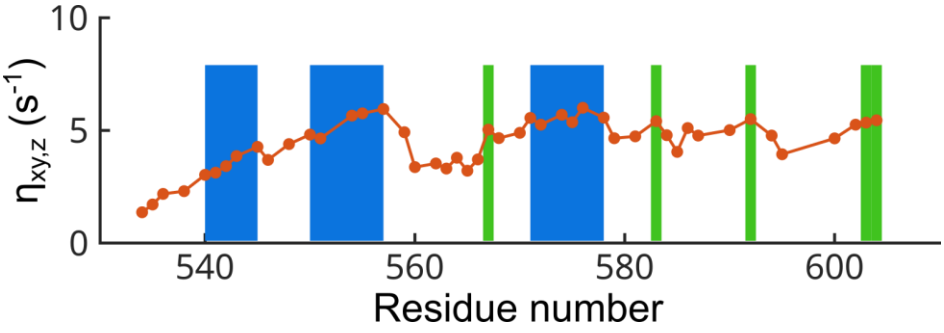
spacing aromates
533-MMSAPGSRFSNSRRGGGYRNNNRDRKWAGGASAGGFGSSRSRDNWSRGGSGGSWDSKSSGGNWSGGSNNSSW-604

N-term Arg cluster
533-MMSAPGSRSRNSRRGGRFGRNRNNRRDYKAGGASAGGWGSSSDNSFGGSGWGSDSKSSGWGNSGGSNNSSWW-604

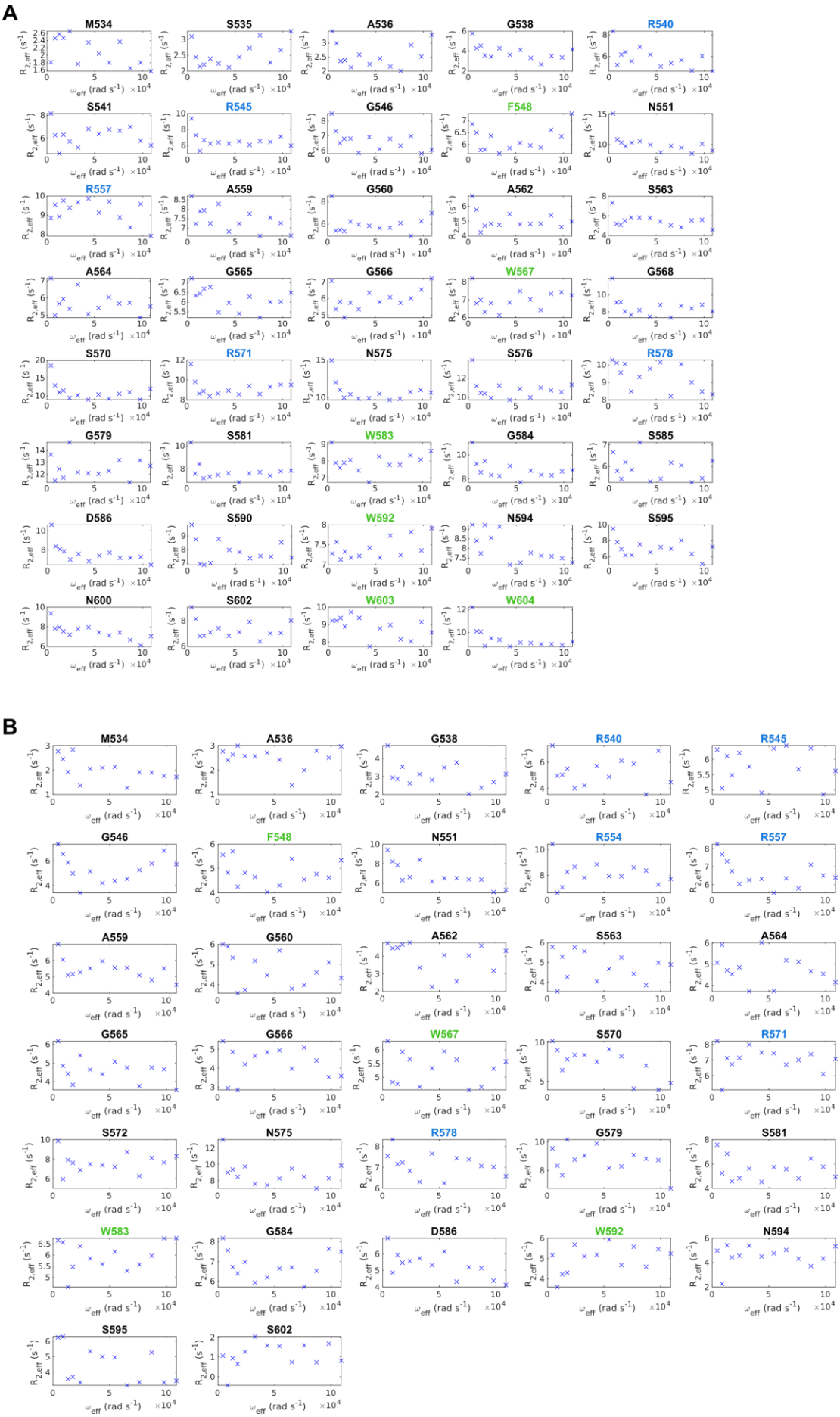
spacing Arg
533-MMSAPRGSSNSGGRFGNNNDYRKAGGASARGGWGSSSRDNSFGGSRGWGSDSKRSSGWGNSRGGSNNSSRWW-604



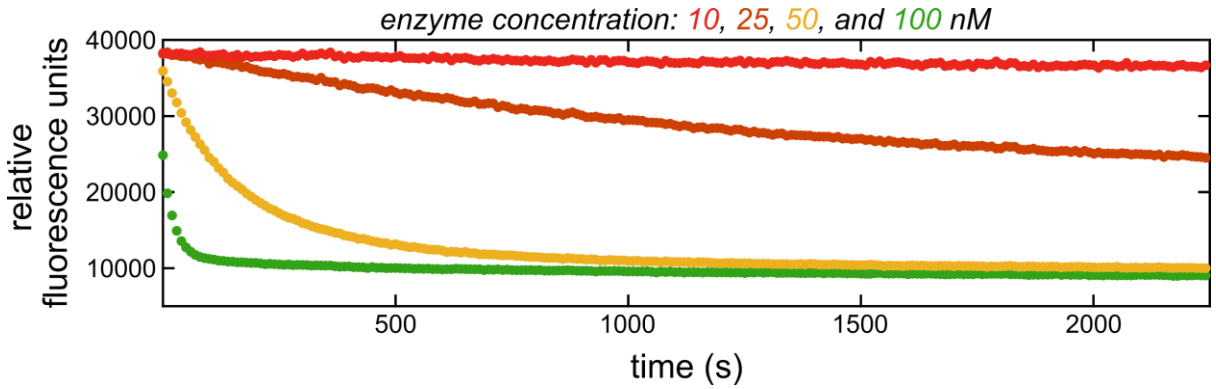
Supplemental Figure 4.2: The order of stickers has no strong influence on the condensation of the Ded1p C-terminal IDR. c_{SAT} was determined for a protein variant with a fully inverted sequence (yellow), for variants in which the tryptophan and all aromatic residues are spaced (green) and for variants in which the positive charges were either clustered or equally spaced (blue). Only the variant with equally spaced aromatic residues shows a slightly reduced propensity for condensation.



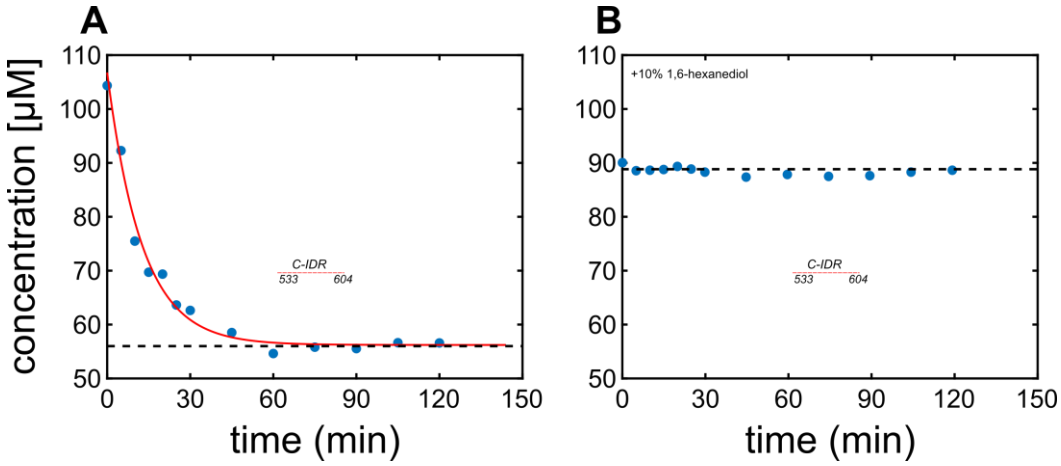
Supplemental Figure 4.3: ^{1}H-^{15}N CSA/DD cross-correlated transverse relaxation rates of the WT C-terminal IDR. The data were recorded at 4°C and 800 MHz. A similar trend as for the ^{15}N R_2 rates is observed with increased rates for tryptophan and arginine residues.


A
M534
S535
A536
G538
R540
S541
R545
G546
F548
N551
R557
A559
G560
A562
S563
A564
G565
G566
W567
G568
S570
R571
N575
S576
R578
G579
S581
W583
G584
S585
D586
S590
W592
N594
S595
N600
S602
W603
W604
B
M534
A536
G538
R540
R545
G546
F548
N551
R554
R557
A559
G560
A562
S563
A564
G565
G566
W567
S570
R571
S572
N575
R578
G579
S581
W583
G584
D586
W592
N594
S595
S602
$R_{2,eff}$ (s^{-1})
ω_{eff} (rad s^{-1}) $\times 10^4$

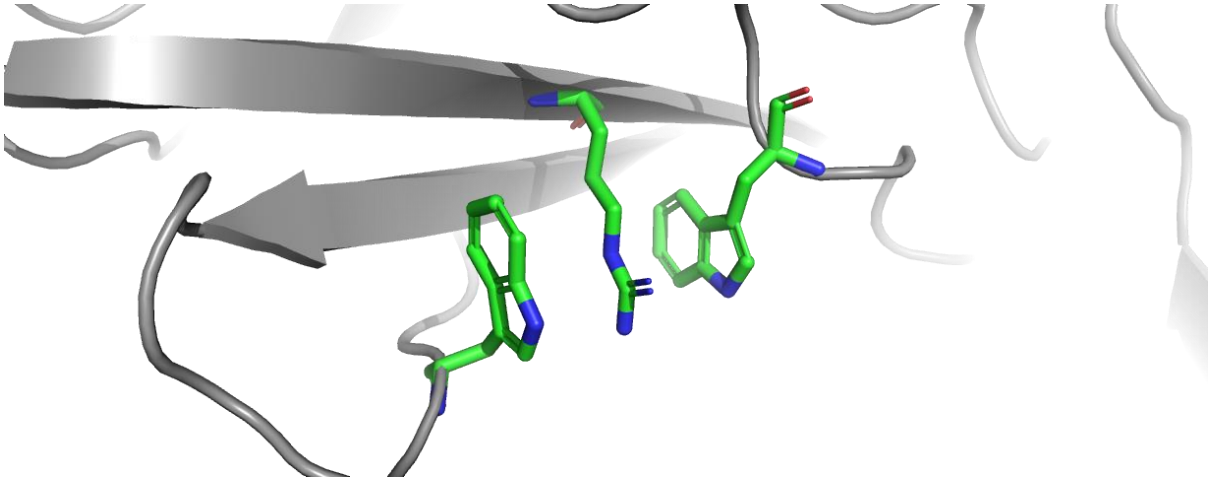
Supplemental Figure 4.4: $^{1}H^{N}$-$R_{1,\rho}$ relaxation dispersion data for the quantifiable resonances of the C-terminal IDR of Ded1p. (**A**) Relaxation dispersion data of the WT C-terminal IDR. (**B**) Relaxation dispersion of the C-terminal IDR lacking the C-terminal WW motif. All data were recorded at 4°C and 800 MHz. No relaxation dispersion was observed for the WT C-terminal IDR or the variant lacking the C-terminal WW motif. Overall, $R_{2,eff}$ rates of the WW mutant are reduced compared to the WT C-terminal IDR.



Supplemental Figure 4.5: Determination of the WT Ded1p helicase activity. The dissociation of dsRNA is monitored by the change in fluorescence intensity over time. The rate of dissociation is strongly dependent on the concentration of Ded1p.



Supplemental Figure 4.6: 1,6-Hexanediol dissolves condensates of the C-terminal IDR of Ded1p. (**A**) The isolated disordered C-terminal IDR alone undergoes condensation above a concentration of 55 µM. (**B**) The addition of 10% (w/v) impairs the self-aggregation (c_{SAT} > 88 µM) of the C-terminal IDR of Ded1p.



Supplemental Figure 4.7: Planar tryptophan-arginine-tryptophan interactions. Example of the structural details of how an arginine residue is stacked via π-π interactions between two tryptophan residues (PDB ID: 1BQU; cytokine-binding region of gp130 (264)).

Supplemental Table 4.1: Sequences of the proteins used in this study. All proteins start with an N-terminal GG that results from TEV cleavage.

Protein	Sequence	Internal Ref.
Ded1p WT (1-604)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFR GGSGWGSDSKSSGWGNSGGSNNSSWW	#1791
Ded1p deltaC (1-535)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPAD IHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLIFV ETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARGLD IPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTEAN QEVPSFLKDAMM	#2228
Ded1p W592G (1-604)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGRGGGSFF SNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPNF QSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVKYSVP IVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVIMA PTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLNDLL ERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPADI QHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLIFV ETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARGLD IPNVTHVINYDLPSDVDDYVHRIG RTGRAGNTGLATAFFNSENSNIVKGLHEILTEANQEVPSFLKDAMMSAPGSRSNSRRG GFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSKSSGGGNSGGSNNSSWW	#2389
Ded1p W604* (1-603) (-W)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFR GGSGWGSDSKSSGWGNSGGSNNSSW	#2395

Ded1p W603* (1-602) (-2W)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFR GGSGWGSDSKSSGWGNSGGSNNSS	#2394
Ded1p +W605 (1-605) (+W)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFR GGSGWGSDSKSSGWGNSGGSNNSSWWW	#2654
Ded1p R540K, R544K, R545K, R550K, R554K R557K, R571K, R573K, R578K (1-604) (RtoK)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSKSNSKKGGFGKNNNKDYKKAGGASAGGWGSSKSKDNSFK GGSGWGSDSKSSGWGNSGGSNNSSWW	#2643
Ded1p R540G, R544G, R545G, R550G, R554G, R557G, R571G, R573G, R578G (1-604) (RtoG)	GGMAELSEQVQNLSINDNNENGYVPPHLRGKPRSARNNSSNYNNNNGGYNGGRGGGSF FSNNRRGGYGNGGFFGGNNGGSRSNGRSGGRWIDGKHVPAPRNEKAEIAIFGVPEDPN FQSSGINFDNYDDIPVDASGKDVPEPITEFTSPPLDGLLLENIKLARFTKPTPVQKYS VPIVANGRDLMACAQTGSGKTGGFLFPVLSESFKTGPSPQPESQGSFYQRKAYPTAVI MAPTRELATQIFDEAKKFTYRSWVKACVVYGGSPIGNQLREIERGCDLLVATPGRLND LLERGKISLANVKYLVLDEADRMLDMGFEPQIRHIVEDCDMTPVGERQTLMFSATFPA DIQHLARDFLSDYIFLSVGRVGSTSENITQKVLYVENQDKKSALLDLLSASTDGLTLI FVETKRMADQLTDFLIMQNFRATAIHGDRTQSERERALAAFRSGAATLLVATAVAARG LDIPNVTHVINYDLPSDVDDYVHRIGRTGRAGNTGLATAFFNSENSNIVKGLHEILTE ANQEVPSFLKDAMMSAPGSGSNSGGGGFGGNNNRDYGKAGGASAGGWGSSGSGDNSFG GGSGWGSDSKSSGWGNSGGSNNSSWW	#2653
Ded1p C-IDR WT (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2000

Ded1p C-IDR F548G (533-604)	GGMMSAPGSRSNSRRGGGGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2397
Ded1p C-IDR Y556G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDGRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2398
Ded1p C-IDR W567G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGGGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2347
Ded1p C-IDR F577G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2348
Ded1p C-IDR W583G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGGGSDSK SSGWGNSGGSNNSSWW	#2396
Ded1p C-IDR W592G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGGGNSGGSNNSSWW	#2349
Ded1p C-IDR W603G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSGW	#2393
Ded1p C-IDR W603G W604G (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSGG	#2392
Ded1p C-IDR W603* (533-602) (-1W)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSW	#2324
Ded1p C-IDR W604* (533-603) (-2W)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSS	#2330
Ded1p C-IDR G532W (533-604)	GWMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2432
Ded1p C-IDR +W605 (533-605)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSGRGGSGWGSDSK SSGWGNSGGSNNSSWWW	#2364
Ded1p C-IDR W567Y, W583Y, W592Y, W603Y, W604Y (533-604) (all WtoY)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGYGSSRSRDNSFRGGSGYGSDSK SSGYGNSGGSNNSSYY	#2532
Ded1p C-IDR W567F, W583F, W592F, W603F, W604F (533-604) (all WtoF)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGFGSSRSRDNSFRGGSGFGSDSK SSGFGNSGGSNNSSFF	#2536
Ded1p C-IDR F548W, Y556W, F577W (533-604) (all FYtoW)	GGMMSAPGSRSNSRRGGWGRNNNRDWRKAGGASAGGWGSSRSRDNSWRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2533
Ded1p C-IDR R540K, R544K, R545K, R550K, R554K, R557K, R571K, R573K, R578K (533-604) (all RtoK)	GGMMSAPGSKSNSKKGGFGKNNNKDYKKAGGASAGGWGSSKSKDNSFKGGSGWGSDSK SSGWGNSGGSNNSSWW	#2410

Ded1p C-IDR R578G (533-604) (-1R)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2618
Ded1p C-IDR R554G R578G (533-604) (-2R)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2617
Ded1p C-IDR R550G, R554G, R557G, R578G (533-604) (-4R)	GGMMSAPGSRSNSRRGGFGGNNNGDYGKAGGASAGGWGSSRSRDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2499
Ded1p C-IDR R550G, R554G, R557G, R573G, R578G (533-604) (-5R)	GGMMSAPGSRSNSRRGGFGGNNNGDYGKAGGASAGGWGSSRSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2486
Ded1p C-IDR R540G, R550G, R554G, R557G, R573G, R578G (533-604) (-6R)	GGMMSAPGSGSNSRGGGFGGNNNGDYGKAGGASAGGWGSSRSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2552
Ded1p C-IDR R540G, R544G, R550G, R554G, R557G, R573G, R578G (533-604) (-7R)	GGMMSAPGSGSNSRGGGFGGNNNGDYGKAGGASAGGWGSSRSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2535
Ded1p C-IDR R540G, R545G, R550G, R554G, R557G, R571G, R573G, R578G (533-604) (-8R)	GGMMSAPGSGSNSRGGGFGGNNNGDYGKAGGASAGGWGSSGSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2534
Ded1p C-IDR R540G, R544G, R545G, R550G, R554G, R557G, R571G, R573G, R578G (533-604) (-9R)	GGMMSAPGSGSNSGGGGFGGNNNGDYGKAGGASAGGWGSSGSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2542
Ded1p C-IDR +R589 (533-605) (+ 1R)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SRSGWGNSGGSNNSSWW	#2604
Ded1p C-IDR +R589 +R596 (533-606) (+2R)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SRSGWGNSGRGSNNSSWW	#2605
Ded1p C-IDR inverted sequence (533-604)	GGWWSSNNSGGSNGWGSSKSDSGWGSGGRFSNDRSRSSGWGGASAGGAKRYDRNNNRG FGGRRSNSRSGPASMM	#2411
Ded1p C-IDR W603 G604 G605 W606 (533-606) (spacing WW motif)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWGGW	#2413
Ded1p C-IDR spacing of aromatic residues (533-604)	GGMMSAPGSRFSNSRRGGGYRNNNRDRKWAGGASAGGFGSSRSRDNWSRGGSGGSWDS KSSGGNWSGGSNNSSW	#2436
Ded1p C-IDR N-terminal Arg cluster (533-604)	GGMMSAPGSRSRNSRRGGRFGRNRNNRRDYKAGGASAGGWGSSSDNSFGGSGWGSDSK SSGWGNSGGSNNSSWW	#2488
Ded1p C-IDR spacing of Arg residues (533-604)	GGMMSAPRGSSNSGGRFGNNNDYRKAGGASARGGWGSSSRDNSFGGSRGWGSDSKRSS GWGNSRGGSNNSSRWW	#2487

Ded1p C-IDR G532C (533-604)	GSMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2366
Ded1p C-IDR N552C (533-604)	GGMMSAPGSRSNSRRGGFGRNCNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2344
Ded1p C-IDR S581C (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGCGWGSDSK SSGWGNSGGSNNSSWW	#2346
Ded1p C-IDR S598C (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGCNNSSWW	#2323
Ded1p C-IDR S602C (533-604)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSCWW	#2377
Ded1p C-IDR R540G, R544G, R545G (533-604) (Δpatch 1)	GGMMSAPGSGSNSGGGGFGRNNNRDYRKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2437
Ded1p C-IDR R550G, R554G, R557G (533-604) (-3R) / (Δpatch 2)	GGMMSAPGSRSNSRRGGFGGNNNGDYGKAGGASAGGWGSSRSRDNSFRGGSGWGSDSK SSGWGNSGGSNNSSWW	#2438
Ded1p C-IDR R571G, R573G, R578G (533-604) (Δpatch 3)	GGMMSAPGSRSNSRRGGFGRNNNRDYRKAGGASAGGWGSSGSGDNSFGGGSGWGSDSK SSGWGNSGGSNNSSWW	#2439
Ded1p C-IDR R540K, R544K, R545K, R550K, R554K, R557K, R571K, R573K, R578K, F548G, Y556G, W567G, F577G, W583G, W592G (533-604) (only WW)	GGMMSAPGSKSNSKKGGGGKNNNKDGKKAGGASAGGGGSSKSKDNSGKGGSGGGSDSK SSGGGNSGGSNNSSWW	#2624

Conclusion and outlook

In this work, NMR spectroscopy and biochemical assays were combined to study the intermolecular interactions leading to LLPS of the DEAD-box RNA helicase Ded1p, a main component of stress granules.

First, studies of the core domains of five different DEAD-box RNA helicases revealed a coupling between ATP and RNA binding that is mediated by a conformational rearrangement of the RecA core domains. The helicase core domains, which tumble independently in the apo state, transition to a closed conformation in the presence of both substrates. The formation of this closed state is required for helicase unwinding activity in all tested DEAD-box RNA helicases.

Second, temperature-dependent interactions between the N- and C-terminal disordered extensions and the folded helicase core were revealed that contribute to the LLPS of Ded1p. For the N-terminal IDR, a short conserved motif between residues 20 and 40 was shown to mediate the interaction with the core. The interaction of the C-terminal IDR with the core resulted in visible LLPS. However, the interaction sites between the RecA1 domain and the C-terminal IDR remain to be mapped.

Third, the molecular grammar governing the driving force of LLPS was identified for the C-terminal IDR. The data revealed that the intermolecular interactions are of highly dynamic nature. Tryptophan was shown to be a highly potent sticker, a residue that forms non-covalent interactions with other residues. Together with arginine residues, efficient sticker-pairs are formed whose interactions drive condensation of the C-terminal IDR of Ded1p. However, protein condensation did not correlate with Ded1p helicase activity.

In summary, these results highlight the importance of IDRs for the biological function and condensation of Ded1p.

In the future, it would be interesting to examine the degree of LLPS and the enzymatic activity of WT Ded1p and its mutants in the absence and presence of components of the translation initiation complex and RNA. It has been reported that the activity of Ded1p is enhanced by the interaction with eIF4A, whereas the interaction of Ded1p with eIF4E and eIF4G has been reported to inhibit both catalytic activity and

oligomerization (242). On could also assess how nucleotides influence condensation and how these nucleotides affect the stability of RNA-helicase interactions.

To this end, these future experiments could provide a comprehensive picture of how Ded1p helicase activity is controlled by clustering of the enzyme in a condensed phase. To support such findings, computational approaches to visualize the Ded1p LLPS propensity could be used in a cooperation. Finally, all *in vitro* findings will be tested in an *in vivo* setting.

References

1. Gorter E, Grendel F (1925) On biomolecular layers of lipoids on the chromocytes of the blood. *The Journal of Experimental Medicine* 41:439–443.
2. Brown R (1833) *Observations on the Organs and Mode of Fecundation in Orchideae and Asclepiadeae* (Taylor).
3. Mullins C (2005) *The biogenesis of cellular organelles* (Eurekah. com and Kluwer Academic/Plenum Publishers).
4. Stillwell W (2016) Membrane Transport. *An Introduction to Biological Membranes*:423–451.
5. Watson H (2015) Biological membranes. *Essays in biochemistry* 59:43–69.
6. Möller MN *et al. (*2016) Solubility and diffusion of oxygen in phospholipid membranes. *Biochimica et biophysica acta* 1858:2923–2930.
7. Kühlbrandt W (2015) Structure and function of mitochondrial membrane protein complexes. *BMC biology* 13:89.
8. Alberti S, Gladfelter A, Mittag T (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 176:419–434.
9. Courchaine EM, Lu A, Neugebauer KM (2016) Droplet organelles? *The EMBO journal* 35:1603–1612.
10. Hirose T, Ninomiya K, Nakagawa S, Yamazaki T (2023) A guide to membraneless organelles and their various roles in gene regulation. *Nature reviews. Molecular cell biology* 24:288–304.
11. Li P *et al. (*2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483:336–340.
12. Wagner R (1835) Einige bemerkungen und fragen über das keimbläschen (vesicular germinativa). *Müller's Archiv Anat Physiol Wissenschaft Med* 268:373–377.
13. Irastortza-Olaziregi M, Amster-Choder O (2021) RNA localization in prokaryotes: Where, when, how, and why. *Wiley interdisciplinary reviews. RNA* 12:e1615.

14. Sołtys K, Tarczewska A, Bystranowska D, Sozańska N (2022) Getting Closer to Decrypting the Phase Transitions of Bacterial Biomolecules. *Biomolecules* 12.

15. Lamond AI, Spector DL (2003) Nuclear speckles: a model for nuclear organelles. *Nature reviews. Molecular cell biology* 4:605–612.

16. Pollock C, Huang S (2010) The perinucleolar compartment. *Cold Spring Harb Perspect Biol* 2:a000679.

17. Lallemand-Breitenbach V, Thé H de (2010) PML nuclear bodies. *Cold Spring Harb Perspect Biol* 2:a000661.

18. Duronio RJ, Marzluff WF (2017) Coordinating cell cycle-regulated histone gene expression through assembly and function of the Histone Locus Body. *RNA biology* 14:726–738.

19. Boke E, Mitchison TJ (2016) The balbiani body and the concept of physiological amyloids. *Cell Cycle* 16:153–154.

20. Liu J-L, Gall JG (2007) U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies. *Proceedings of the National Academy of Sciences of the United States of America* 104:11655–11659.

21. Voronina E, Seydoux G, Sassone-Corsi P, Nagamori I (2011) RNA Granules in Germ Cells. *Cold Spring Harb Perspect Biol* 3.

22. Zeng M *et al. (*2016) Phase Transition in Postsynaptic Densities Underlies Formation of Synaptic Complexes and Synaptic Plasticity. *Cell* 166:1163-1175.e12.

23. Hirokawa N (2006) mRNA Transport in Dendrites: RNA Granules, Motors, and Tracks. *The Journal of Neuroscience* 26:7139–7142.

24. Banani SF, Lee HO, Hyman AA, Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews. Molecular cell biology* 18:285–298.

25. Su X *et al. (*2016) Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science (New York, N.Y.)* 352:595–599.

26. Kilic S *et al. (*2019) Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments. *The EMBO journal* 38:e101379.

27. Yoshizawa T *et al. (*2018) Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell* 173:693-705.e22.

28. Labokha AA *et al. (*2013) Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *The EMBO journal* 32:204–218.

29. Woodruff JB *et al. (*2017) The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell* 169:1066-1077.e10.

30. Boehning M *et al. (*2018) RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature structural & molecular biology* 25:833–840.

31. Cho W-K *et al. (*2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (New York, N.Y.)* 361:412–415.

32. Levone BR *et al. (*2021) FUS-dependent liquid-liquid phase separation is important for DNA repair initiation. *The Journal of cell biology* 220.

33. Hernandez-Verdun D (2011) Assembly and disassembly of the nucleolus during the cell cycle. *Nucleus (Austin, Tex.)* 2:189–194.

34. Lafontaine DLJ, Riback JA, Bascetin R, Brangwynne CP (2021) The nucleolus as a multiphase liquid condensate. *Nature reviews. Molecular cell biology* 22:165–182.

35. Platani M, Goldberg I, Swedlow JR, Lamond AI (2000) In vivo analysis of Cajal body movement, separation, and joining in live human cells. *The Journal of cell biology* 151:1561–1574.

36. Cioce M, Lamond AI (2005) Cajal bodies: a long history of discovery. *Annual review of cell and developmental biology* 21:105–131.

37. Buchan JR (2014) mRNP granules. Assembly, function, and connections with disease. *RNA biology* 11:1019–1030.

38. Jain S, Parker R (2013) The discovery and analysis of P Bodies. *Advances in experimental medicine and biology* 768:23–43.

39. Anderson P, Kedersha N, Ivanov P (2015) Stress granules, P-bodies and cancer. *Biochimica et biophysica acta* 1849:861–870.

40. Begovich K, Wilhelm JE (2020) An In Vitro Assembly System Identifies Roles for RNA Nucleation and ATP in Yeast Stress Granule Formation. *Molecular cell* 79:991-1007.e4.

41. Decker CJ, Parker R (2012) P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol* 4:a012286.

42. O'Flynn BG, Mittag T (2021) The role of liquid-liquid phase separation in regulating enzyme activity. *Current opinion in cell biology* 69:70–79.

43. Peeples W, Rosen MK (2020) *Phase Separation Can Increase Enzyme Activity by Concentration and Molecular Organization*.

44. Feric M, Brangwynne CP (2013) A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. *Nature cell biology* 15:1253–1259.

45. Fournier M-J, Gareau C, Mazroui R (2010) The chemotherapeutic agent bortezomib induces the formation of stress granules. *Cancer cell international* 10:12.

46. Iserman C *et al. (*2020) Condensation of Ded1p Promotes a Translational Switch from Housekeeping to Stress Protein Production. *Cell* 181:818-831.e19.

47. Guillén-Boixet J *et al. (*2020) RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* 181:346-361.e17.

48. Kedersha N, Anderson P (2002) Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. *Biochemical Society transactions* 30:963–969.

49. Buchan JR, Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. *Molecular cell* 36:932–941.

50. Kedersha NL, Gupta M, Li W, Miller I, Anderson P (1999) RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. *The Journal of cell biology* 147:1431–1442.

51. Ivanov P, Kedersha N, Anderson P (2019) Stress Granules and Processing Bodies in Translational Control. *Cold Spring Harb Perspect Biol* 11.

52. Fromm SA *et al. (*2014) In vitro reconstitution of a cellular phase-transition process that involves the mRNA decapping machinery. *Angewandte Chemie (International ed. in English)* 53:7354–7359.

53. Andrei MA *et al. (*2005) A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *RNA (New York, N.Y.)* 11:717–727.

54. Wurm JP, Holdermann I, Overbeck JH, Mayer PHO, Sprangers R (2017) Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proceedings of the National Academy of Sciences of the United States of America* 114:6034–6039.

55. Kedersha N *et al. (*2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *The Journal of cell biology* 169:871–884.

56. Kearly A, Nelson ADL, Skirycz A, Chodasiewicz M (2022) Composition and function of stress granules and P-bodies in plants. *Seminars in cell & developmental biology*.

57. Brangwynne CP, Mitchison TJ, Hyman AA (2011) Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences of the United States of America* 108:4334–4339.

58. Brangwynne CP *et al. (*2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.)* 324:1729–1732.

59. Feric M *et al. (*2016) Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 165:1686–1697.

60. Bremer A *et al. (*2022) Quantifying Coexistence Concentrations in Multi-Component Phase-Separating Systems Using Analytical HPLC. *Biomolecules* 12.

61. Wheeler JR, Matheny T, Jain S, Abrisch R, Parker R (2016) Distinct stages in stress granule assembly and disassembly. *eLife* 5.

62. Kroschwald S *et al. (*2015) Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *eLife* 4:e06807.

63. Coletta A *et al. (*2010) Low-complexity regions within protein sequences have position-dependent roles. *BMC Systems Biology* 4:43.

64. Sanders DW *et al. (*2020) Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* 181:306-324.e28.

65. Molliex A *et al. (*2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163:123–133.

66. Babu MM (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society transactions* 44:1185–1200.

67. Nott TJ *et al. (*2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* 57:936–947.

68. Dumetz AC, Chockla AM, Kaler EW, Lenhoff AM (2008) Effects of pH on protein-protein interactions and implications for protein phase behavior. *Biochimica et biophysica acta* 1784:600–610.

69. Félix SS, Laurents DV, Oroz J, Cabrita EJ (2023) Fused in sarcoma undergoes cold denaturation: Implications for phase separation. *Protein science : a publication of the Protein Society* 32:e4521.

70. Krainer G *et al. (*2021) Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nature communications* 12:1085.

71. Hankins HM, Baldridge RD, Xu P, Graham TR (2014) Role of flippases, scramblases, and transfer proteins in phosphatidylserine subcellular distribution. *Traffic (Copenhagen, Denmark)* 16:35–47.

72. Lin Y, Protter DSW, Rosen MK, Parker R (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular cell* 60:208–219.

73. Patel A *et al. (*2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162:1066–1077.

74. Damman R *et al. (*2019) Atomic-level insight into mRNA processing bodies by combining solid and solution-state NMR spectroscopy. *Nature communications* 10:4536.

75. Wang B *et al. (*2021) Liquid-liquid phase separation in human health and diseases. *Signal transduction and targeted therapy* 6:290.

76. Babinchak WM, Surewicz WK (2020) Liquid-Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. *Journal of molecular biology* 432:1910–1925.

77. Ambadipudi S, Biernat J, Riedel D, Mandelkow E, Zweckstetter M (2017) Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* 8:275.

78. Shin Y, Brangwynne CP (2017) Liquid phase condensation in cell physiology and disease. *Science (New York, N.Y.)* 357.

79. Huai Y *et al. (*2022) How do RNA binding proteins trigger liquid-liquid phase separation in human health and diseases? *Bioscience trends* 16:389–404.

80. Wheeler RJ *et al. (*2019) *Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease*.

81. Ren J, Zhang Z, Zong Z, Zhang L, Zhou F (2022) Emerging Implications of Phase Separation in Cancer. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)* 9:e2202855.

82. Deviri D, Safran SA (2021) Physical theory of biological noise buffering by multicomponent phase separation. *Proceedings of the National Academy of Sciences of the United States of America* 118.

83. Peran I, Martin EW, Mittag T (2020) Walking Along a Protein Phase Diagram to Determine Coexistence Points by Static Light Scattering. *Methods in molecular biology (Clifton, N.J.)* 2141:715–730.

84. Qian D, Michaels TCT, Knowles TPJ (2022) Analytical Solution to the Flory-Huggins Model. *The journal of physical chemistry letters* 13:7853–7860.

85. Ganser LR, Myong S (2020) Methods to Study Phase-Separated Condensates and the Underlying Molecular Interactions. *Trends in biochemical sciences* 45:1004–1005.

86. Murthy AC, Fawzi NL (2020) The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *The Journal of biological chemistry* 295:2375–2384.

87. Arter WE *et al. (*2022) Biomolecular condensate phase diagrams with a combinatorial microdroplet platform. *Nature communications* 13:7845.

88. Martin EW, Mittag T (2018) Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* 57:2478–2487.

89. Flory PJ (1953) *Principles of polymer chemistry* (Cornell university press).

90. Uversky VN, Kuznetsova IM, Turoverov KK, Zaslavsky B (2015) Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS letters* 589:15–22.

91. Chong PA, Vernon RM, Forman-Kay JD (2018) RGG/RG Motif Regions in RNA Binding and Phase Separation. *Journal of molecular biology* 430:4650–4665.

92. Murthy AC *et al. (*2019) Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nature structural & molecular biology* 26:637–648.

93. Ryan VH *et al. (*2018) Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Molecular cell* 69:465-479.e7.

94. Martin EW *et al. (*2020) Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science (New York, N.Y.)* 367:694–699.

95. Bremer A *et al. (*2022) Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nature chemistry* 14:196–207.

96. Choi J-M, Holehouse AS, Pappu RV (2020) Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual review of biophysics* 49:107–133.

97. Connor JP, Quinn SD, Schaefer C (2022) Sticker-and-spacer model for amyloid beta condensation and fibrillation. *Frontiers in Molecular Neuroscience* 15.

98. Posey AE, Holehouse AS, Pappu RV (2018) Phase Separation of Intrinsically Disordered Proteins. *Methods in enzymology* 611:1–30.

99. Borcherds W, Bremer A, Borgia MB, Mittag T (2021) How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Current opinion in structural biology* 67:41–50.

100. Vernon RM *et al. (*2018) Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* 7.

101. Brangwynne CP, Tompa P, Pappu RV (2015) Polymer physics of intracellular phase transitions. *Nature Phys* 11:899–904.

102. Ho W-L, Huang J-R (2022) The return of the rings: Evolutionary convergence of aromatic residues in the intrinsically disordered regions of RNA-binding proteins for liquid-liquid phase separation. *Protein science : a publication of the Protein Society* 31:e4317.

103. Urry DW, Trapane TL, Prasad KU (1985) Phase-structure transitions of the elastin polypentapeptide-water system within the framework of composition-temperature studies. *Biopolymers* 24:2345–2356.

104. Ribeiro SS, Samanta N, Ebbinghaus S, Marcos JC (2019) The synergic effect of water and biomolecules in intracellular phase separation. *Nat Rev Chem* 3:552–561.

105. Quiroz FG, Chilkoti A (2015) Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature materials* 14:1164–1171.

106. Simon JR, Carroll NJ, Rubinstein M, Chilkoti A, López GP (2017) Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nature chemistry* 9:509–515.

107. Kovach IS (1995) The importance of polysaccharide configurational entropy in determining the osmotic swelling pressure of concentrated proteoglycan solution and the bulk compressive modulus of articular cartilage. *Biophysical chemistry* 53:181–187.

108. Das S, Lin Y-H, Vernon RM, Forman-Kay JD, Chan HS (2020) Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* 117:28795–28805.

109. Wang J *et al. (*2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174:688-699.e16.

110. Dyson HJ (2016) Making Sense of Intrinsically Disordered Proteins. *Biophysical journal* 110:1013–1016.

111. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in biochemical sciences* 33:2–8.

112. Necci M, Piovesan D, Tosatto SCE (2016) Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein science : a publication of the Protein Society* 25:2164–2174.

113. van der Lee R *et al. (*2014) Classification of intrinsically disordered regions and proteins. *Chemical reviews* 114:6589–6631.

114. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America* 93:11504–11509.

115. Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, Dahlquist FW (1997) The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. *Nature structural biology* 4:285–291.

116. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology* 293:321–331.

117. Alderson TR, Kay LE (2021) NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. *Cell* 184:577–595.

118. Malinovska L, Kroschwald S, Alberti S (2013) Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochimica et biophysica acta* 1834:918–931.

119. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular cell biology* 16:18–29.

120. Lin Y, Currie SL, Rosen MK (2017) Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *The Journal of biological chemistry* 292:19110–19120.

121. Charlier C *et al. (*2017) Structure and Dynamics of an Intrinsically Disordered Protein Region That Partially Folds upon Binding by Chemical-Exchange NMR. *Journal of the American Chemical Society* 139:12219–12227.

122. Kosol S, Contreras-Martos S, Cedeño C, Tompa P (2013) Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* 18:10802–10828.

123. Tsoi PS *et al. (*2021) Electrostatic modulation of hnRNPA1 low-complexity domain liquid-liquid phase separation and aggregation. *Protein science : a publication of the Protein Society* 30:1408–1417.

124. Kim TH *et al. (*2021) Interaction hot spots for phase separation revealed by NMR studies of a CAPRIN1 condensed phase. *Proceedings of the National Academy of Sciences of the United States of America* 118.

125. Li H-R, Chiang W-C, Chou P-C, Wang W-J, Huang J-R (2018) TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *The Journal of biological chemistry* 293:6090–6098.

126. Li H-R *et al. (*2018) The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43. *Biochimica et biophysica acta. Proteins and proteomics* 1866:214–223.

127. Conicella AE, Zerze GH, Mittal J, Fawzi NL (2016) ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure (London, England : 1993)* 24:1537–1549.

128. Garcia Garcia C, Patkar SS, Jovic N, Mittal J, Kiick KL (2021) Alteration of Microstructure in Biopolymeric Hydrogels via Compositional Modification of Resilin-Like Polypeptides. *ACS biomaterials science & engineering* 7:4244–4257.

129. Baruch Leshem A *et al. (*2023) Biomolecular condensates formed by designer minimalistic peptides. *Nature communications* 14:421.

130. Qamar S *et al. (*2018) FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* 173:720-734.e15.

131. Schuster BS *et al. (*2020) Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of*

the National Academy of Sciences of the United States of America 117:11421–11431.

132. Gobbi A, Frenking G (1993) Y-Conjugated compounds: the equilibrium geometries and electronic structures of guanidine, guanidinium cation, urea, and 1,1-diaminoethylene. *J. Am. Chem. Soc.* 115:2362–2372.

133. Hong Y *et al. (*2022) Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins. *Nature communications* 13:7326.

134. Gao B, Wyttenbach T, Bowers MT (2009) Protonated arginine and protonated lysine: hydration and its effect on the stability of salt-bridge structures. *The journal of physical chemistry. B* 113:9995–10000.

135. Bah A, Forman-Kay JD (2016) Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *The Journal of biological chemistry* 291:6696–6705.

136. Hofweber M *et al. (*2018) Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* 173:706-719.e13.

137. Banjade S, Rosen MK (2014) Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife* 3.

138. Jayabalan AK *et al. (*2016) NEDDylation promotes stress granule assembly. *Nature communications* 7:12125.

139. Jongjitwimol J, Baldock RA, Morley SJ, Watts FZ (2016) Sumoylation of eIF4A2 affects stress granule formation. *Journal of cell science* 129:2407–2415.

140. Dao TP *et al. (*2018) Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Molecular cell* 69:965-978.e6.

141. McGurk L *et al. (*2018) Poly(ADP-Ribose) Prevents Pathological Phase Separation of TDP-43 by Promoting Liquid Demixing and Stress Granule Localization. *Molecular cell* 71:703-717.e9.

142. Altmeyer M *et al. (*2015) Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nature communications* 6:8088.

143. Schaffert L-N, Carter WG (2020) Do Post-Translational Modifications Influence Protein Aggregation in Neurodegenerative Diseases: A Systematic Review. *Brain Sciences* 10.

144. Jankowsky E (2011) RNA helicases at work: binding and rearranging. *Trends in biochemical sciences* 36:19–29.

145. Woodson SA (2010) Taming free energy landscapes with RNA chaperones. *RNA biology* 7:677–686.

146. Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* 76:23–50.

147. Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic acids research* 30:1427–1464.

148. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic acids research* 17:4713–4730.

149. Jankowsky E, Fairman ME (2007) RNA helicases-one fold for many functions. *Current opinion in structural biology* 17:316–324.

150. Byrd AK, Raney KD (2012) Superfamily 2 helicases. *Frontiers in bioscience (Landmark edition)* 17:2070–2088.

151. Linder P, Fuller-Pace FV (2013) Looking back on the birth of DEAD-box RNA helicases. *Biochimica et biophysica acta* 1829:750–755.

152. Caruthers JM, McKay DB (2002) Helicase structure and mechanism. *Current opinion in structural biology* 12:123–133.

153. Linder P (2006) Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic acids research* 34:4168–4180.

154. Henn A, Bradley MJ, La Cruz EM de (2012) ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annual review of biophysics* 41:247–267.

155. Putnam AA, Jankowsky E (2013) DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochimica et biophysica acta* 1829:884–893.

156. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell* 125:287–300.

157. Fairman-Williams ME, Guenther U-P, Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Current opinion in structural biology* 20:313–324.

158. Andreou AZ, Klostermeier D (2012) Conformational changes of DEAD-box helicases monitored by single molecule fluorescence resonance energy transfer. *Methods in enzymology* 511:75–109.

159. Sharma D, Jankowsky E (2014) The Ded1/DDX3 subfamily of DEAD-box RNA helicases. *Critical reviews in biochemistry and molecular biology* 49:343–360.

160. Gulay S, Gupta N, Lorsch JR, Hinnebusch AG (2020) Distinct interactions of eIF4A and eIF4E with RNA helicase Ded1 stimulate translation in vivo. *eLife* 9.

161. Klostermeier D, Rudolph MG (2009) A novel dimerization motif in the C-terminal domain of the Thermus thermophilus DEAD box helicase Hera confers substantial flexibility. *Nucleic acids research* 37:421–430.

162. Del Campo M, Lambowitz AM (2009) Structure of the Yeast DEAD box protein Mss116p reveals two wedges that crimp RNA. *Molecular cell* 35:598–609.

163. Wurm JP, Glowacz K-A, Sprangers R (2021) Structural basis for the activation of the DEAD-box RNA helicase DbpA by the nascent ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 118.

164. Linder P, Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nature reviews. Molecular cell biology* 12:505–516.

165. Yang Q, Jankowsky E (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nature structural & molecular biology* 13:981–986.

166. Weis K, Hondele M (2022) The Role of DEAD-Box ATPases in Gene Expression and the Regulation of RNA-Protein Condensates. *Annual review of biochemistry* 91:197–219.

167. Sengoku T, Nureki O, Dohmae N, Nakamura A, Yokoyama S (2004) Crystallization and preliminary X-ray analysis of the helicase domains of Vasa

complexed with RNA and an ATP analogue. *Acta crystallographica. Section D, Biological crystallography* 60:320–322.

168. Liu F, Putnam A, Jankowsky E (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proceedings of the National Academy of Sciences of the United States of America* 105:20209–20214.

169. Yang Q, Del Campo M, Lambowitz AM, Jankowsky E (2007) DEAD-box proteins unwind duplexes by local strand separation. *Molecular cell* 28:253–263.

170. Bono F, Ebert J, Lorentzen E, Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126:713–725.

171. Jarmoskaite I, Russell R (2011) DEAD-box proteins as RNA helicases and chaperones. *Wiley interdisciplinary reviews. RNA* 2:135–152.

172. Yang Q, Jankowsky E (2005) ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* 44:13591–13601.

173. Iost I, Dreyfus M, Linder P (1999) Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. *The Journal of biological chemistry* 274:17677–17683.

174. Chuang RY, Weaver PL, Liu Z, Chang TH (1997) Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science (New York, N.Y.)* 275:1468–1471.

175. Hilliker A, Gao Z, Jankowsky E, Parker R (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Molecular cell* 43:962–972.

176. La Cruz J de, Iost I, Kressler D, Linder P (1997) The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 94:5201–5206.

177. Guenther U-P *et al. (*2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* 559:130–134.

178. Lai M-C, Sun HS, Wang S-W, Tarn W-Y (2016) DDX3 functions in antiviral innate immunity through translational control of PACT. *The FEBS journal* 283:88–101.

179. Valiente-Echeverría F, Hermoso MA, Soto-Rifo R (2015) RNA helicase DDX3: at the crossroad of viral replication and antiviral immunity. *Reviews in medical virology* 25:286–299.

180. Zhao L *et al. (*2016) Multifunctional DDX3: dual roles in various cancer development and its related signaling pathways. *American Journal of Cancer Research* 6:387–402.

181. He Y *et al. (*2018) A double-edged function of DDX3, as an oncogene or tumor suppressor, in cancer progression (Review). *Oncology reports* 39:883–892.

182. Tarn W-Y, Chang T-H (2009) The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA biology* 6:17–20.

183. Abdelhaleem M (2004) Do human RNA helicases have a role in cancer? *Biochimica et biophysica acta* 1704:37–46.

184. Beckham C *et al. (*2008) The DEAD-box RNA helicase Ded1p affects and accumulates in Saccharomyces cerevisiae P-bodies. *Molecular biology of the cell* 19:984–993.

185. Hondele M *et al. (*2019) DEAD-box ATPases are global regulators of phase-separated organelles. *Nature* 573:144–148.

186. Floor SN, Condon KJ, Sharma D, Jankowsky E, Doudna JA (2016) Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3. *The Journal of biological chemistry* 291:2412–2421.

187. Jumper J *et al. (*2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.

188. Varadi M *et al. (*2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* 50:D439-D444.

189. Schrödinger LL (2015) *The PyMOL Molecular Graphics System, Version 1.8*.

190. Overbeck JH, Stelzig D, Fuchs A-L, Wurm JP, Sprangers R (2022) Observation of conformational changes that underlie the catalytic cycle of Xrn2. *Nature chemical biology* 18:1152–1160.

191. Keeler J (2012) *Understanding NMR spectroscopy* (John Wiley & Sons, Ltd. Publication, Chichester).

192. Zerbe O, Jurt S (2014) *Applied NMR spectroscopy for chemists and life scientists* (Wiley-VCH, Weinheim).

193. Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *The FEBS journal* 278:687–703.

194. Sattler M, Fesik SW (1996) Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure (London, England : 1993)* 4:1245–1249.

195. Webb GA, ed (2017) *Modern Magnetic Resonance* (Springer International Publishing, Cham).

196. Cvetkovic MA, Sprangers R (2017) in *Modern Magnetic Resonance*, ed Webb GA. (Springer International Publishing, Cham), pp 1–15.

197. Schütz S, Sprangers R (2020) Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in nuclear magnetic resonance spectroscopy* 116:56–84.

198. Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magn. Reson. Chem.* 41:843–852.

199. Monneau YR *et al. (*2017) Automatic methyl assignment in large proteins by the MAGIC algorithm. *Journal of biomolecular NMR* 69:215–227.

200. Hartlmüller C, Spreitzer E, Göbl C, Falsone F, Madl T (2019) NMR characterization of solvent accessibility and transient structure in intrinsically disordered proteins. *Journal of biomolecular NMR* 73:305–317.

201. Burke KA, Janke AM, Rhine CL, Fawzi NL (2015) Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular cell* 60:231–241.

202. Brady JP *et al. (*2017) Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings*

of the National Academy of Sciences of the United States of America 114:E8194-E8203.

203. Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical reviews* 109:4108–4139.

204. Theissen B, Karow AR, Köhler J, Gubaev A, Klostermeier D (2008) Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proceedings of the National Academy of Sciences of the United States of America* 105:548–553.

205. Cao W *et al. (*2011) Mechanism of Mss116 ATPase reveals functional diversity of DEAD-Box proteins. *Journal of molecular biology* 409:399–414.

206. Henn A, Bradley MJ, La Cruz EM de (2012) ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annual review of biophysics* 41:247–267.

207. Lorsch JR, Herschlag D (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* 37:2180–2193.

208. Wurm JP (2023) Structural basis for RNA-duplex unwinding by the DEAD-box helicase DbpA. *RNA (New York, N.Y.)*.

209. Mallam AL, Del Campo M, Gilman B, Sidote DJ, Lambowitz AM (2012) Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490:121–125.

210. Putnam AA, Jankowsky E (2013) AMP sensing by DEAD-box RNA helicases. *Journal of molecular biology* 425:3839–3845.

211. Yount RG, Babcock D, Ballantyne W, Ojala D (1971) Adenylyl imidodiphosphate, an adenosine triphosphate analog containing a P--N--P linkage. *Biochemistry* 10:2484–2489.

212. Jin Y, Richards NG, Waltho JP, Blackburn GM (2017) Metal Fluorides as Analogues for Studies on Phosphoryl Transfer Enzymes. *Angewandte Chemie (International ed. in English)* 56:4110–4128.

213. Lacabanne D *et al. (*2020) ATP Analogues for Structural Investigations: Case Studies of a DnaB Helicase and an ABC Transporter. *Molecules* 25.

214. Andreou AZ, Klostermeier D (2013) The DEAD-box helicase eIF4A: Paradigm or the odd one out? *RNA biology* 10:19–32.

215. Tieg B, Krebber H (2013) Dbp5 - from nuclear export to translation. *Biochimica et biophysica acta* 1829:791–798.

216. Moeller H von, Basquin C, Conti E (2009) The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nature structural & molecular biology* 16:247–254.

217. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of biomolecular NMR* 33:199–211.

218. Delaglio F *et al. (*1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 6:277–293.

219. Kiianitsa K, Solinger JA, Heyer W-D (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Analytical biochemistry* 321:266–271.

220. Tsu CA, Uhlenbeck OC (1998) Kinetic analysis of the RNA-dependent adenosinetriphosphatase activity of DbpA, an Escherichia coli DEAD protein specific for 23S ribosomal RNA. *Biochemistry* 37:16989–16996.

221. Sun Y *et al. (*2014) Single-molecule kinetics of the eukaryotic initiation factor 4AI upon RNA unwinding. *Structure (London, England : 1993)* 22:941–948.

222. Wurm JP (2021) Assignment of the Ile, Leu, Val, Met and Ala methyl group resonances of the DEAD-box RNA helicase DbpA from E. coli. *Biomolecular NMR assignments* 15:121–128.

223. Rudolph MG, Heissmann R, Wittmann JG, Klostermeier D (2006) Crystal structure and nucleotide binding of the Thermus thermophilus RNA helicase Hera N-terminal domain. *Journal of molecular biology* 361:731–743.

224. Samatanga B, Klostermeier D (2014) DEAD-box RNA helicase domains exhibit a continuum between complete functional independence and high thermodynamic

coupling in nucleotide and RNA duplex recognition. *Nucleic acids research* 42:10644–10654.

225. Song H, Ji X (2019) The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. *Nature communications* 10:3085.

226. Wallace EWJ *et al. (*2015) Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* 162:1286–1298.

227. Gibson DG *et al. (*2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6:343–345.

228. Keller R (2004) *The computer aided resonance assignment tutorial*.

229. Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics (Oxford, England)* 31:1325–1327.

230. Kabsch W (2010) XDS. *Acta crystallographica. Section D, Biological crystallography* 66:125–132.

231. Liebschner D *et al. (*2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta crystallographica. Section D, Structural biology* 75:861–877.

232. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66:486–501.

233. Waterhouse A *et al. (*2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46:W296-W303.

234. Klostermeier D (2013) Rearranging RNA structures at 75°C? Toward the molecular mechanism and physiological function of the Thermus thermophilus DEAD-box helicase Hera. *Biopolymers* 99:1137–1146.

235. Nerli S, Paula VS de, McShan AC, Sgourakis NG (2021) Backbone-independent NMR resonance assignments of methyl probes in large proteins. *Nature communications* 12:691.

236. Cherkasov V *et al. (*2013) Coordination of translational control and protein homeostasis during severe heat stress. *Current biology : CB* 23:2452–2462.

237. Yoo H, Bard JAM, Pilipenko EV, Drummond DA (2022) Chaperones directly and efficiently disperse stress-triggered biomolecular condensates. *Molecular cell* 82:741-755.e11.

238. Alberti S, Hyman AA (2021) Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nature reviews. Molecular cell biology* 22:196–213.

239. Boeynaems S *et al. (*2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology* 28:420–435.

240. Schütz S, Nöldeke ER, Sprangers R (2017) A synergistic network of interactions promotes the formation of in vitro processing bodies and protects mRNA against decapping. *Nucleic acids research* 45:6911–6922.

241. Tauber D *et al. (*2020) Modulation of RNA Condensation by the DEAD-Box Protein eIF4A. *Cell* 180:411-426.e16.

242. Putnam AA *et al. (*2015) Division of Labor in an Oligomer of the DEAD-Box RNA Helicase Ded1p. *Molecular cell* 59:541–552.

243. Schwalbe H *et al. (*1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* 36:8977–8991.

244. Abyzov A *et al. (*2016) Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J. Am. Chem. Soc.* 138:6240–6251.

245. Ban D, Iconaru LI, Ramanathan A, Zuo J, Kriwacki RW (2017) A Small Molecule Causes a Population Shift in the Conformational Landscape of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* 139:13692–13700.

246. Pak CW *et al. (*2016) Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Molecular cell* 63:72–85.

247. Khan SN *et al. (*2015) Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophysical journal* 109:988–999.

248. Kroenke CD, Loria JP, Lee LK, Rance M, Palmer AG (1998) Longitudinal and Transverse 1 H− 15 N Dipolar/ 15 N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules. *J. Am. Chem. Soc.* 120:7905–7915.

249. Tjandra N, Szabo A, Bax A (1996) Protein Backbone Dynamics and 15 N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects. *J. Am. Chem. Soc.* 118:6986–6991.

250. Gallego LD *et al. (*2020) Phase separation directs ubiquitination of gene-body nucleosomes. *Nature* 579:592–597.

251. Dzuricky M, Rogers BA, Shahid A, Cremer PS, Chilkoti A (2020) De novo engineering of intracellular condensates using artificial disordered proteins. *Nature chemistry* 12:814–825.

252. Greig JA *et al. (*2020) Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Molecular cell* 77:1237-1250.e4.

253. Tesei G *et al. (*2017) Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proceedings of the National Academy of Sciences of the United States of America* 114:11428–11433.

254. Fisher RS, Elbaum-Garfinkle S (2020) Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature communications* 11:4628.

255. Tsang B *et al. (*2019) Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proceedings of the National Academy of Sciences of the United States of America* 116:4218–4227.

256. Düster R, Kaltheuner IH, Schmitz M, Geyer M (2021) 1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities. *The Journal of biological chemistry* 296:100260.

257. Chiu Y-P *et al. (*2020) Liquid-liquid phase separation and extracellular multivalent interactions in the tale of galectin-3. *Nature communications* 11:1229.

258. Gallivan JP, Dougherty DA (1999) Cation-pi interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America* 96:9459–9464.

259. Nadzirin N, Willett P, Artymiuk PJ, Firdaus-Raih M (2013) IMAAAGINE: a webserver for searching hypothetical 3D amino acid side chain arrangements in the Protein Data Bank. *Nucleic acids research* 41:W432-40.

260. Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochimica et biophysica acta* 1804:1231–1264.

261. Currie SL *et al. (*2023) Quantitative reconstitution of yeast RNA processing bodies. *Proceedings of the National Academy of Sciences of the United States of America* 120:e2214064120.

262. Currie SL, Rosen MK (2022) Using quantitative reconstitution to investigate multicomponent condensates. *RNA (New York, N.Y.)* 28:27–35.

263. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* 11:319–324.

264. Bravo J, Staunton D, Heath JK, Jones EY (1998) Crystal structure of a cytokine-binding region of gp130. *The EMBO journal* 17:1665–1674.

Acknowledgments

I am deeply grateful to all those who have supported and guided me throughout my PhD journey.

First, I would like to express my thanks to **Remco Sprangers** for giving me the opportunity to work in his lab and for his guidance, expertise, and continuous encouragement.

I am especially grateful for the support given by **Philip Wurm,** who offered great help in the lab, with the NMR machines, and for his constant motivation.

Further thanks go to **Jobst Liebau**, **Jan Overbeck**, and **Daniela Lazzaretti** for their helpful thoughts and for proof-reading this thesis.

Next, I want to thank **Nadine Stefan** and **Johanna Stöfl** for their technical assistance and for knowing everything in the lab.

I would also like to thank **Patricia Albrecht** for taking care of everything outside the lab.

Thanks also to **Alexander Schmalix** for sharing the lab, being a great colleague, hilarious conversations, and all our trips to the Mensa and the Cafeteria.

Furthermore, I want to thank **Werner Kremer** and **Simon Grunert** for taking care of the magnets.

I would also like to thank my students **Valentin Thron** and **Jakob Schneider** for their excellent support in the lab.

I want to thank **Prof. Dr. Reinhard Sterner** and **Prof. Dr. Birte Höcker** for their mentoring in the RIGeL program. Also, I want to thank **Carolin Apfel** and **Kinga Ay** for their great support in the RIGeL program and for always being available for questions.

Also, I want to thank all the past and present members of the Sprangers and Wiesner group for always being helpful, the awesome time, and the exciting kicker matches. In random order: **Silke**, **Christina**, **David**, **Martin**, **Antonia**, **Kathi**, **Andrea**, **Anna**, **Carsten**, **Olga**, **Zeliha**, **Hendrik**, **Thomas**.

To all my friends, no matter where you are, thanks a lot that you have always been there and that I can always relay on you.

Finally, I want to thank my beloved family for always being supportive, for your unconditional support, for believing in me, and for never putting pressure on me.