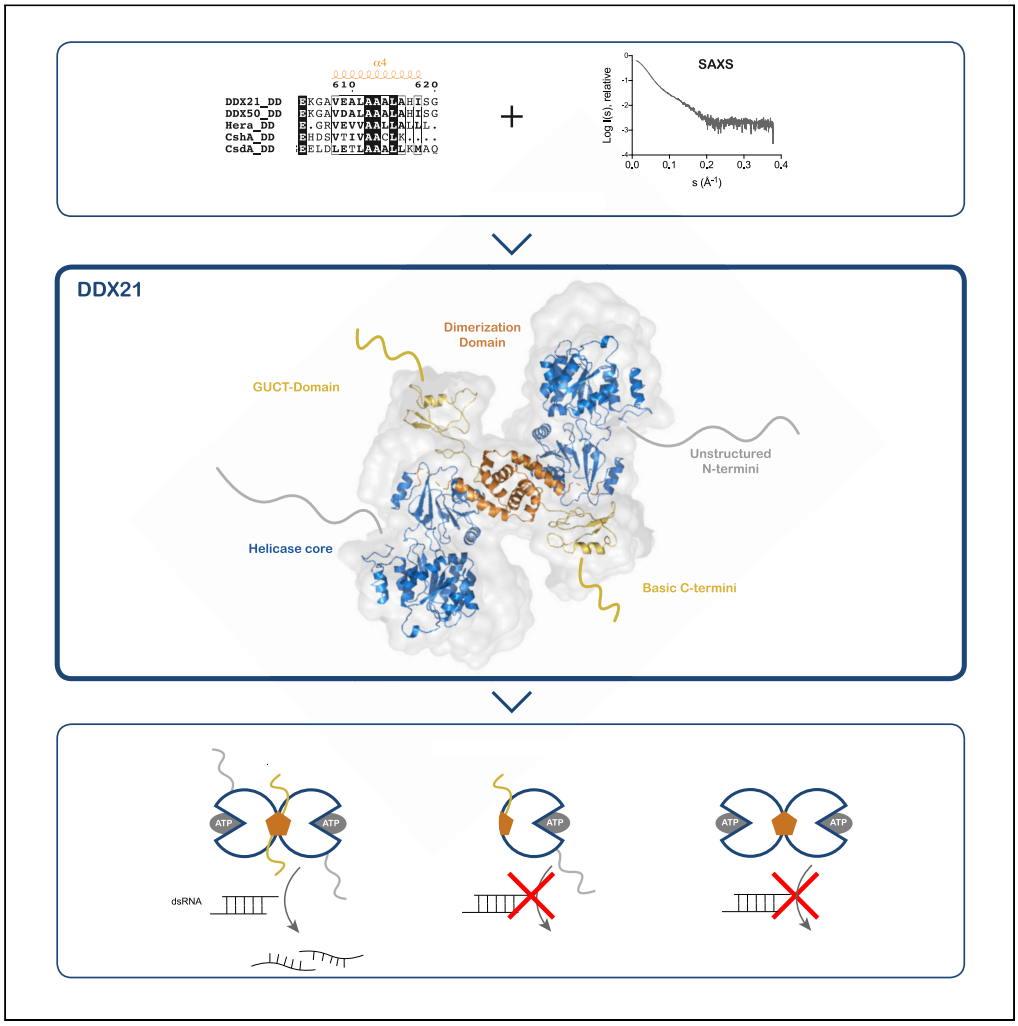
iScience

CellPress
OPEN ACCESS

Article

The Human RNA Helicase DDX21 Presents a Dimerization Interface Necessary for Helicase Activity



Maria J. Marcaida, Annamaria Kauzlaric, Alice Duperrex, ..., Suliana Manley, Didier Trono, Matteo Dal Peraro

maria.marcaidalopez@epfl.ch (M.J.M.)
matteo.dalperaro@epfl.ch (M.D.P.)

HIGHLIGHTS

The human RNA DEAD-box helicase DDX21 is dimeric

DDX21 dimerization is mediated by a hydrophobic central core domain

SAXS-based modeling reveals that DDX21 is intrinsically flexible

Dimerization and C-terminal domains mediate G-quadruplex and dsRNA unwinding

Marcaida et al., iScience *23*, 101811
December 18, 2020
https://doi.org/10.1016/j.isci.2020.101811

iScience

CellPress
OPEN ACCESS

Article

The Human RNA Helicase DDX21 Presents a Dimerization Interface Necessary for Helicase Activity

Maria J. Marcaida,[1,*] Annamaria Kauzlaric,[2] Alice Duperrex,[1] Jenny Sülzle,[3] Martin C. Moncrieffe,[4] Damilola Adebajo,[1] Suliana Manley,[3] Didier Trono,[2] and Matteo Dal Peraro[1,5,*]

SUMMARY

Members of the DEAD-box helicase family are involved in all fundamental processes of RNA metabolism, and as such, their malfunction is associated with various diseases. Currently, whether and how oligomerization impacts their biochemical and biological functions is not well understood. In this work, we show that DDX21, a human DEAD-box helicase with RNA G-quadruplex resolving activity, is dimeric and that its oligomerization state influences its helicase activity. Solution small-angle X-ray scattering (SAXS) analysis uncovers a flexible multidomain protein with a central dimerization domain. While the Arg/Gly rich C termini, rather than dimerization, are key to maintaining high affinity for RNA substrates, *in vitro* helicase assays indicate that an intact dimer is essential for both DDX21 ATP-dependent double-stranded RNA unwinding and ATP-independent G-quadruplex remodeling activities. Our results suggest that oligomerization plays a key role in regulating RNA DEAD-box helicase activity.

INTRODUCTION

RNA DEAD-box helicases perform structural rearrangements of RNA and RNA/protein complexes in many cellular processes including ribosome biogenesis, transcription, translation, and RNA editing (Linder and Jankowsky, 2011). Their names are derived from a characteristic sequence of amino acids (D-E-A-D) that is involved in ATP binding. Members of the DEAD-box family are present in all domains of life and share a common helicase core (HC) composed of two domains homologous to the bacterial single-stranded DNA-binding protein, RecA (Story et al., 2001) (Figure S1). A flexible linker connects the two RecA-like domains (RecA-N and RecA-C). The HC is able to couple the chemical energy from ATP binding and hydrolysis to unwind double-stranded RNA (dsRNA) in a non-processive manner, thanks to several conserved short motifs that mediate either ATP or RNA binding, or intramolecular interactions. The flexible linker allows for large variations in the orientation and distance of the RecA-like domains with respect to each other during the catalytic cycle. Some of these conformations have been observed in crystal structures, depicting the unwinding reaction: the apo, pre-unwound, and post-unwound states (Chen et al., 2020; Henn et al., 2012; Ozgur et al., 2015). In the apo state, the two RecA-like domains are in the open conformation without a defined relative orientation. In the pre-unwound state, dsRNA binding causes structural rearrangements and domain closure (Mallam et al., 2012; Song and Ji, 2019). Further dramatic changes take place to unwind the duplex, resulting in the post-unwound state where one HC is bound to ATP and single-stranded RNA (ssRNA), as seen, for example, for the enzyme Vasa and for DDX21 (Chen et al., 2020; Sengoku et al., 2006) (Figure S1). For some enzymes, strand separation requires only ATP binding, while ATP hydrolysis causes the release of RNA and the recycling of the enzyme (Liu et al., 2008). Despite the high level of structural and sequence conservation, the nucleotide-dependent RNA-binding mode and the modular function of the RecA-like domains are not universal and can vary among DEAD-box proteins (Samatanga and Klostermeier, 2014). Similarly, the oligomerization state of the enzyme on its own or bound to RNA is a factor that varies among family members and could even change during the catalytic cycle. Therefore oligomerization needs to be considered when studying the mechanism of action of DEAD-box helicases, but it has thus far been investigated only for a few members of the family (Huen et al., 2017; Klostermeier, 2013; Minshall and Standart, 2004; Ogilvie et al., 2003; Putnam et al., 2015; Song and Ji, 2019; Xu et al., 2017).

In many cases, DEAD-box helicases also contain additional N- or C- terminal domains that contribute to higher enzymatic efficiency or mediate specific protein/RNA interactions (Del Campo et al., 2009; Grohman

[1]Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015 Switzerland

[2]Global Health Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015 Switzerland

[3]Laboratory for Experimental Biophysics, Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015 Switzerland

[4]Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK

[5]Lead contact

*Correspondence: maria.marcaidalopez@epfl.ch (M.J.M.), matteo.dalperaro@epfl.ch (M.D.P.)

https://doi.org/10.1016/j.isci.2020.101811

et al., 2007; Jarmoskaite and Russell, 2014; Linden et al., 2008; Mallam et al., 2011; Mohr et al., 2008). DEAD box proteins in general display little sequence specificity for RNA, although in some examples, the accessory domains recognize specific RNA structural elements (Kossen et al., 2002; Steimer et al., 2013). The human DEAD-box RNA helicase DDX21, also known as nucleolar RNA helicase II/Guα, falls into this latter class. DDX21 can bind and remodel RNA G-quadruplexes using amino acids at the C terminus of the protein, outside the HC (McRae et al., 2017, 2018). DDX21 is involved in multiple steps of ribosome biogenesis like rRNA processing (Henning et al., 2003) and the recruitment of small nucleolar RNAs into pre-ribosome complexes (Sloan et al., 2015). Furthermore, it has been implicated in transcription by RNA Polymerase I and II, through direct binding of transcription factors like c-Jun (Holmström et al., 2008; Westermarck et al., 2002; Zhang et al., 2014) or through formation of large RNA/protein complexes with long non-coding RNAs such as 7SK or SLERT (Calo et al., 2015; Xing et al., 2017). It is also one of the human DEAD-box helicases implicated in viral infection (Dong et al., 2016; Hammond et al., 2018; Naji et al., 2012; Zhang et al., 2011). For example, DDX21 inhibits influenza A viral replication by inhibiting the viral RNA polymerase (Chen et al., 2014), but it is required for efficient HIV viral production via Rev-related mechanisms (Hammond et al., 2018; Naji et al., 2012). This broad spectrum of functions might reflect the fact that DDX21, in addition to being a *bona fide* helicase able to unwind dsRNA, can also resolve other nucleic acid structures like R-loops or RNA G-quadruplexes (Argaud et al., 2019; Hao et al., 2019; McRae et al., 2017, 2018, 2020; Song et al., 2017). R-loops are cellular structures composed of a ssRNA:dsDNA hybrid that are dynamically formed and resolved at specific loci in association with different transcription phases (Santos-Pereira and Aguilera, 2015). When they form, the exposed ssDNA can be susceptible to DNA damage, and therefore mechanisms that resolve R-loops are important to maintain genomic integrity. Similarly, G-quadruplexes are also non-canonical nucleic acid secondary structures found in sequences rich in guanine residues. Resolving such structures in a timely and regulated manner is key to maintaining genome stability and thus avoiding certain diseases (Agarwala et al., 2015; Santos-Pereira and Aguilera, 2015). Indeed, DDX21 dysregulation has been observed in multiple human cancers (Kim et al., 2019; Zhang et al, 2014, 2018) and was linked to developmental disorders (Calo et al., 2018). Despite growing evidence of its prominent roles, there is still very little known about how the helicase uses its various auxiliary domains to modulate the unwinding and remodeling of the different RNA substrates, thus regulating its function.

This work provides the first structural and biochemical characterization of the accessory domains of DDX21, revealing that it is a homodimer and identifying its dimerization interface. Because of the inherent flexibility of the multi-domain organization of DDX21, we used an integrative approach (Degiacomi et al., 2013; Fonti et al., 2019; Tamò et al., 2015) combining biochemical and biophysical analyses as well as solution small-angle X-ray scattering (SAXS) data and flexible fitting. In light of our results, we propose that an intact dimerization interface is essential for maintaining two DDX21 activities: dsRNA unwinding and G-quadruplex remodeling. We further discuss the implications of oligomerization for the current understanding of the modular function of DEAD-box helicases.

RESULTS

DDX21 Is a Multi-Domain Homodimer

Previous work has shown that in addition to the DEAD-box HC, DDX21 contains a Gu C-terminal (GUCT) domain and a C-terminal basic tail (also described as an RGG/RG motif containing four repeats of the sequence FRGQR/PRGQR, Figure 1A) (Flores-Rozas and Hurwitz, 1993; Valdez, 2000). All domains are necessary for HeLa cell growth and cell cycle progression (Ou et al., 1999). The GUCT domain adopts the fold of an RNA recognition motif (RRM) (Daubner et al., 2013) with a central four-stranded β-sheet and two α-helices (Figure S2) (PDB code 2M3D, unpublished). Comparison of the electrostatic potential distribution on the surface of the GUCT/RRM domains from different enzymes might suggest they have different functions and will be discussed later in the context of RNA binding.

Structure-based sequence alignments and homology modeling template searches have allowed us to identify a putative dimerization domain (DD) (residues 568–620) (Figure 1B). The templates used are three dimeric bacterial proteins (Huen et al., 2017; Klostermeier, 2013; Xu et al., 2017) whose sequence alignment is shown in Figures 1B and S3. On the basis of the structural inferences derived from the three bacterial homodimeric helicases, we hypothesized that the DD in human DDX21 also drives self-association, as is the case for the bacterial enzymes. Indeed, when full-length DDX21 ($DDX21_{FL}$) and several N- and C-terminal truncation constructs were expressed and purified (Figure 1C), they behaved as dimers, as shown by multi-angle light scattering coupled to size exclusion chromatography (SEC-MALS) experiments

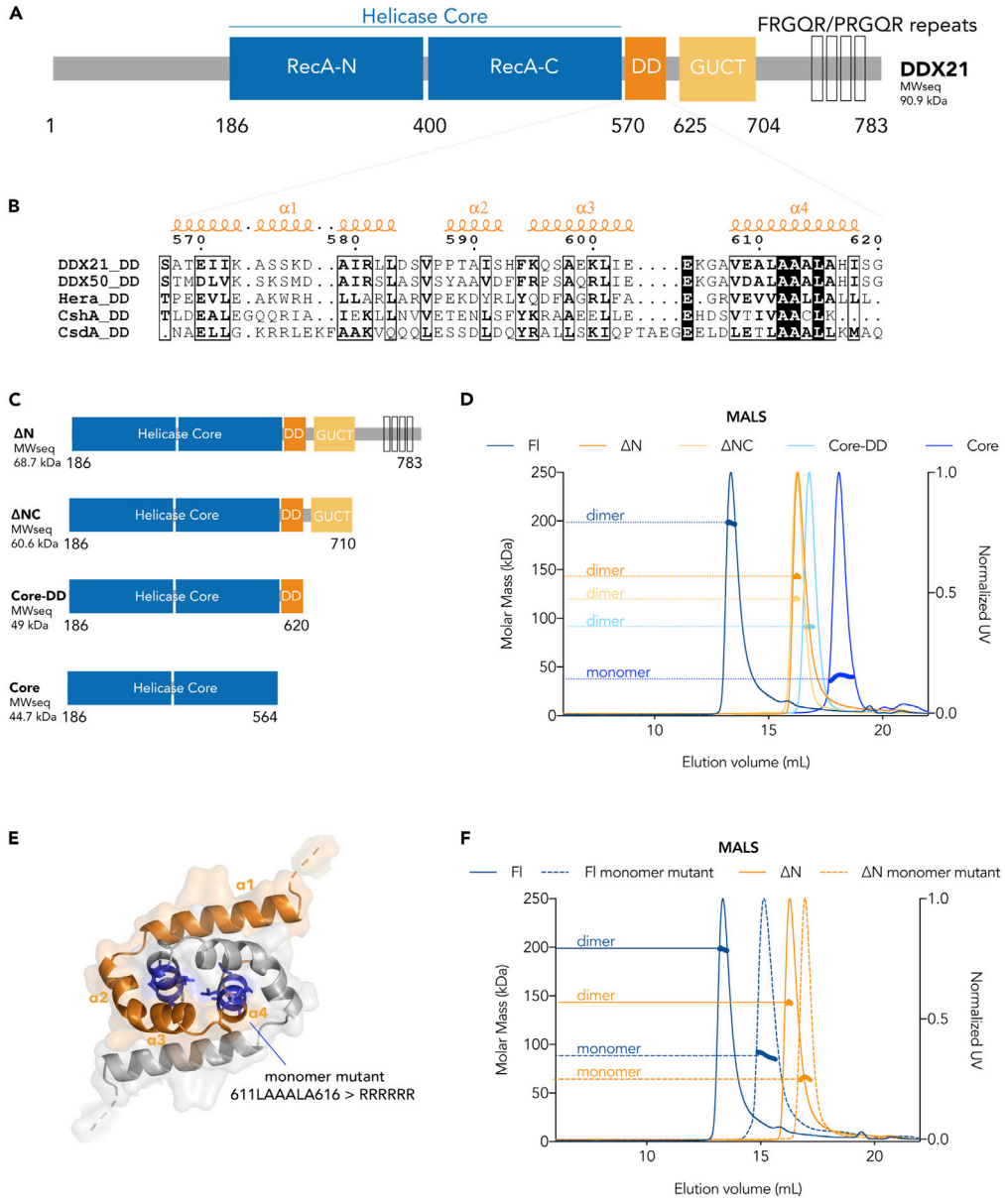


Figure 1. Identification of the Dimerization Domain of DDX21

(A) Numbered linear sequence with the different domains colored in blue (HC), orange (DD), and yellow (GUCT) and the C-terminal tail with the FRGQR/PRGQR repeats in black boxes. The molecular weight of this sequence is MW_{seq} = 90.9 kDa (cleaved affinity tag) (see also Figures S1, S2and S7).

(B) Sequence alignment of the DDs from the two human paralogs DDX21 and DDX50, as well as the bacterial dimeric DEAD-box helicases Hera, CshA, and CsdA (see also Figure S3).

(C) Schematic representation of the truncation constructs used in this study with their corresponding monomeric MW_{seq} calculated from their sequences.

(D) SEC-MALS data show that all the constructs except $DDX21_{Core}$ are dimers (theoretical molar masses [MW_{seq}] for Fl, ΔN, ΔNC, Core-DD, and Core are 90.9, 68.7, 60.6, 49, and 44.7 kDa, respectively). Experimental SEC-MALS molar masses are Fl, 198 kDa; ΔN, 143 kDa; ΔNC, 120 kDa; Core-DD, 91 kDa; and Core, 41 kDa) (see also Figures S4 and S5).

(E) Homology model of the DD of DDX21 in cartoon representation, with helices α1 to α4 of each protomer colored in orange and gray, respectively. The residues in α4 mutated to create the monomeric mutants are shown in blue stick representation.

(F) SEC-MALS of the Fl and ΔN monomeric mutants (in dashed lines) with molar masses of 88 and 66 kDa, compared with intact $DDX21_{Fl}$ (blue) and $DDX21_{\Delta N}$ (orange) with molar masses of 198 and 143 kDa.

(Figure 1D). On the contrary, as expected, the HC by itself is monomeric (Figure 1D). DDX21$_{Fl}$ dimerization was further confirmed by cross-linking and analytical ultracentrifugation equilibrium sedimentation experiments (Figure S4).

In the bacterial structures, the DD consists of four α-helices that each pack compactly with the corresponding helices of the dimeric partner, forming the so-called "knot" that is stabilized by hydrophobic and hydrogen bonding interactions (Huen et al., 2017; Klostermeier and Rudolph, 2009; Xu et al., 2017). Knots are rare but have been reported for the DD of the three bacterial helicase crystal structures, and the pattern of amino acids that support the dimers of the bacterial proteins also occurs in the human DDX21 DD (Figure 1B). Based on these templates, a model for the DD of DDX21 was generated (Figure 1E). The homology model suggests that the hydrophobic core of the DD forms the main hydrophobic interface of the dimer. This is composed of residues of the two α4 helices, specifically residues 608 to 616. To test this hypothesis, DDX21$_{Fl}$ and DDX21$_{\Delta N}$ mutants were generated in which the conserved hydrophobic residues in the α4 helices (residues 611–616 [LAAALA]) were mutated to charged arginines, to create repulsion upon dimerization (Figures 1B and 1E). The two mutants were expressed, purified, and well folded, as were all the other DDX21 constructs (confirmed by circular dichroism, Figure S5). Both DDX21 mutants behaved as monomers in SEC-MALS experiments (Figure 1F), confirming that DDX21 dimerization requires the hydrophobic interface provided by the α4 helices in the DD to be intact.

DDX21 Binds R-Loops as a Dimer

We next investigated the oligomeric state of DDX21 in the presence of RNA. The protein-RNA complexes were unstable under the conditions suitable for SEC-MALS analysis; therefore we used mass photometry to estimate the molecular mass of the macromolecular assemblies in solution (Sonn-Segev et al., 2020). The basis of mass photometry is the optical detection of the scattering signal generated by a single particle at the glass-water interface. This signal scales linearly with the mass of the particle. After calibrating the signal with a sample of known mass, mass photometry accurately provides mass distributions. The histograms in Figures 2A–2C show the measured distribution of molecular mass estimates for individual DDX21 macromolecular complexes and RNA molecules. Because mass photometry is limited to single events, the concentration of macromolecules is limited to the low nanomlolar range. Thus, we used 30 and 15 nM, respectively. DDX21$_{Fl}$ was detected as two peaks of 110 $\pm$ 10 and 216 $\pm$ 13 kDa, whereas the monomer mutant showed only one peak at 86 $\pm$ 15 kDa (Figure 2A). This indicates that at this low concentration, DDX21$_{Fl}$ displays a monomer-dimer equilibrium, whereas the mutant is mainly monomeric. Further mass photometric analysis of the truncation mutants corroborated the molecular masses measured by MALS: both DDX21$_{\Delta N}$ and DDX21$_{Core\text{-}DD}$ display two peaks that are consistent with a monomer-dimer equilibrium, whereas the system without the DD (DDX21$_{Core}$) is mainly monomeric (Figure 2B).

As a nucleic acid substrate we choose the R-loop described by Song and colleagues (Song et al., 2017), because its molecular weight (45 kDa) allows for the mass change upon complex formation to be detectable by this technique. As shown by fluorescence polarization (FP) binding experiments, purified DDX21$_{Fl}$ binds the R-loop substrate with nM affinity (9.2 $\pm$ 0.7 nM, Figure 2D), which is within the affinity range suitable for mass photometric analysis. The R-loop alone was detected as a single peak at 65 $\pm$ 13 kDa. This difference is likely due to the different refractive indices and charges of proteins and nucleic acids and the fact that 45 kDa is at the low mass detection limit of the technique. The DDX21$_{Fl}$-R-loop complex at 30 nM was detected as three peaks: the first peak corresponds to excess R-loop (61 $\pm$ 10 kDa), the second peak to the protein dimer (216 $\pm$ 21 kDa), and the third peak to a protein dimer bound to two R-loop molecules (302 $\pm$ 33 kDa) (Figures 2C and S6).

These experiments suggest that DDX21 is a dimer in the apo conformation as well as in complex with RNA substrates. DDX21 appears therefore structurally homologous to the bacterial enzymes (i.e., Hera, CshA, and CsdA) and dimerizes via a small hydrophobic helical domain. Inspection of its DD amino acid composition suggests that the DDX21 paralog DDX50 might also be a dimer (Figures 1B and S7). Altogether, we identified a previously unknown dimerization domain for DDX21 yet potentially important for helicase activity.

DDX21 Dimerization Is Essential for dsRNA Unwinding

Using the acquired structural information, we sought to understand the influence of dimerization and each accessory domain in RNA binding and unwinding. The binding affinities of the DDX21 variants to ssRNA,

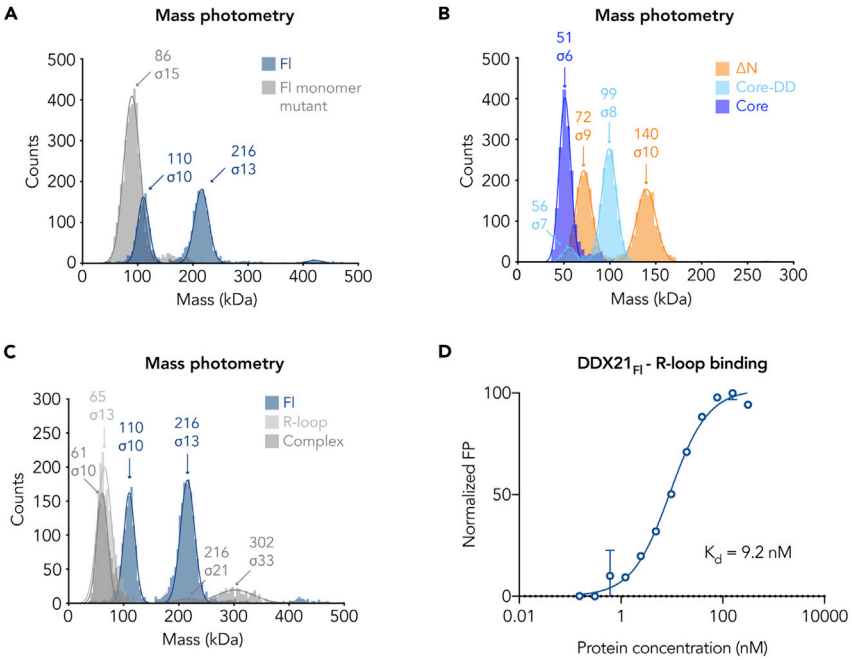


Figure 2. The DDX21 Dimer Binds Two R-loop Molecules

(A) Mass photometric profiles obtained for $DDX21_{Fl}$ (in blue) and $DDX21_{Fl}$ monomer mutant (in gray) with the determined average molecular mass indicated above each peak. The theoretical molar mass for $DDX21_{Fl}$ (with uncleaved affinity tag) is 104.9 kDa and for the $DDX21_{Fl}$ monomer mutant is 90 kDa. The wild-type protein shows a monomer-dimer equilibrium, whereas the mutant is mainly monomeric.

(B) Mass photometric profiles for $DDX21_{\Delta N}$, $DDX21_{Core\text{-}DD}$, and $DDX21_{core}$ indicate that $DDX21_{\Delta N}$ displays a monomer-dimer equilibrium as seen for $DDX21_{Fl}$, whereas $DDX21_{Core\text{-}DD}$ and $DDX21_{core}$ are mainly dimeric and monomeric, respectively.

(C) Mass photometric profile for the $DDX21_{Fl}$-R-loop complex at 30 nM. (D) Fluorescence polarization binding assays measure a binding affinity of $DDX21_{Fl}$ to the R-loop substrate of 9.2 $\pm$ 0.7 nM. Error bars represent the standard deviation of three independent measurements. The mass photometric experiments were performed in duplicates for the protein samples and in triplicates for the R-loop complex (see also Figure S6).

dsRNA, and RNA G-quadruplex were measured using fluorescence polarization experiments (see Transparent Methods, Figures 3A, 3B, and 4A, and Table S1). $DDX21_{Fl}$ binds RNA in the absence of ATP with the highest affinity for the RNA G-quadruplex (Q2 RNA) (2.6 nM Figure 4A, Table S1), whereas the affinity decreases to 30.7 and 10 nM, for ssRNA and dsRNA, respectively (Figures 3A and 3B). All constructs bind RNA with nanomolar affinity, but $DDX21_{Fl}$ exhibits the strongest affinity, indicating that all the domains cooperate in stabilizing the complexes. Removal of the N-terminal 185-amino acid region has a slight effect on the binding to the ssRNA and the dsRNA (~1.5-fold decrease) and a ~3-fold decrease to the RNA G-quadruplex binding. These results indicate that this domain, most likely unstructured (see below), makes only a small contribution to RNA binding. On the other hand, deletion of the C-terminal basic tail significantly reduces the affinity by at least one order of magnitude for all substrates ($DDX21_{\Delta NC}$). The additional deletion of the GUCT domain in the $DDX21_{Core\text{-}DD}$ construct slightly recovers this drop (by 1.5-fold), still remaining far below the $DDX21_{Fl}$ binding affinity values. $DDX21_{Core}$ is the weakest binder, whereas the monomeric DDX21 mutants bind RNA as tightly as the wild-type protein, emphasizing that the C-terminal basic tail is the key for RNA recognition, whereas dimerization seems not to be essential.

We next tested the dsRNA unwinding activity of the DDX21 mutants using an *in vitro* helicase assay (Figures 3C–3G and S8). Unwinding of the dsRNA substrate by DDX21 could not be observed directly by gel electrophoresis, as the single strands reannealed quickly (lane 4 in Figure 3C). Strand separation was observed only when boiling the samples at 95°C (lane 1 in Figure 3C). To observe the unwinding reaction, RNase T1 was used to cleave the resulting single RNA strands at G residues while the unwinding reaction was taking place (lane 5 in Figure 3C). Importantly, RNase T1 cleavage depends on the presence of both DDX21 and ATP (lane 5 versus lanes 4 and 6 in Figure 3C) demonstrating that DDX21 uses ATP hydrolysis for dsRNA

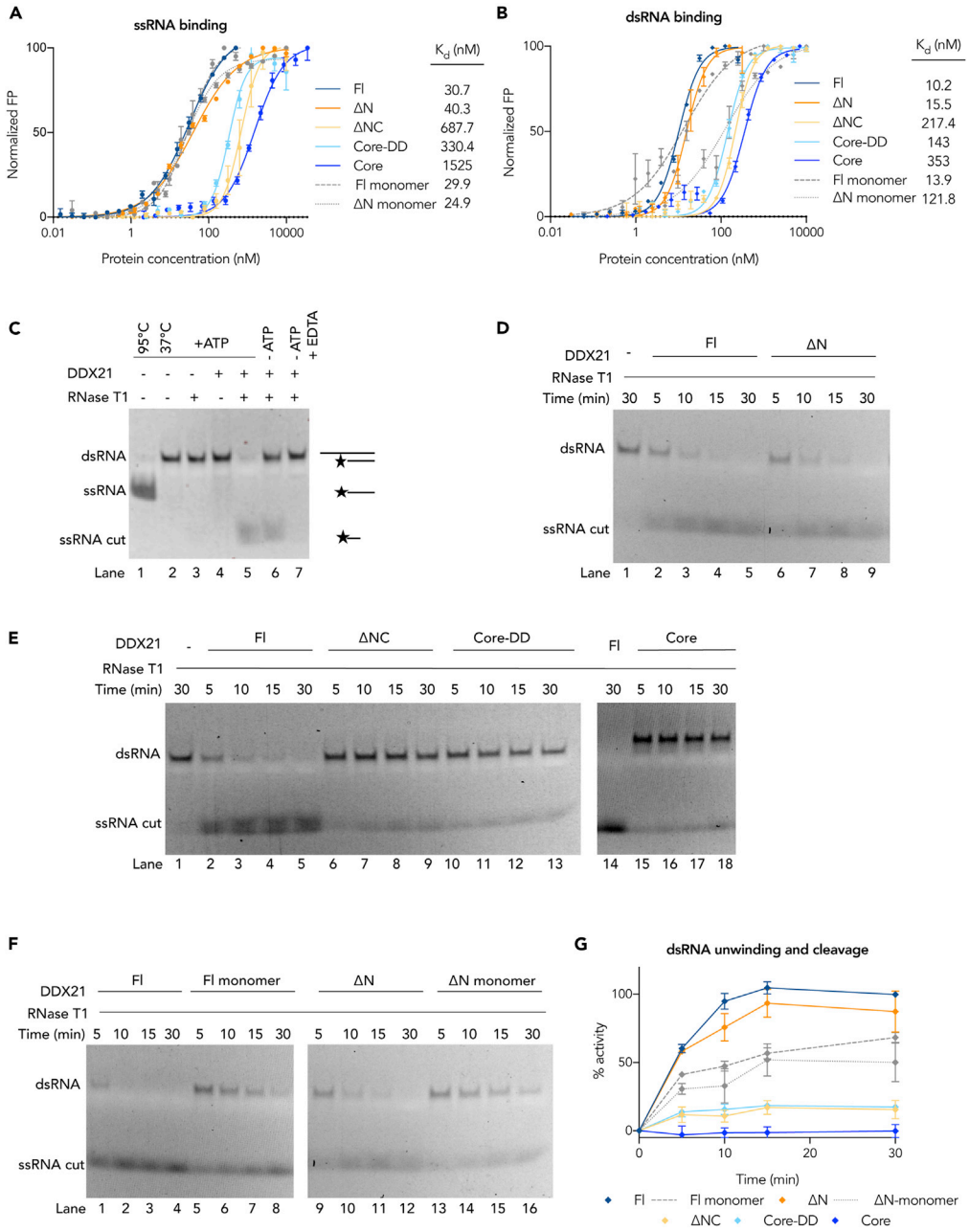


Figure 3. dsRNA Binding and Unwinding by DDX21

(A and B) FP experiments comparing the binding affinity of the DDX21 mutants to the (A) 15-mer ssRNA and the (B) dsRNA with a 3′ 15-base overhang, respectively. Error bars represent the standard deviation of three independent measurements (see also Table S1).

(C) dsRNA helicase assay control experiment (time = 5 min) showing that upon addition of ATP, DDX21 can unwind the dsRNA, such that the exposed ssRNA can be cut by RNase T1 (lane 5). Without DDX21, RNase T1 cannot cut the substrate (lane 3). As described by Hammond et al. (2018) and Valdez et al. (1997) in the absence of ATP (lane 6), DDX21 has some residual helicase activity, which disappears upon addition of 5 mM EDTA (lane 7). The nature of the RNA bands in the gel is depicted on the right, where the star represents the 5′FAM modification.

(D–F) dsRNA helicase assay comparing the activity of the different DDX21 mutants at different time points (5, 10, 15, and 30 min). In native condition polyacrylamide gels, the 5′-FAM-labeled RNA was visualized by measuring the fluorescent signal at 535 nm.

Figure 3. *Continued*
(G) The helicase assays shown in (D–F) were repeated in triplicate (see also Figure S8); the activity was calculated by measuring the intensity of the product (cut) RNA bands, with respect to $DDX21_{Fl}$ activity; and the quantification is shown in the graph, where the points represent the mean of three independent measurements and the error bars represent the standard error.

unwinding. The helicase activity of the DDX21 mutants was compared with $DDX21_{Fl}$ activity using a time course experiment (Figures 3D–3F). $DDX21_{\Delta N}$, without the N-terminal unstructured region, has ~87% $DDX21_{Fl}$ activity, confirming that this region is not necessary for RNA unwinding (Figure 3D). However, as shown in Figure 3E, deletion of the C-terminal basic tail reduces the activity to ~15% of $DDX21_{Fl}$. Additional removal of the GUCT domain does not decrease the activity further. $DDX21_{Core}$ has almost undetectable activity under the same conditions. Interestingly, the monomeric DDX21 mutants, with intact GUCT domains and basic tails, have only ~50% of the dimeric protein activity (Figure 3F), indicating that an intact DD is crucial for maintaining full dsRNA unwinding activity.

DDX21 Requires Dimerization to Remodel RNA G-Quadruplexes

G-quadruplexes are non-canonical nucleic acid secondary structures found in sequences rich in guanine residues (Agarwala et al., 2015). Despite the importance of RNA G-quadruplexes homeostasis in cells, the understanding of the mechanisms underlying G-quadruplex processing by helicases is still poor. As DDX21 has the highest affinity to this RNA structure (Figures 3A, 3B and 4A), we investigated the implications of our structural model in the context of RNA G-quadruplex binding and remodeling. FP RNA-binding experiments show that ionic strength greatly affects the binding affinity of all DDX21 variants to such structures (Figure S9 and Table S2). The flexibility of DDX21 may allow modulation of the accessibility of charged patches on its surface to bind certain protein or RNA partners. The C-terminal basic tail is essential for the recognition of this RNA structure (Figure 4A), as shown by previous studies (McRae et al., 2017, 2018). Further deletion of the GUCT domain strengthens the interaction to a moderate extent. The GUCT domain thus seems to contribute little to RNA binding and remodeling by itself, but may support protein-protein interactions and the formation of an RNA binding unit together with the C-terminal basic tail (see next section). Such a role in protein-protein interactions for the GUCT domain was proposed for DDX50, by comparison of its pI with that of the rest of the protein and by observation of its electrostatic potential distribution (Ohnishi et al., 2009) (Figures S2 and S7). Therefore, although sharing a similar structure as other RRMs of bacterial enzymes, the GUCT domain of DDX21 might not share their RNA-binding modes.

Interestingly, the specificity for the RNA G-quadruplex is already present in the $DDX21_{Core\text{-}DD}$ and $DDX21_{Core}$ constructs, which bind this substrate with an order of magnitude higher affinity than all other substrates (Figures 3A, 3B and 4A). The aforementioned disruption of the dimerization interface does not affect the binding of DDX21 to RNA G-quadruplexes (Figure 4A), so we sought to determine whether oligomerization influences their remodeling.

DDX21 has been shown to remodel a G-rich RNA sequence from one G-quadruplex conformation to another (McRae et al., 2017). Circular dichroism spectroscopy and dynamic light scattering were used to assess the formation of the expected parallel G-quadruplex structure by the 5′Fluorescein amidite (FAM) Q2 RNA (Figure S10), which was maintained even in the presence of $DDX21_{Fl}$ (Figure 4B). To monitor the remodeling activity of DDX21, we used a similar approach as before, using RNase T1 (Figures 4C–4E and S11). In this case, RNase T1 cuts the substrate in the absence of DDX21, with the presence of the helicase accentuating this activity, showing that DDX21 changes the conformation of this RNA structure increasing the accessibility of the G bases. As for dsRNA unwinding, $DDX21_{Fl}$ has the highest activity, whereas only $DDX21_{\Delta N}$ and the monomeric mutants have activity above background. All the other truncation mutants are compromised (Figures 4C–4E). As opposed to dsRNA unwinding, G-quadruplex remodeling is not dependent on ATP (Figure 4C).

Taken together the data indicate that DDX21 uses its C-terminal accessory domains and dimerization to unwind dsRNA and specifically recognize and remodel RNA G-quadruplex structures.

DDX21 Domain Architecture and Conformational Flexibility

Owing to the flexible nature of helicases, the combination of solution scattering (SAXS) data and flexible fitting of individual domain structures is a powerful tool to understand the conformation and the

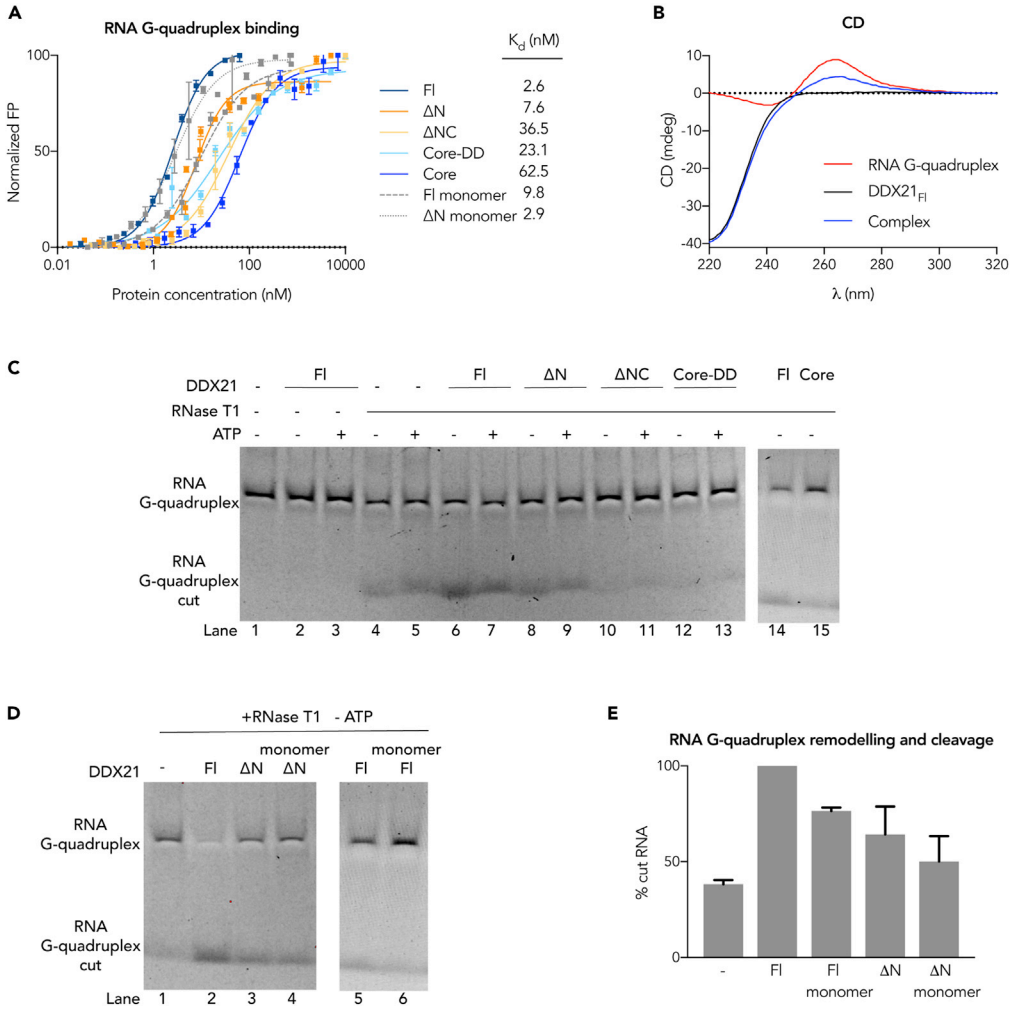


Figure 4. DDX21 Activity on RNA G-Quadruplexes

(A) FP binding curves for the different DDX21 variants and the RNA G-quadruplex Q2. Error bars represent the standard deviation of three independent measurements (see also Figure S9 and Table S2).

(B) The circular dichroism (CD) spectra of the RNA G-quadruplex (in red) with a parallel structure with its characteristic maximum at 260 nm and minimum at 240 nm. In the presence of $DDX21_{Fl}$ (blue), this structure is maintained, as showed by the maximum at 260 nm that is not present in the spectrum of the protein alone (black) (see also Figure S10).

(C and D) RNA G-quadruplex remodeling assay comparing the activity of the different DDX21 mutants. In native condition polyacrylamide gels, the 5′-FAM-labeled RNA was visualized by measuring the fluorescent signal at 535 nm. The activity was calculated by measuring the intensity of the product (cut) RNA bands, with respect to $DDX21_{Fl}$ activity.

(E) Quantification of the G-quadruplex remodeling assay for the constructs with activity above background: Fl, ΔN, and the monomeric ΔN and Fl mutants. The graph on the right summarizes the results from three independent experiments (see also Figure S11). Error bars represent the standard deviation.

structure-function relationship in these multi-domain systems (Huen et al., 2017; Mallam et al., 2011; Rambo, 2015; Xu et al., 2017). We, therefore, used this method to explore the quaternary organization of DDX21 and study the spatial arrangement of the domains with respect to each other to better rationalize the role of dimerization for DDX21 activity.

SAXS was performed coupled to size exclusion chromatography (SEC-SAXS) to avoid aggregation effects (Figure 5 and S12, the relevant parameters in Table S3). The Guinier analysis of the scattering curves (Figure S13) shows good linearity indicating neither aggregation nor polydispersity effects. According to SAXS molecular weights estimates, all the constructs (except the ΔN monomer mutant) are dimers, as expected (Table S3). The Guinier approximation gives an estimated radius of gyration (R_g) of 70 Å for Fl, 48 Å for ΔN,

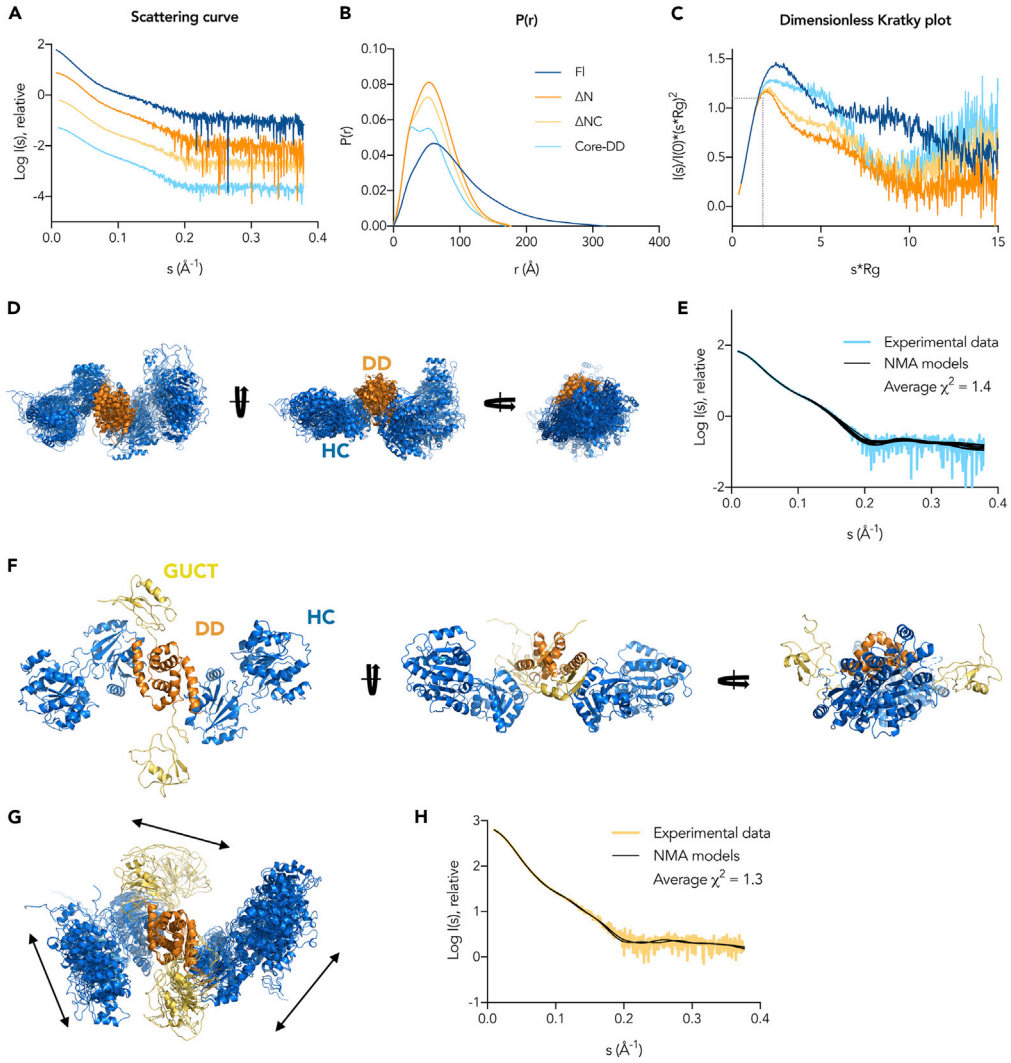


Figure 5. SAXS-Guided Modeling and DDX21 Conformational Flexibility

(A) Experimental scattering curves generated from the SEC-SAXS traces shown in Figure S12A.

(B) Pair distance distribution functions indicating the maximum dimension of the particles, D_{max}.

(C) Dimensionless Kratky plot shows the characteristic profile for flexible multi-domain proteins (see also Figures S12 and S13).

(D) The ensemble of the outcome structures from the NOLB NMA for $DDX21_{Core\text{-}DD}$ displays a linear arrangement of the HC and DD domains (see also Figure S14).

(E) Comparison between the calculated scattering curves from the ensemble models (black) and the $DDX21_{Core\text{-}DD}$ experimental scattering curve (light blue), giving an excellent fit with χ^2 of 1.4.

(F) Representative $DDX21_{\Delta NC}$ model of the ensembles created by flexible fitting, depicting the flatness of apo $DDX21_{\Delta NC}$ in solution.

(G) Ensemble of the outcome structures from the NOLB NMA for $DDX21_{\Delta NC}$. The arrows indicate the domain movements after aligning the conformers by the DD. The HC domains are shown in blue, the DD in orange, and the GUCT domains in yellow.

(H) Comparison between the calculated scattering curves from the ensemble models (black) and the $DDX21_{\Delta NC}$ experimental scattering curve (yellow), giving an excellent fit with χ^2 of 1.3 (see also Figures S15–S17).

46 Å for ΔNC, and 44 Å for Core-DD. The large difference between the R_g value of the $DDX21_{Fl}$ and the rest of the constructs suggests the presence of a core globular structure and an elongated flexible N terminus. This is apparent when comparing the pair distance distribution functions P(r) and the maximum dimensions (D_{max}) for the $DDX21_{Fl}$ and the variants in Figure 5B and Table S3. $DDX21_{Fl}$ shows a profile characteristic of elongated structures with a very large D_{max} of 319 Å and a main peak at a much shorter radius (dark blue in

Figure 5B). When the N-terminal 185 amino acids are removed, the P(r) function resembles that of a globular structure with a much smaller D_{max} of about 177 Å (orange in Figure 5B). In addition, the dimensionless Kratky plot (Figure 5C) contains features typical of multi-domain proteins with flexible linkers for all constructs, as expected, but indicates a larger flexibility for $DDX21_{Fl}$ (dark blue). As a reference, compact, globular molecules show a peak at the Guinier-Kratky point (1.1, √3) where the dashed lines in Figure 5C intercept (Receveur-Brechot and Durand, 2012).

Compared with the compact globular shape previously described for the monomeric $DDX21_{Core}$ (Chen et al., 2020), $DDX21_{Core-DD}$ uncovers a flexible, elongated, and bimodal structure (as seen by the two peaks in the P(r) function in Figure 5B light blue), fitting to a dimer of two HC domains (Figures 5D and S14). In solution, $DDX21_{Core-DD}$ does not form the "inverted V-shaped" structure as seen, for example, in the CshA crystal structure (Huen et al., 2017) (Figures S14A and S14B), but a more linear arrangement with the HC domains extending away from the DD (Figures 5D and S14C and S14D). Interestingly, the dimensionless Kratky plot in Figure 5C suggests that $DDX21_{Core-DD}$ is less compact than $DDX21_{\Delta NC}$ and $DDX21_{\Delta N}$. Therefore, the SAXS data suggest that the presence of the GUCT domain makes the protein more compact. More importantly, $DDX21_{\Delta N}$, which contains the C-terminal basic tails with the F/PRGQR repeats, is still very compact, suggesting that the C-terminal tail provides structure. The basic tail expressed and purified alone in solution is soluble and unfolded (Figure S15). However, the R_g and D_{max} values for the $DDX21_{\Delta NC}$ and $DDX21_{\Delta N}$ constructs are very similar, suggesting that the basic tail may pack against the GUCT domain, and perhaps, contributes to the higher thermal stability of $DDX21_{\Delta N}$ with respect to $DDX21_{\Delta NC}$ (T_m (ΔN) = 62°C versus T_m (ΔNC) = 38°C, Figure S5B). In line with this hypothesis, inspection of the GUCT domain electrostatic surface clearly shows an exposed acidic surface (Figure S2C) that could be binding the basic tail.

We have reconstructed a complete model for $DDX21_{Fl}$ without the unstructured N- and C- terminal tails, using the SAXS data to guide the modeling and analyze its flexibility (Figure 5E). We compared two approaches that can deal with multi-domain flexible systems in solution: the ensemble optimization method (EOM) (Bernadó et al., 2007; Tria et al., 2015) and the nonlinear rigid block normal-mode analysis method (NOLB NMA) (Hoffmann and Grudinin, 2017). EOM creates an ensemble of structures that together represent the scattering macromolecule (Figures S16). For the NOLB NMA procedure, $DDX21_{\Delta NC}$ models were modified to fit the SAXS experimental curve by efficiently exploring the configurations of the flexible linkers and rigid domains (see Transparent Methods and Figures 5E, 5F, and S17). Both EOM and NOLB NMA generate flexible conformational ensembles where the HC domains can be found in a range of open conformations characteristic of apo helicases. The two protomers are linked by the small helical DD, which becomes the only point of contact between them (Figure 5). From this central point, the GUCT domains also extend out in the same plane as the HC domains. The variations within the ensembles (depicted by arrows in Figures 5G and S17) result from the opening and closing of the HC and the movements of the GUCT domains with respect to the DD, highlighting the origin of flexibility. This structural organization places the DD at the center of the protein dimer, acting as a structural linker between the HC domains and the GUCT-basic tail units, coordinating their joined activities onto the RNA substrates, explaining why maintaining the dimer interface is essential for enzymatic activity.

DISCUSSION

We identify human DDX21 as a dimer and suggest a model of its quaternary structure relating these findings to its RNA-binding and unwinding activities. Using integrative modeling that combines biochemistry experiments, homology modeling, and SAXS, we found that the human DDX21 helicase is a homodimer, in which residues 570–620 form a hydrophobic dimerization domain similar to those found in other bacterial DEAD-box helicases. Sequence alignments and structural comparison using the Dali server (Holm, 2019) were unable to identify other human DEAD-box helicases containing an analogous dimerization domain. Consequently, this is a unique characteristic of the two human paralogs DDX21 and DDX50.

DDX21 can unwind dsRNA substrates with both 3′ (this work) and 5′ single-stranded overhangs (Valdez et al., 1997) in an ATP-dependent manner. We propose that DDX21 uses dimerization and the basic C-termini to recognize and unwind dsRNA substrates. Dimerization on a dsRNA substrate has only been observed in the crystal structure of the truncated HC of the DEAD-box helicase DDX3X in the pre-unwound state (Song and Ji, 2019). As observed for DDX3X and also the homologous bacterial CsdA (Xu et al., 2017), DDX21 dimerization allows for cooperativity and enhances the unwinding activity of duplexes.

However, as opposed to DDX3X HC that binds dsRNA with a 2:1 protein:RNA stoichiometry, full-length DDX21 is able to form a complex with 2:2 stoichiometry (as seen for the R-loop substrate, Figure 2). Within the DDX21 dimer, there are two HCs and two GUCT-basic tail units that can bind two substrates simultaneously or perhaps also work synergistically to unwind or remodel one substrate RNA. The different stoichiometries may depend on the nature (shape and length) of the substrate and may have important functions in the cell.

$DDX21_{Fl}$ binds RNA G-quadruplex with the highest affinity, when compared with other RNA structures, and this preference already exists in the $DDX21_{Core\text{-}DD}$ and $DDX21_{Core}$ constructs. Interestingly, the bacterial DEAD-box helicase CsdA also shows the same behavior, given that the DD-Core dimer of CsdA has specificity for the G-rich substrates over other RNA structures tested (Huen et al., 2017). As these interactions are ionic strength dependent, this preference for the G-quadruplex disk-like structures might be determined by the electrostatic surface charge distribution of the HC dimer. The accessory domains will contribute to enhancing the affinity of the dimer, modulating the electrostatic effects that govern RNA binding, as suggested here for the GUCT-basic tail unit of DDX21. As opposed to dsRNA unwinding, RNA G-quadruplex remodeling does not require ATP, indicating that the protein is using a different mechanism than that of canonical DEAD-box helicases. Still, an intact dimer interface is essential for maintaining this activity. *In vitro* experiments described by McRae et al. (2017, 2018) complement these findings as they show that the C-terminal domain of DDX21 can remodel G-quadruplexes in an ATP-independent manner. Their experiments used a construct covering the DD-GUCT and basic tail. In our experiments, we have separated the contributions of the C-terminal basic tail from the GUCT and dimerization domains, which allows us to show that dimerization in DDX21 is important also for its ATP-independent G-quadruplex remodeling function.

The combination of dimerization and C-terminal accessory domains seems to be a recurring feature within the DEAD-box family (as seen in the helicases Hera, CsdA, DDX21, and DDX50). This indicates that oligomerization cannot be overlooked when studying the mechanism of action of this helicase family. Further evidence on the role of oligomerization comes from the association of other helicases in cells (DDX5 and DDX17, Ogilvie et al., 2003) or *in vitro* (human DDX3X and its yeast ortholog Ded1p, Putnam et al., 2015; Sharma et al., 2017; Song and Ji, 2019). DDX3X and Ded1p are reported to function as trimers based on the estimation of their rate constants in unwinding reactions that show a sigmoidal dependence on helicase concentration (Sharma et al., 2017; Yang and Jankowsky, 2006). Both of them contain a low-complexity region at their C termini, which turned out to be essential for oligomerization. Human DDX21 differs from these eukaryotic enzymes as, based on the present results, it functions as a dimer, containing a dimerization domain that is shared with bacterial helicase homologs. Therefore, eukaryotic helicases seem to have the capacity to define and control their activity using different accessory domains able to promote specific oligomeric states.

DEAD-box RNA helicases have indispensable roles in cell homeostasis; however, our understanding of their RNA recognition and function in the cell at the molecular and structural levels is sparse primarily because these enzymes are challenging to characterize structurally given their flexible and multi-domain nature. As further studies are necessary for understanding the implications of oligomerization on the catalytic mechanism, as well as on the recognition of binding partners (in the case of DDX21 these could be phosphatases, viral proteins, or the 7SK RNP complex, for instance, Calo et al., 2015; Chen et al., 2014; De Wever et al., 2012), these findings will contribute to shed light on the molecular mechanisms of DEAD-box helicases in the cell and their potential relevance as therapeutic targets.

Limitations of the Study

The fact that the activity assays presented here were carried out using purified protein and *in vitro* conditions might be a limitation of the work, as we are not studying the enzyme in the cellular environment in conjunction with protein or nucleic acid partners. In addition, the models presented are at medium-low resolution given the experimental data available and have to be interpreted at this resolution level.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Matteo Dal Peraro (matteo.dalperaro@epfl.ch).

Materials Availability

All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

Data and Code Availability

SAXS data and models were deposited in the Small Angle Scattering Biological Data Bank SASBDB with accession codes SASDGU9, SASDGV9, SASDGW9, SASDGX9, and SASDGY9.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101811.

ACKNOWLEDGMENTS

We thank the EPFL Protein Production and Structure Core Facility for providing the equipment for the biophysical characterization of the protein complexes; Bruno Correia lab for the use of the MALS machine; Cy Jeffries and Mark Tully for their assistance with SAXS data preparation and analysis; Amédé Larabi, Jule Goike, and Eliane Duperrex for their assistance with protein purification; Luciano Abriata, Florence Pojer, Kelvin Lau, and Ben F. Luisi for helpful discussions and advice; BM29 staff scientists for their assistance in SAXS data collection; and staff at the EPFL Biomolecular screening facility for excellent support and training. This work was supported by the EPFL and the Swiss National Science Foundation (200021_157217 to M.D.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.J.M. and M.D.P.; Investigation, M.J.M., A.K., A.D., J.S., M.C.M., and D.A.; Funding Acquisition, M.D.P.; Writing –Original Draft, M.J.M; Writing – Review & Editing, M.J.M., A.K., J.S., M.C.M., D.A., S.M., D.T., and M.D.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 25, 2020
Revised: September 2, 2020
Accepted: November 11, 2020
Published: December 18, 2020

REFERENCES

Agarwala, P., Pandey, S., and Maiti, S. (2015). The tale of RNA G-quadruplex. Org. Biomol. Chem. *13*, 5570–5585.

Argaud, D., Boulanger, M.-C., Chignon, A., Mkannez, G., and Mathieu, P. (2019). Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus prevents R-loop formation and promotes transcription. Nucleic Acids Res. *47*, 8424–8438.

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J. Am. Chem. Soc. *129*, 5656–5664.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249–253.

Calo, E., Gu, B., Bowen, M.E., Aryan, F., Zalc, A., Liang, J., Flynn, R.A., Swigut, T., Chang, H.Y., Attardi, L.D., et al. (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. Nature *554*, 112–117.

Chen, G., Liu, C.-H., Zhou, L., and Krug, R.M. (2014). Cellular DDX21 RNA helicase inhibits influenza A virus replication but is counteracted by the viral NS1 protein. Cell Host Microbe *15*, 484–493.

Chen, Z., Li, Z., Hu, X., Xie, F., Kuang, S., Zhan, B., Gao, W., Chen, X., Gao, S., Li, Y., et al. (2020). Structural basis of human helicase DDX21 in RNA binding, unwinding, and antiviral signal activation. Adv. Sci. (Weinh.) *7*, 2000532.

Daubner, G.M., Cléry, A., and Allain, F.H.-T. (2013). RRM-RNA recognition: NMR or crystallography…and new findings. Curr. Opin. Struct. Biol. *23*, 100–108.

De Wever, V., Lloyd, D.C., Nasa, I., Nimick, M., Trinkle-Mulcahy, L., Gourlay, R., Morrice, N., and Moorhead, G.B.G. (2012). Isolation of human mitotic protein phosphatase complexes: identification of a complex between protein phosphatase 1 and the RNA helicase Ddx21. PLoS One *7*, e39510.

Degiacomi, M.T., Iacovache, I., Pernot, L., Chami, M., Kudryashev, M., Stahlberg, H., van der Goot, F.G., and Dal Peraro, M. (2013). Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. Nat. Chem. Biol. *9*, 623–629.

Del Campo, M., Mohr, S., Jiang, Y., Jia, H., Jankowsky, E., and Lambowitz, A.M. (2009). Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. J. Mol. Biol. *389*, 674–693.

Dong, Y., Ye, W., Yang, J., Han, P., Wang, Y., Ye, C., Weng, D., Zhang, F., Xu, Z., and Lei, Y. (2016). DDX21 translocates from nucleus to cytoplasm

and stimulates the innate immune response due to dengue virus infection. Biochem. Biophys. Res. Commun. *473*, 648–653.

Flores-Rozas, H., and Hurwitz, J. (1993). Characterization of a new RNA helicase from nuclear extracts of HeLa cells which translocates in the 5′ to 3′ direction. J. Biol. Chem. *268*, 21372–21383.

Fonti, G., Marcaida, M.J., Bryan, L.C., Träger, S., Kalantzi, A.S., Helleboid, P.-Y.J., Demurtas, D., Tully, M.D., Grudinin, S., Trono, D., et al. (2019). KAP1 is an antiparallel dimer with a functional asymmetry. Life Sci. Alliance *2*, e201900349.

Grohman, J.K., Del Campo, M., Bhaskaran, H., Tijerina, P., Lambowitz, A.M., and Russell, R. (2007). Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. Biochemistry *46*, 3013–3022.

Hammond, J.A., Zhou, L., Lamichhane, R., Chu, H.-Y., Millar, D.P., Gerace, L., and Williamson, J.R. (2018). A survey of DDX21 activity during Rev/RRE complex formation. J. Mol. Biol. *430*, 537–553.

Hao, H., Han, T., Xuan, B., Sun, Y., Tang, S., Yue, N., and Qian, Z. (2019). Dissecting the role of DDX21 in regulating HCMV replication. J. Virol. *93*, e01222-19.

Henn, A., Bradley, M.J., and De La Cruz, E.M. (2012). ATP utilization and RNA conformational rearrangement by DEAD-box proteins. Annu. Rev. Biophys. *41*, 247–267.

Henning, D., So, R.B., Jin, R., Lau, L.F., and Valdez, B.C. (2003). Silencing of RNA helicase II/Gualpha inhibits mammalian ribosomal RNA production. J. Biol. Chem. *278*, 52307–52314.

Hoffmann, A., and Grudinin, S. (2017). NOLB: nonlinear rigid block normal-mode analysis method. J. Chem. Theor. Comput. *13*, 2123–2134.

Holm, L. (2019). Benchmarking fold detection by DaliLite v.5. Bioinformatics *35*, 5326–5327.

Holmström, T.H., Mialon, A., Kallio, M., Nymalm, Y., Mannermaa, L., Holm, T., Johansson, H., Black, E., Gillespie, D., Salminen, T.A., et al. (2008). c-Jun supports ribosomal RNA processing and nucleolar localization of RNA helicase DDX21. J. Biol. Chem. *283*, 7046–7053.

Huen, J., Lin, C.-L., Golzarroshan, B., Yi, W.-L., Yang, W.-Z., and Yuan, H.S. (2017). Structural insights into a unique dimeric DEAD-box helicase CshA that promotes RNA decay. Structure *25*, 469–481.

Jarmoskaite, I., and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. Annu. Rev. Biochem. *83*, 697–725.

Kim, D.-S., Camacho, C.V., Nagari, A., Malladi, V.S., Challa, S., and Kraus, W.L. (2019). Activation of PARP-1 by snoRNAs controls ribosome biogenesis and cell growth via the RNA helicase DDX21. Mol. Cell *75*, 1270–1285.e14.

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.

Klostermeier, D., and Rudolph, M.G. (2009). A novel dimerization motif in the C-terminal domain of the Thermus thermophilus DEAD box helicase Hera confers substantial flexibility. Nucleic Acids Res. *37*, 421–430.

Kossen, K., Karginov, F.V., and Uhlenbeck, O.C. (2002). The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA. J. Mol. Biol. *324*, 625–636.

Linden, M.H., Hartmann, R.K., and Klostermeier, D. (2008). The putative RNase P motif in the DEAD box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. Nucleic Acids Res. *36*, 5800–5811.

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

Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. Proc. Natl. Acad. Sci. U. S. A. *105*, 20209–20214.

Mallam, A.L., Jarmoskaite, I., Tijerina, P., Del Campo, M., Seifert, S., Guo, L., Russell, R., and Lambowitz, A.M. (2011). Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. Proc. Natl. Acad. Sci. U. S. A. *108*, 12254–12259.

Mallam, A.L., Del Campo, M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. Nature *490*, 121–125.

McRae, E.K.S., Booy, E.P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S.A. (2017). Human DDX21 binds and unwinds RNA guanine quadruplexes. Nucleic Acids Res. *45*, 6656–6668.

McRae, E.K.S., Davidson, D.E., Dupas, S.J., and McKenna, S.A. (2018). Insights into the RNA quadruplex binding specificity of DDX21. Biochim. Biophys. Acta Gen. Subj. *1862*, 1973–1979.

McRae, E.K.S., Dupas, S.J., Booy, E.P., Piragasam, R.S., Fahlman, R.P., and McKenna, S.A. (2020). An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21. RNA *26*, 44–57.

Minshall, N., and Standart, N. (2004). The active form of Xp54 RNA helicase in translational repression is an RNA-mediated oligomer. Nucleic Acids Res. *32*, 1325–1334.

Mohr, G., Del Campo, M., Mohr, S., Yang, Q., Jia, H., Jankowsky, E., and Lambowitz, A.M. (2008). Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. J. Mol. Biol. *375*, 1344–1364.

Naji, S., Ambrus, G., Cimermančič, P., Reyes, J.R., Johnson, J.R., Filbrandt, R., Huber, M.D., Vesely, P., Krogan, N.J., Yates, J.R., et al. (2012). Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. Mol. Cell. Proteomics *11*, M111.015313.

Ogilvie, V.C., Wilson, B.J., Nicol, S.M., Morrice, N.A., Saunders, L.R., Barber, G.N., and Fuller-Pace, F.V. (2003). The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. Nucleic Acids Res. *31*, 1470–1480.

Ohnishi, S., Pääkkönen, K., Koshiba, S., Tochio, N., Sato, M., Kobayashi, N., Harada, T., Watanabe, S., Muto, Y., Güntert, P., et al. (2009). Solution structure of the GUCT domain from human RNA helicase II/Gu beta reveals the RRM fold, but implausible RNA interactions. Proteins *74*, 133–144.

Ou, Y., Fritzler, M.J., Valdez, B.C., and Rattner, J.B. (1999). Mapping and characterization of the functional domains of the nucleolar protein RNA helicase II/Gu. Exp. Cell Res. *247*, 389–398.

Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015). The conformational plasticity of eukaryotic RNA-dependent ATPases. FEBS J. *282*, 850–863.

Putnam, A.A., Gao, Z., Liu, F., Jia, H., Yang, Q., and Jankowsky, E. (2015). Division of labor in an oligomer of the DEAD-Box RNA helicase Ded1p. Mol. Cell *59*, 541–552.

Rambo, R.P. (2015). Resolving individual components in protein-RNA complexes using small-angle X-ray scattering experiments. Methods Enzymol. *558*, 363–390.

Receveur-Brechot, V., and Durand, D. (2012). How random are intrinsically disordered proteins? A small angle scattering perspective. Curr. Protein Pept. Sci. *13*, 55–75.

Samatanga, B., and 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 Res. *42*, 10644–10654.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat. Rev. Genet. *16*, 583–597.

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

Sharma, D., Putnam, A.A., and Jankowsky, E. (2017). Biochemical differences and similarities between the DEAD-box helicase orthologs DDX3X and Ded1p. J. Mol. Biol. *429*, 3730–3742.

Sloan, K.E., Leisegang, M.S., Doebele, C., Ramírez, A.S., Simm, S., Safferthal, C., Kretschmer, J., Schorge, T., Markoutsa, S., Haag, S., et al. (2015). The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. Nucleic Acids Res. *43*, 553–564.

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

Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev. *31*, 1370–1381.

Sonn-Segev, A., Belacic, K., Bodrug, T., Young, G., VanderLinden, R.T., Schulman, B.A., Schimpf, J., Friedrich, T., Dip, P.V., Schwartz, T.U., et al. (2020). Quantifying the heterogeneity of macromolecular machines by mass photometry. Nat. Commun. *11*, 1772.

Steimer, L., Wurm, J.P., Linden, M.H., Rudolph, M.G., Wöhnert, J., and Klostermeier, D. (2013). Recognition of two distinct elements in the RNA substrate by the RNA-binding domain of the T. thermophilus DEAD box helicase Hera. Nucleic Acids Res. *41*, 6259–6272.

Story, R.M., Li, H., and Abelson, J.N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii. Proc. Natl. Acad. Sci. U. S. A. *98*, 1465–1470.

Tamò, G.E., Abriata, L.A., and Dal Peraro, M. (2015). The importance of dynamics in integrative modeling of supramolecular assemblies. Curr. Opin. Struct. Biol. *31*, 28–34.

Tria, G., Mertens, H.D.T., Kachala, M., and Svergun, D.I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ *2*, 207–217.

Valdez, B.C. (2000). Structural domains involved in the RNA folding activity of RNA helicase II/Gu protein. Eur. J. Biochem. *267*, 6395–6402.

Valdez, B.C., Henning, D., Perumal, K., and Busch, H. (1997). RNA-unwinding and RNA-folding activities of RNA helicase II/Gu–two activities in separate domains of the same protein. Eur. J. Biochem. *250*, 800–807.

Westermarck, J., Weiss, C., Saffrich, R., Kast, J., Musti, A.-M., Wessely, M., Ansorge, W., Séraphin, B., Wilm, M., Valdez, B.C., et al. (2002). The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. EMBO J. *21*, 451–460.

Xing, Y.-H., Yao, R.-W., Zhang, Y., Guo, C.-J., Jiang, S., Xu, G., Dong, R., Yang, L., and Chen, L.-L. (2017). SLERT regulates DDX21 rings associated with Pol I transcription. Cell *169*, 664–678.e6.

Xu, L., Wang, L., Peng, J., Li, F., Wu, L., Zhang, B., Lv, M., Zhang, J., Gong, Q., Zhang, R., et al. (2017). Insights into the structure of dimeric RNA helicase CsdA and indispensable role of its C-terminal regions. Structure *25*, 1795–1808.e5.

Yang, Q., and Jankowsky, E. (2006). The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. Nat. Struct. Mol. Biol. *13*, 981–986.

Zhang, H., Zhang, Y., Chen, C., Zhu, X., Zhang, C., Xia, Y., Zhao, Y., Andrisani, O.M., and Kong, L. (2018). A double-negative feedback loop between DEAD-box protein DDX21 and Snail regulates epithelial-mesenchymal transition and metastasis in breast cancer. Cancer Lett. *437*, 67–78.

Zhang, Y., Baysac, K.C., Yee, L.-F., Saporita, A.J., and Weber, J.D. (2014). Elevated DDX21 regulates c-Jun activity and rRNA processing in human breast cancers. Breast Cancer Res. *16*, 449.

Zhang, Z., Kim, T., Bao, M., Facchinetti, V., Jung, S.Y., Ghaffari, A.A., Qin, J., Cheng, G., and Liu, Y.-J. (2011). DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. Immunity *34*, 866–878.

iScience, Volume *23*

Supplemental Information

The Human RNA Helicase DDX21 Presents a Dimerization Interface Necessary for Helicase Activity

Maria J. Marcaida, Annamaria Kauzlaric, Alice Duperrex, Jenny Sülzle, Martin C. Moncrieffe, Damilola Adebajo, Suliana Manley, Didier Trono, and Matteo Dal Peraro

Content

- Supplemental Figures S1-S17
- Supplemental Tables S1-S3
- Transparent Methods
- Supplemental References

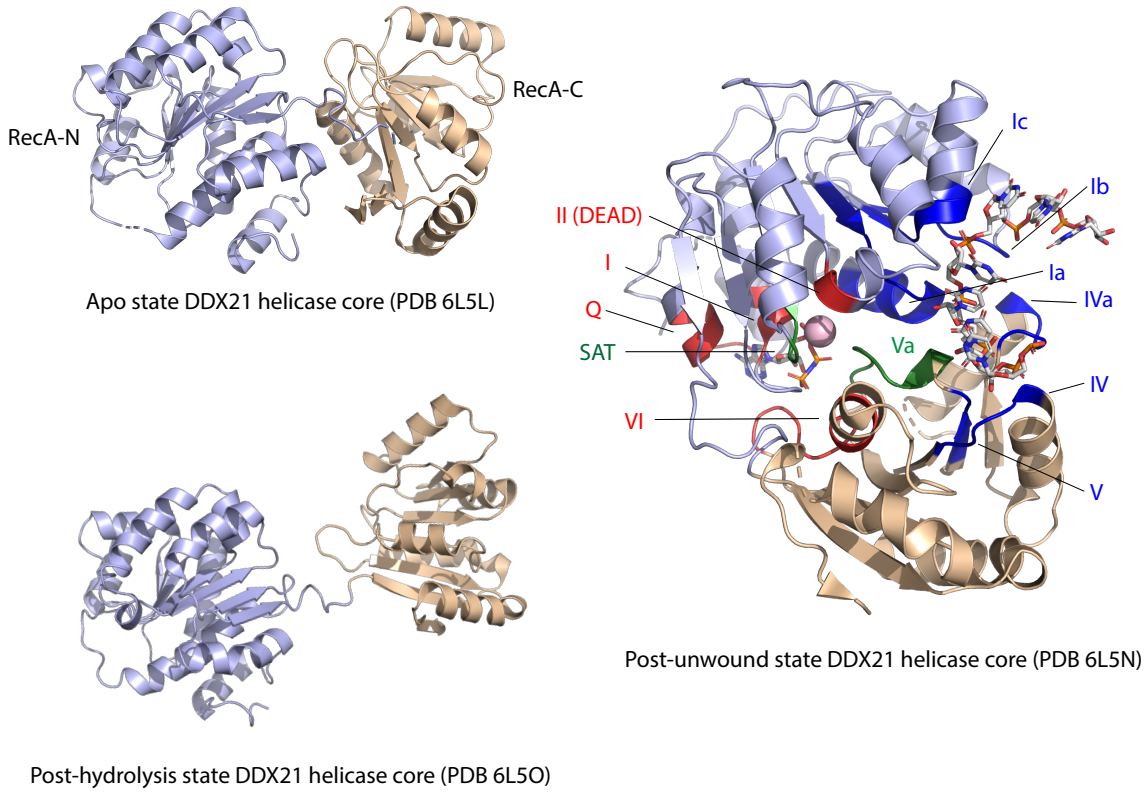


Figure S1. Crystal structure of the DDX21 HC (Related to Figure 1A). Cartoon representation of the DDX21 helicase core domain in three conformations of the two RecA-like domains (RecA-N in light blue and RecA-C in wheat colour) (Chen et al., 2020). On the left, two crystal structures are shown as examples of the various open conformations (apo state PDB 6L5L and post-hydrolysis state PDB 6L5O) to highlight the range of movement allowed by the flexible linker. On the right, the crystal structure of DDX21 HC in complex with ssRNA, magnesium (pink sphere) and an ATP analogue, AMPPNP (PDB 6L5N) represents the post-unwound (closed) conformation. The characteristic motifs of DEAD-box helicases are indicated as follows: motifs interacting with ATP in red, motifs interacting with RNA in blue, and motifs involved in intramolecular interactions in green (Linder and Jankowsky, 2011).

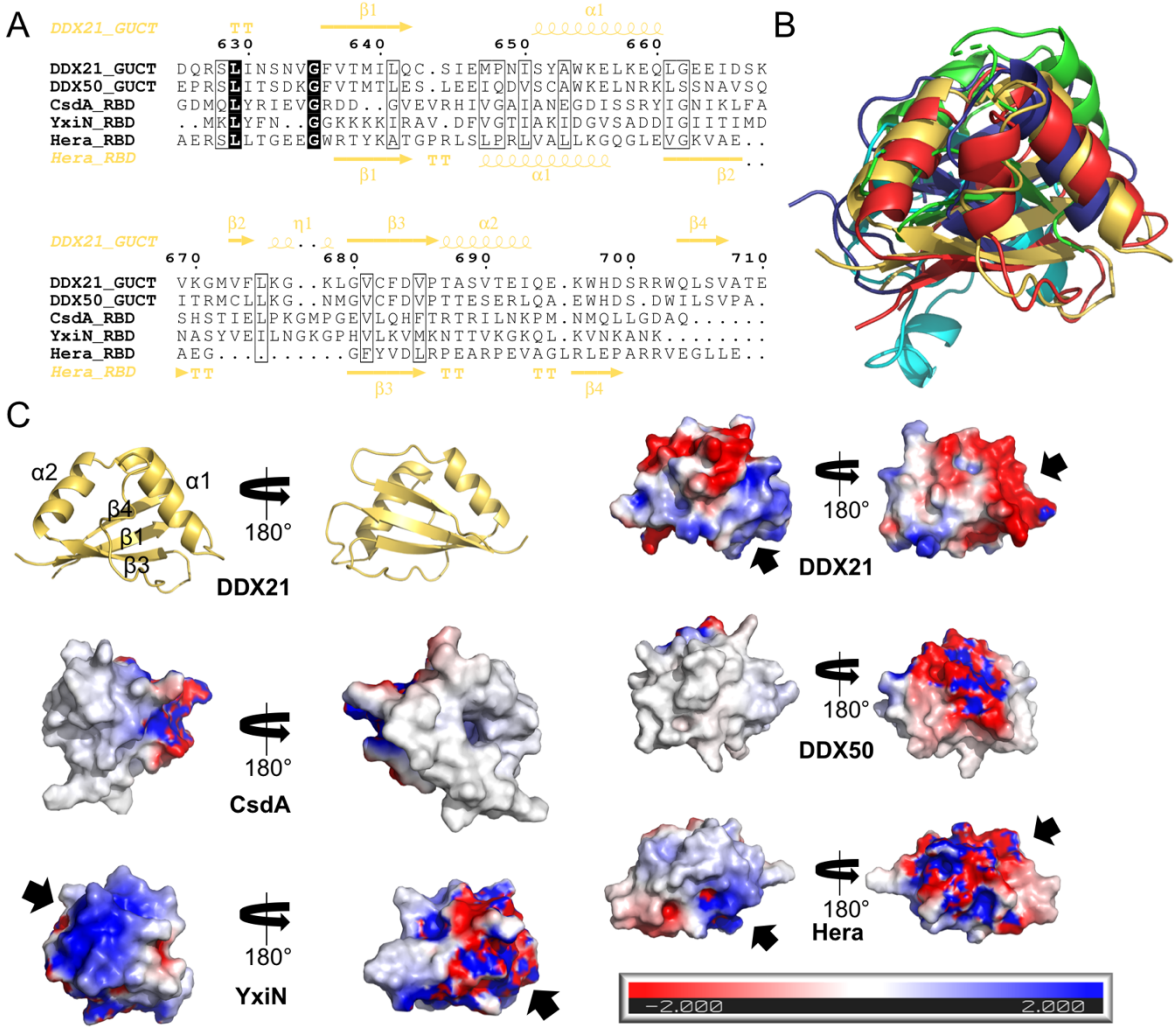


Figure S2. Analysis of the GUCT domain (Related to Figure 1A). A. Sequence alignment of five RNA binding domains (RBDs) or RNA recognition motifs (RRMs) from DEAD-box helicases shows very little sequence conservation. UniProt codes: DDX21: Q9NR30; DDX50: Q9BQ39; CsdA: P0A9P6; YxiN: P42305 and Hera: Q72GF3. Strictly conserved residues are highlighted in black, and residues with a global similarity score above 0.7 (from 0 to 1) are boxed. The secondary structure of the DDX21 GUCT domain is depicted on top, coloured in yellow. The sequence identity between DDX21 GUCT and the RRMs from Hera, CsdA and YxiN is ~30% and ~50% with the GUCT domain of DDX50. **B**. Superposition of the five RBD structures shows an overall fold conservation but Hera, YxiN and CsdA RBDs seem to have different RNA binding modes (Hardin et al., 2010; Ohnishi et al., 2009; Steimer et al., 2013; Xu et al., 2017). Some of the secondary structure elements are more conserved than others, namely α1-helix and strands β1, β3 and β4. The RBD from Hera is shown in blue, from CsdA in cyan, from DDX50 in red, from YxiN in green and from DDX21 in yellow. **C**. Cartoon representation of the GUCT domain of DDX21 showing the position of the secondary structure elements and comparison of the electrostatic potential surfaces of the different RBDs (in the same views as in the cartoon representation). The electrostatic potential distribution was calculated using APBS Electrostatics and the negative and positive potential values (from -2 to 2 kT/e) are shown in red and blue, respectively. RBDs from CsdA and DDX50 do not display obvious charged patches on their surface, but a large hydrophobic area. On the other hand, Hera, Yxin and DDX21 present positively charged regions and DDX21 shows a very large negatively charged patch (all indicated by black arrows).

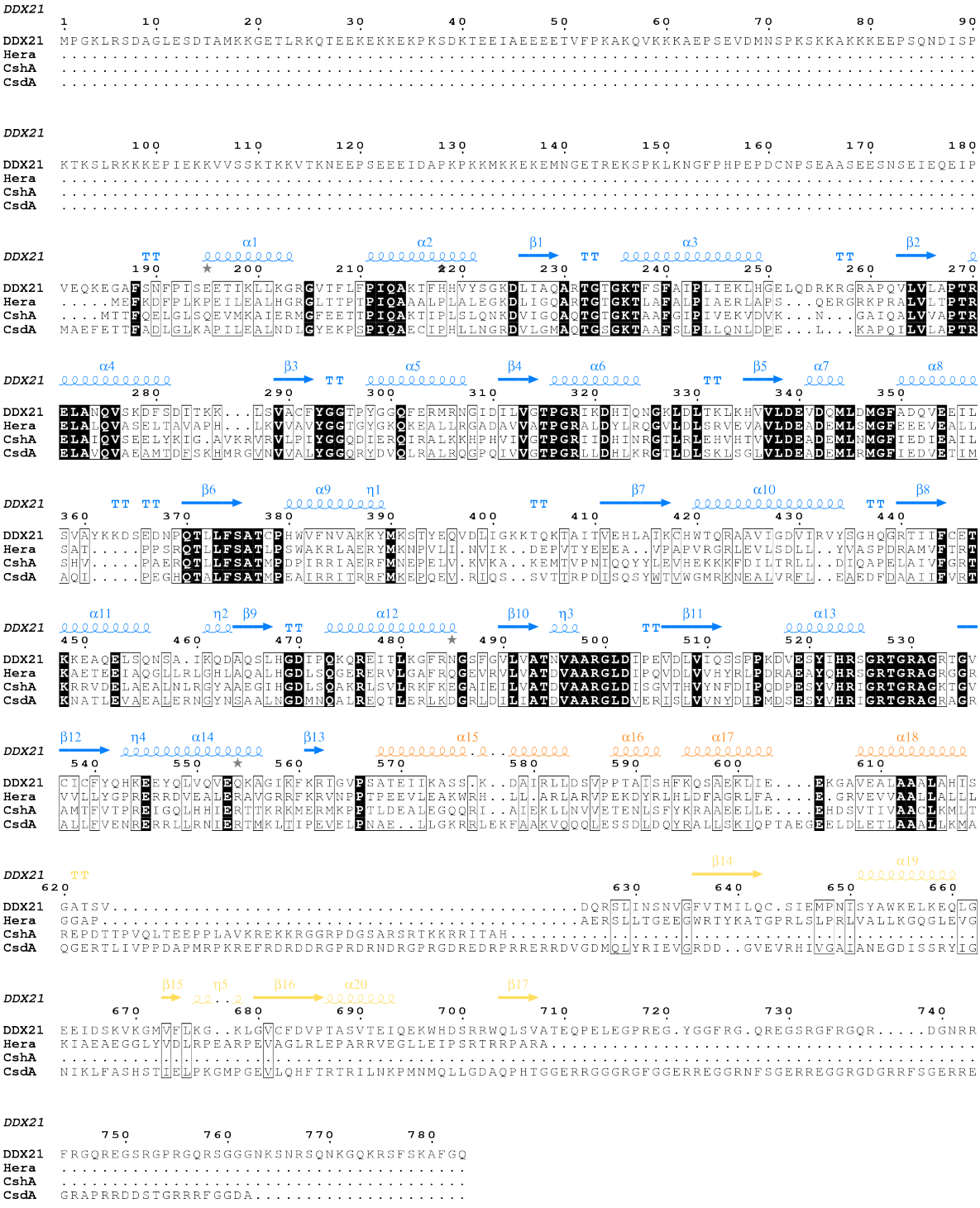


Figure S3. Sequence alignment (Related to Figure 1). Sequence alignment of DDX21 and the three bacterial dimeric DEAD-box helicases used as templates for homology modelling: Hera (Heat resistant RNA-dependent ATPase) from *Thermus thermophilus* (Klostermeier, 2013); CshA (Cold shock helicase A) from the Gram-positive bacteria *Geobacillus stearothermophilus* (Huen et al., 2017); and CsdA (Cold-shock DEAD-box protein A) from *Escherichia coli* (*E. coli*) (UniProt codes: DDX21: Q9NR30; Hera: Q72GF3; CshA: A0A0K2H973 and CsdA: P0A9P6). Strictly conserved residues are highlighted in black, and residues with a global similarity score above 0.7 (from 0 to 1) are boxed. The secondary structure of the DDX21 model is depicted on top, coloured as the domains in the main figures, namely, helicase core in blue, dimerization domain in orange and GUCT domain in yellow.

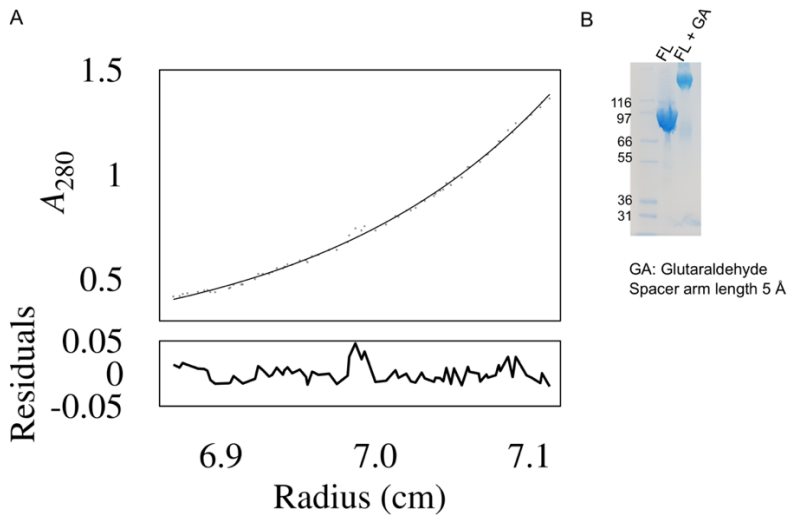
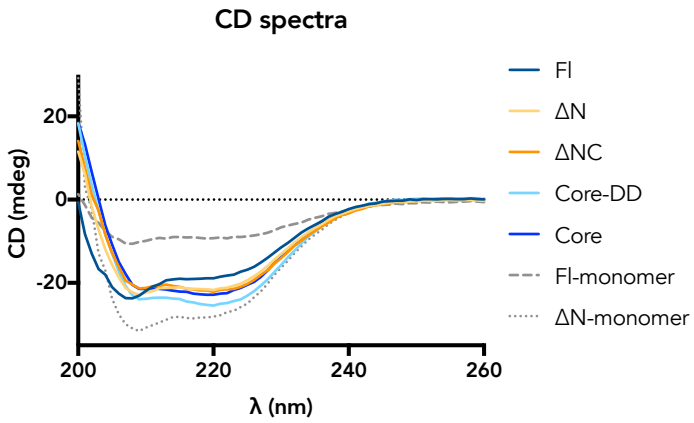


Figure S4. DDX21 is a homodimer (Related to Figure 1). A. Analytical ultracentrifugation equilibrium sedimentation experiment measures a molecular weight (Mw) for $DDX21_{FI}$ of 154 kDa, agreeing with a dimer species (15% error as dimer Mw is 182 kDa). **B**. $DDX21_{FI}$ was covalently crosslinked with glutaraldehyde, resulting in protein oligomers having molecular weights consistent with dimeric forms on SDS-PAGE.

A



B

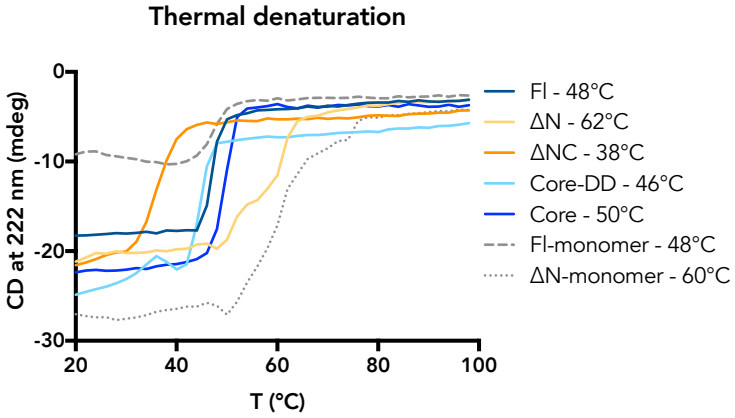


Figure S5. Folding and thermal stability of the DDX21 variants measured by circular dichroism (CD) (Related to Figure 1). A. CD spectra indicating that all the variants have secondary structure. **B**. Thermal denaturation curves monitored by measuring the CD signal at 222 nm across a temperature range from 20 to 98°C. The melting temperature T_m was calculated: T_m (Fl) = 48°C; T_m (ΔN) = 62°C; T_m (ΔNC) = 38°C; T_m (Core-DD) = 46°C; T_m (Core) = 50°C; T_m (Fl-monomer) = 48°C; T_m (ΔN-monomer) = 60°C. The T_m measured for $DDX21_{Fl}$ is the same as the one measured previously (McRae et al., 2017).

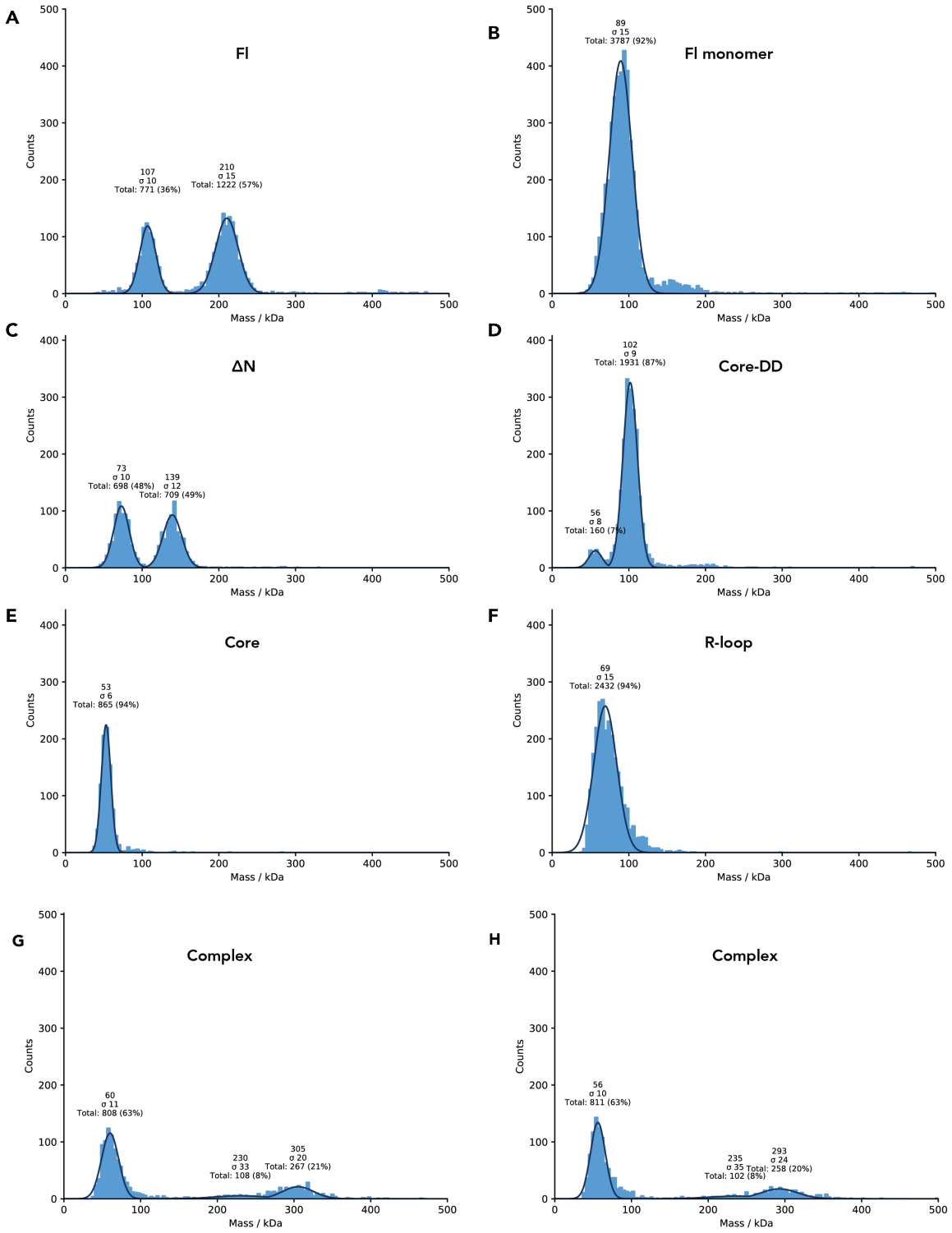


Figure S6. Replicas of the mass photometry measurements (Related to Figure 2). The samples were measured at the following concentrations: Fl at 30nM, Fl monomer mutant at 15nM; ΔN at 30nM, Core-DD at 15nM, Core at 15nM, R-loop at 15nM and the Fl-R-loop complex at 30nM.

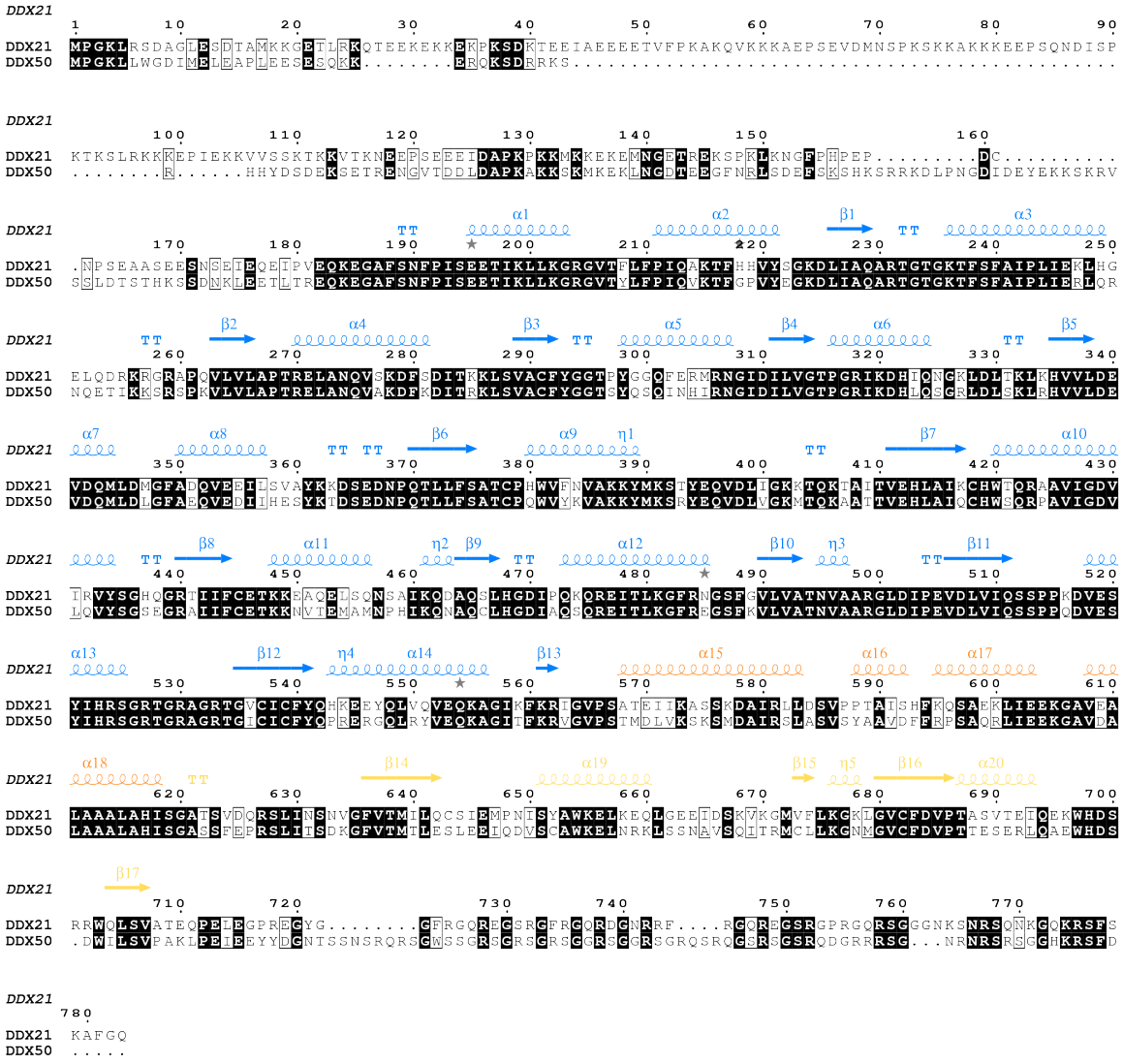
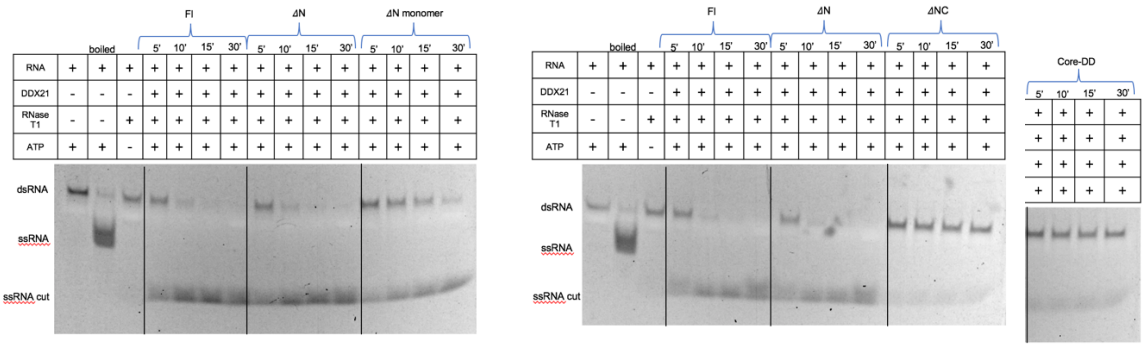
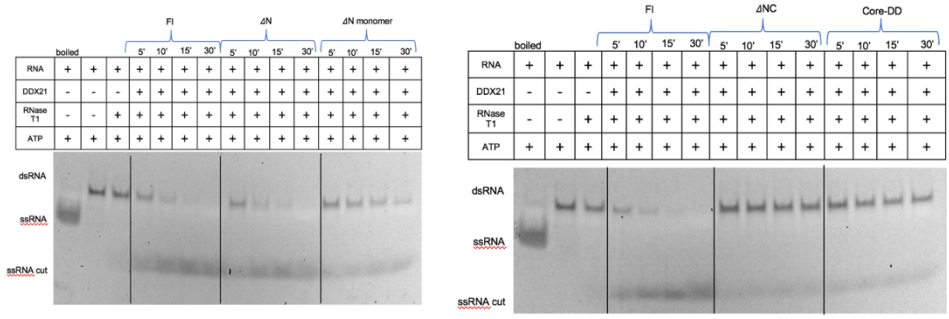


Figure S7. Sequence alignment of the two human paralogues, RH-II/Guα and RH-II/Guβ, i.e., DDX21 and DDX50 (Related to Figure 1). DDX50 localizes to nucleoli and nuclear speckles and may be involved in ribosomal RNA production and pre-mRNA splicing (Valdez et al., 2002a, 2002b). DDX21 and DDX50 share a similar domain architecture with an overall sequence identity of 55.6%. We suggest DDX50 may also form dimers. Another point worth highlighting is that the characteristic (F/P)RGQR repeats in the C-terminal tail of DDX21 are replaced by an arginine-serine-rich sequence in the C-terminal tail of DDX50 (Ohnishi et al., 2009; Valdez et al., 2002a). This may be the reason why DDX50 did not have RNA-folding activity (or misinterpreted G-quadruplexes helicase activity) *in vitro* (McRae et al., 2017; Valdez et al., 2002a). Colour coding and details of the alignment are as in **Figure S3.**

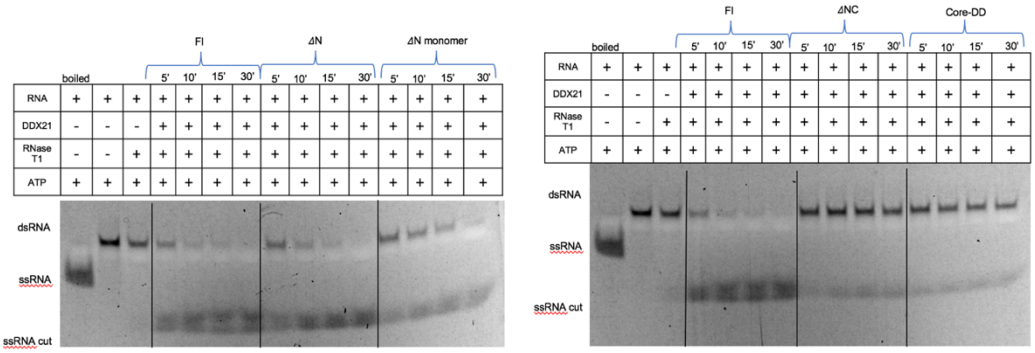
FI, ΔN, ΔN monomer, ΔNC and Core-DD dsRNA unwinding assay 1:



FI, ΔN, ΔN monomer, ΔNC and Core-DD dsRNA unwinding assay 2:



FI, ΔN, ΔN monomer, ΔNC and Core-DD dsRNA unwinding assay 3:



FI monomer and Core dsRNA unwinding assays:

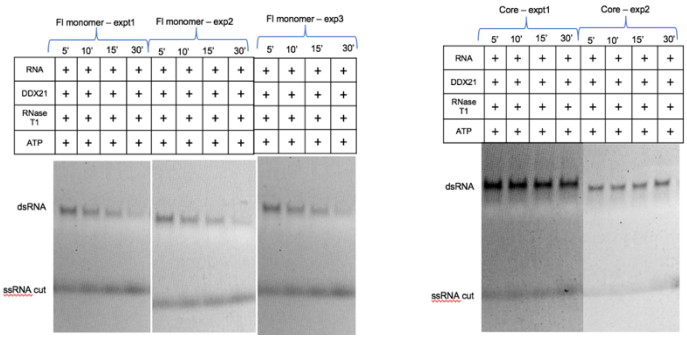


Figure S8. dsRNA helicase assays in triplicates (Related to Figure 3).

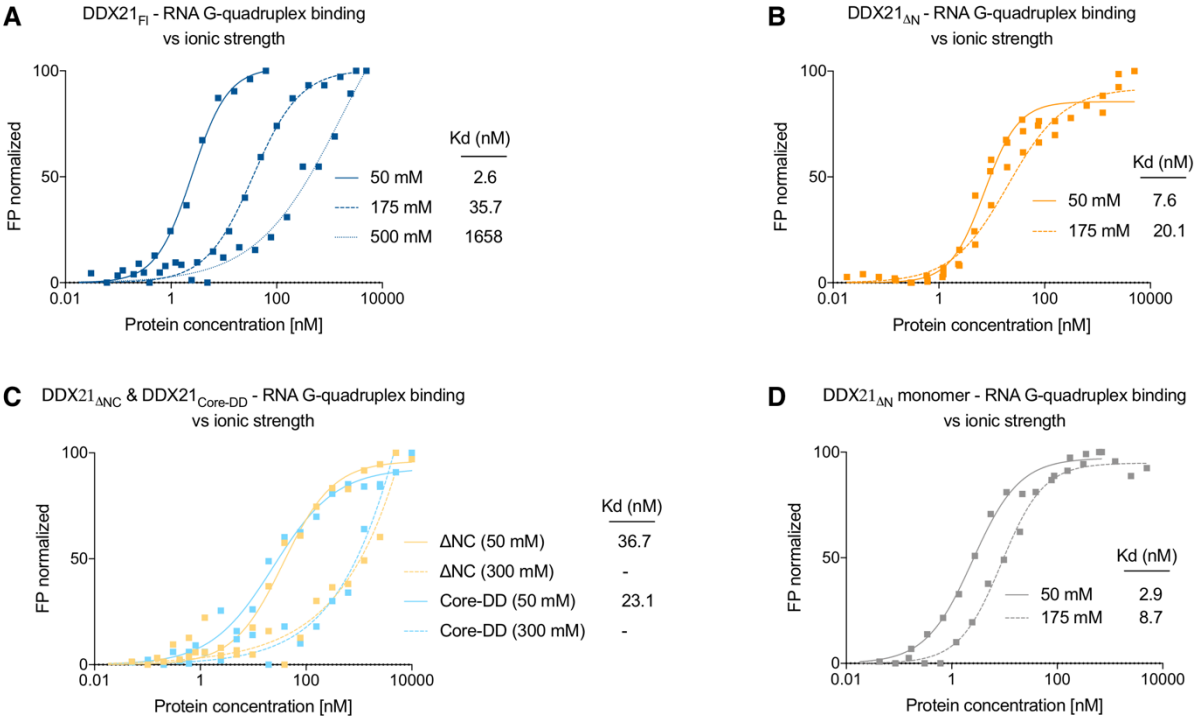


Figure S9. DDX21FI - RNA G-quadruplex binding affinity versus ionic strength (Related to Figure 4). **A-D.** FP binding curves for the DDX21 variants and the RNA G-quadruplex under different salt concentrations. The DDX21FI affinity for the RNA G-quadruplex decreases 14- and over 600-fold when the salt concentration increases from 50 mM to 175 mM and 500 mM, respectively (**Table S2**). This effect is also observed for all the mutants indicating that the recognition of the RNA is based on ionic interactions, as expected.

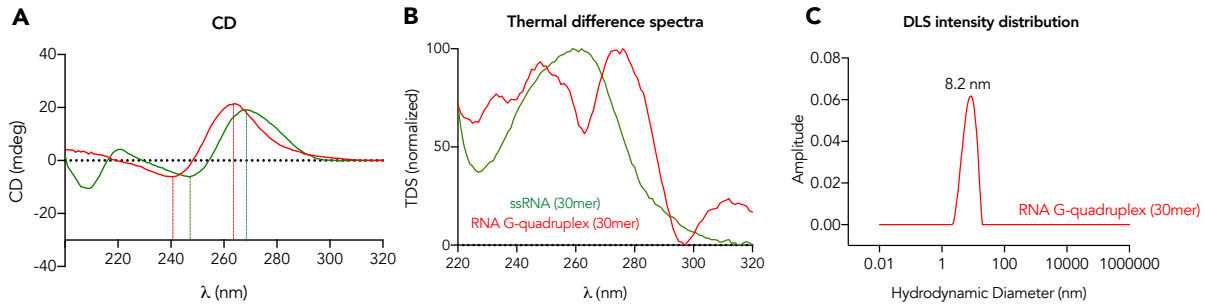


Figure S10. RNA G-quadruplex formation (Related to Figure 4B). A. CD spectra of the Q2 RNA G-quadruplex (30-mer- GUU GGG GCG GGC GUU GGG UUU GGG GGG ACG) used in the remodelling assays (in red) and a 30-mer ssRNA with sequence GCG UCG AUC CGA AAC UAU ACU UAA UUU UAA (in green) that does not form a G-quadruplex. The overlay clearly shows that the minima and maxima do not coincide. **B.** Thermal difference spectra of the two RNA molecules display the characteristic profiles of ssRNA and RNA G-quadruplexes (Mergny et al., 2005). **C.** Dynamic light scattering (DLS) was used to detect a globular structure with a hydrodynamic diameter of ~8.2 nm for the Q2 RNA G-quadruplex, in the same order of magnitude as that described by (Ariyo et al., 2015).

FI, FI monomer, ΔN and ΔN monomer RNA G-quadruplex remodelling assays

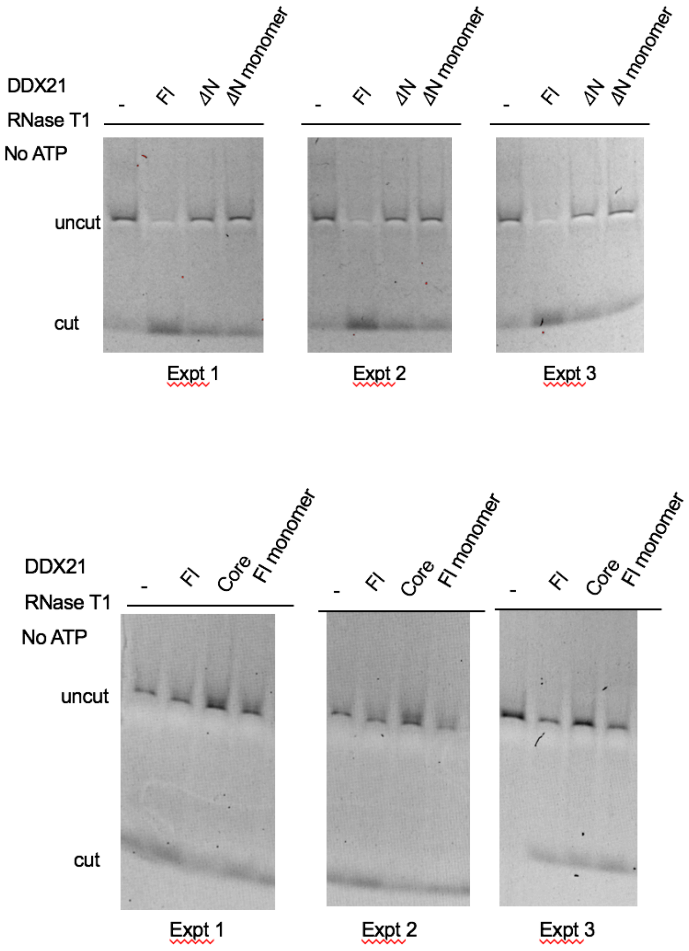


Figure S11. RNA G-quadruplex remodelling assays in triplicates (Related to Figure 4).

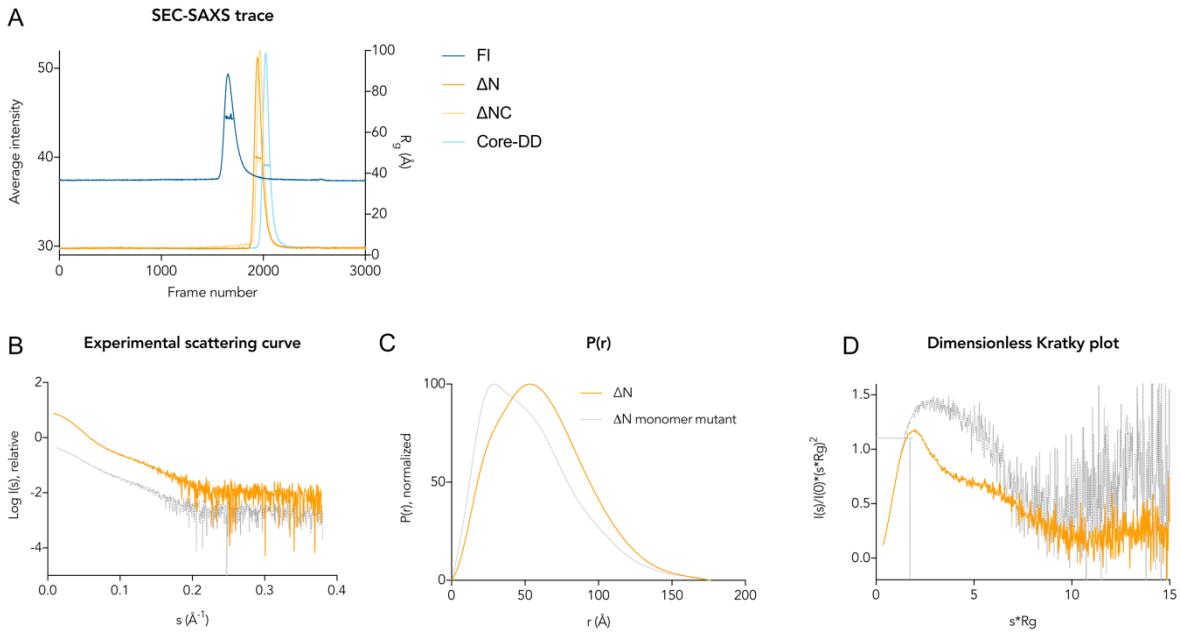


Figure S12. SEC-SAXS analysis of DDX21 (Related to Figure 5). **A**. SEC-SAXS elution profiles used to generate the data in **Figure 5**. The estimated R_g values for the peak regions are also shown. **B-D**. Comparison between DDX21$_{\Delta N}$ and its monomeric mutant (residues 611 to 616 (LAAALA) mutated to arginines) in orange and grey, respectively, using SAXS. **B**. Experimental scattering curve. **C**. Pair distance distribution functions indicating the D_{max} of both proteins is around 177 Å. **D**. Dimensionless Kratky plot shows the characteristic profile for multidomain proteins for ΔN and indicates more flexibility for the monomeric mutant.

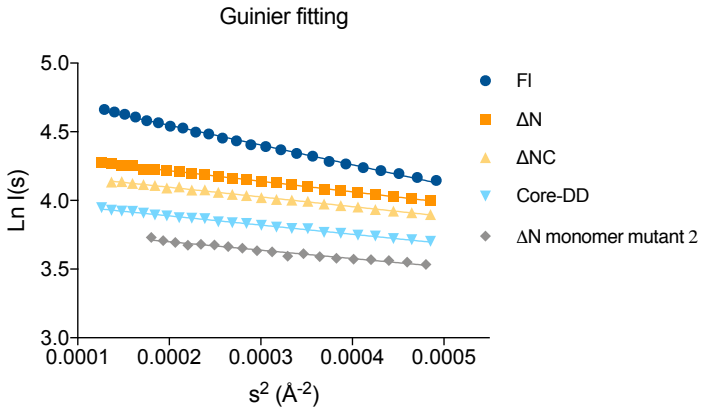


Figure S13. Guinier fitting of the SAXS scattering curves for the DDX21 variants shows no interparticle aggregation effects (Related to Figure 5).

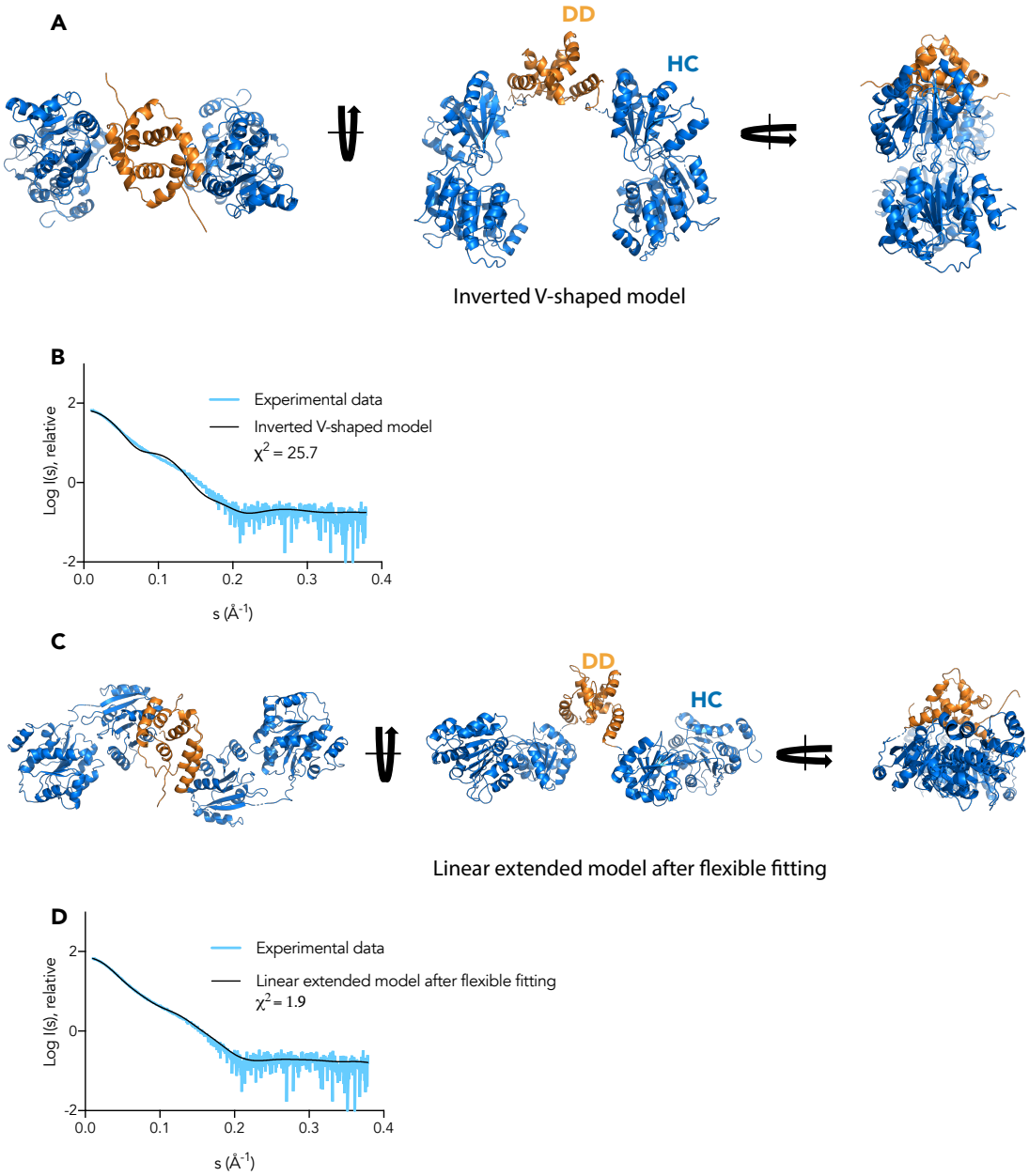


Figure S14. **DDX21$_{Core-DD}$ conformation in solution (Related to Figure 5). A.** Cartoon representation of the model of DDX21$_{Core-DD}$ based on the inverted V-shape conformation displayed by the CshA crystal structure (PDB ID 5IVL, (Huen et al., 2017)). **B.** The calculated scattering profile from the structure shown in **A** (black line) does not fit the experimental SAXS scattering curve (light blue line) (χ^2 = 25.7). Most likely, DDX21$_{Core-DD}$ adopts a more linear arrangement of its domains as shown in **C.**, whose calculated scattering profile fits much better to the experimental data (χ^2 = 1.9), shown in **D.**

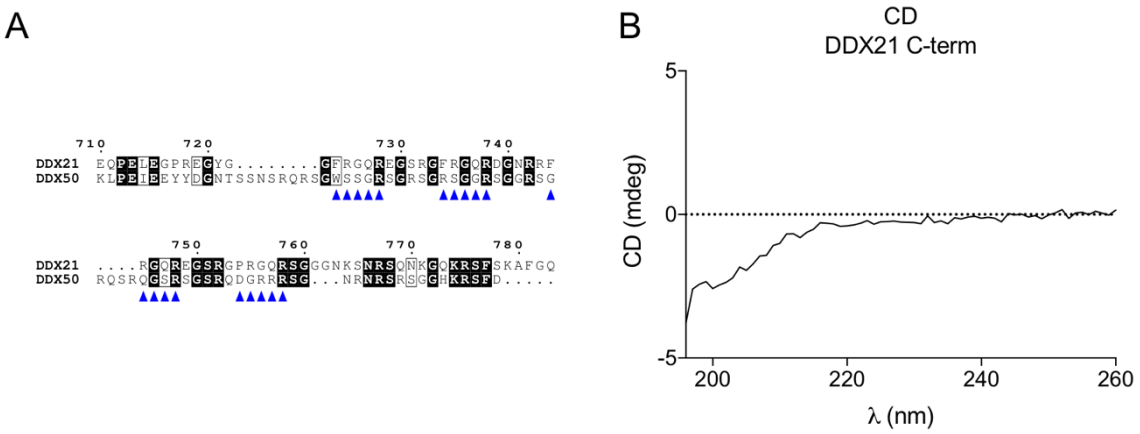


Figure S15. The C-terminal amino acids of DDX21 are unfolded (Related to Figure 5). **A**. C-terminal amino acids of DDX21 aligned to those from DDX50. The F/PRGQR repeats are highlighted by blue triangles. **B**. CD spectrum of the purified C-terminal domain alone shows it has no secondary structure, agreeing with predictions based on sequence alignments (JPred4 (Drozdetskiy et al., 2015)).

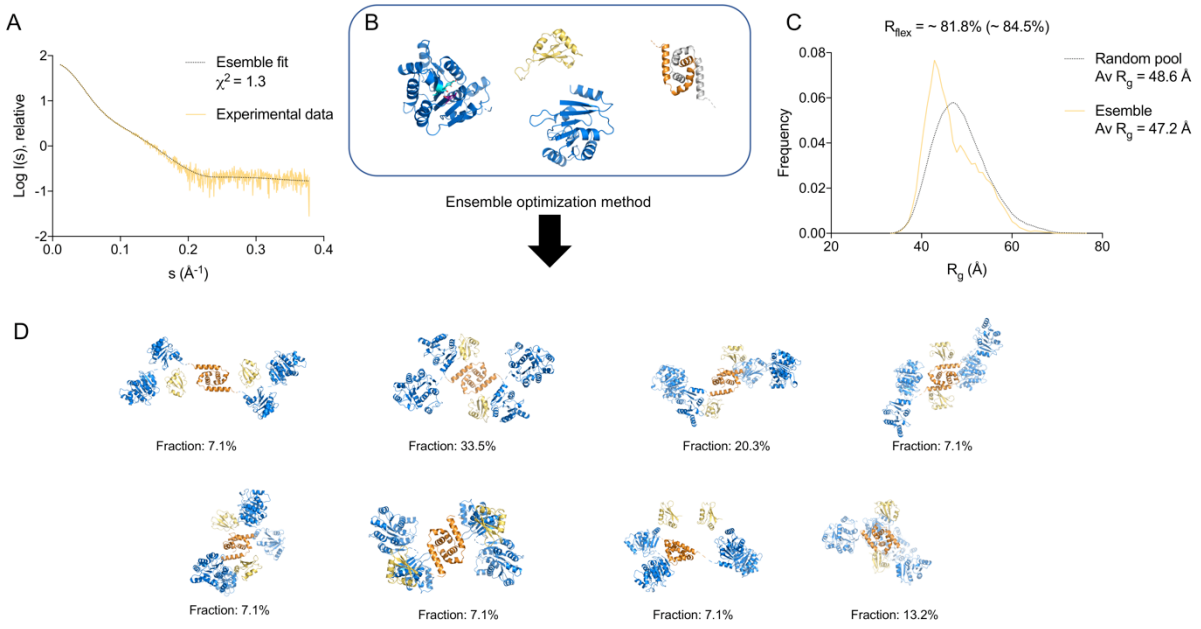


Figure S16. EOM analysis of DDX21$_{\Delta NC}$ (Related to Figure 5). A. Comparison between the calculated scattering curve from the whole EOM ensemble (black) and the DDX21$_{\Delta NC}$ experimental scattering curve (yellow), giving an excellent fit with χ^2 of 1.3. **B**. Input models for the globular domains. **C**. The R_g distribution for the random pool of models (dotted black) and for the ensemble of conformations that altogether give a good fit to the scattering intensity curve (yellow) are shown, with R_{flex} values of 84 and 81%, respectively. The distributions of the calculated R_g values are very close to those of the random pools and R_{flex} for the ensembles are slightly smaller to those from the random pools, indicating that the protein is highly flexible. **D**. Cartoon representation of the eight conformers in the EOM ensemble, showing the conserved flat structure. The frequency at which each conformer is present within the ensemble is shown as % fraction below each model. The HC is coloured in blue, the DD in orange and the GUCT domain in yellow.

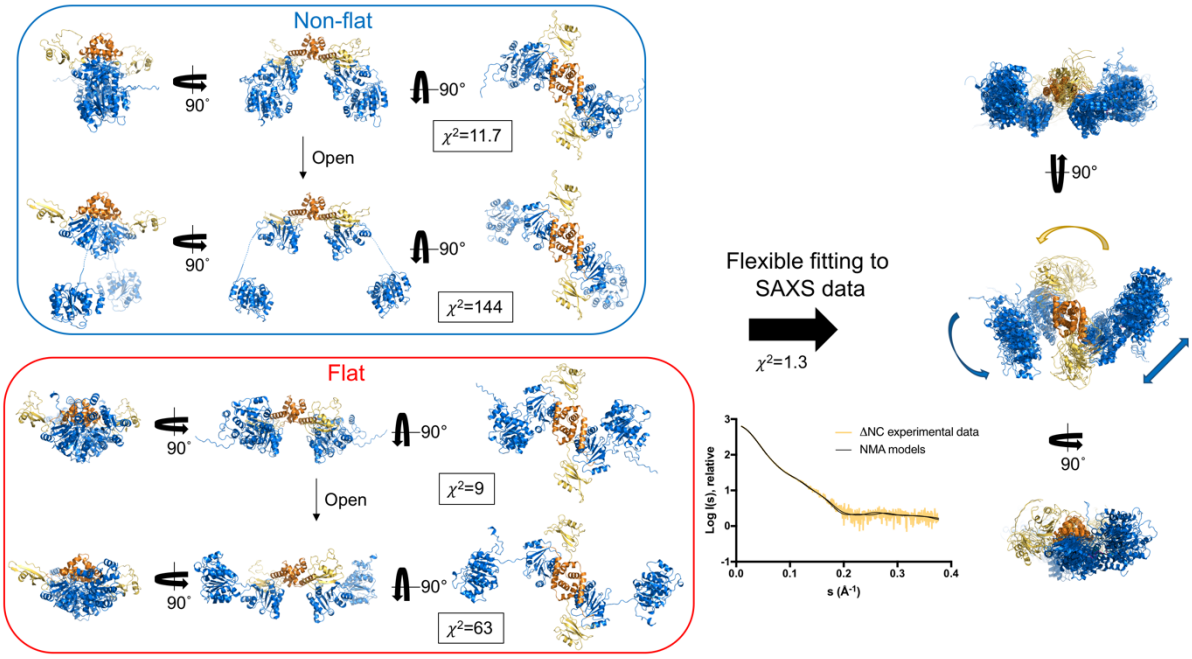


Figure S17. NOLB NMA of DDX21$_{\Delta NC}$ (Related to Figure 5). DDX21$_{\Delta NC}$ input models for the NOLB NMA procedure are shown in cartoon representation, grouped by the conformation of the helicase core (blue) with respect to the DD (orange): flat or not-flat ("inverted V-shaped"). The χ^2 values before the flexible fitting are indicated next to each model. The ensemble of resulting structures is shown on the right, where the arrows highlight the large movements of the domains. The final ensembles have a good fit to the experimental scattering curve with an average χ^2 value of 1.3.

Table S1. K_d values (nM) obtained from FP measurements (Related to Figures 3 and 4). The mean values from three independent measurements and their standard deviation were used to make the non-linear fit to the equation "Specific binding with Hill slope" in Prism, to calculate the K_d values and their uncertainties.

	Fl	⊿N	⊿NC	Core-DD	Core	Fl monomer	⊿N monomer
G-quadruplex	2.6±0.1	7.6 ±0.5	36.5±2.7	23.1±3.2	62.5 ± 4.9	9.8 ±0.9	2.9±0.3
ssRNA	30.7 ± 2.8	40.3±2.6	687.7±52.7	330.4±19.1	1525 ± 112.2	29.9 ± 2.7	24.9±1.7
dsRNA	10.2±0.4	15.5±0.8	217.4±10.4	143.2±6.1	353.3 ± 27.9	13.9 ± 2.5	121.8±19.3

Table S2. K_d values (nM) obtained from FP measurements using RNA G-quadruplex and different salt concentrations (Related to Figure 4). The mean values from three independent measurements and their standard deviation were used to make the non-linear fit to the equation "Specific binding with Hill slope" in Prism, to calculate the K_d values and their uncertainties.

	Fl	⊿N	⊿NC	Core-DD	⊿N monomer
50 mM*	2.6±0.1	7.6 ± 0.5	36.7±2.7	23.1±3.2	2.9±0.3
175/300 mM**	35.7±2.8	20.1±3.8	-	-	8.7±0.8
500 mM***	1658±1532	NM	NM	NM	NM

K_d values in nM. NM = Not measured; - = weak binding, K_d cannot be fit. *50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$. **50mM Tris pH 7.5, 300 mM NaCl/175 mM KCl, 5 mM $MgCl_2$. ***25 mM Tris, 10 mM Hepes, pH 7.5, 25 mM KCl, 500 mM NaCl, 2.5 mM $MgCl_2$, 5% glycerol.

Table S3. SAXS parameters (Related to Figure 5).

	FI	ΔN	ΔNC	Core-DD	ΔN-monomer
SASPDB accession code	SASDGU9	SASDGV9	SASDGW9	SASDGX9	SASDGY9
Data collection					
Instrument			ESRF BM29		
Beam size at sample (µm)			700x700		
Wavelength (Å)			0.992		
s range ($Å^{-1}$)			0.0025-0.500		
Detector			Pilatus 1M		
Detector distance (m)			2.867		
Exposure (s per image)			1		
Column			Superose 6 Increase 10/300		
Flow rate (ml/min)			0.5		
Sample volume (µl)			100		
Sample concentration (mg/ml)			9-12		
Temperature (K)			293		
Structural parameters					
R_g (Å) Guinier	70	48	46	44	41
R_g (Å) P(r)	76	49	47	46	44
D_{max} (Å)	319	176	177	169	177
Porod volume ($Å^3$)	681000	297000	238000	148000	111000
Molecular mass determination					
Theoretical MW (kDa)	90	68.7	60.6	49	69
MALLS MW (kDa)	197.7	143.2	120.5	91.4	68
DATPOROD MW (kDa)	400	174	140	87	65
DATVC MW (kDa)	227	156	122	88	68
DATMOW MW (kDa)	232	165	120	91	74
Data analysis software					
Data reduction			PRIMUS & ScÅtter		
Ab initio modelling			DAMMIF		
Homology modelling			Swiss Model		
Theoretical scattering curve calculation from models			Pepsi-SAXS		
3D graphics representation			Pymol/Chimera		

Transparent Methods

Protein expression and purification

The human DDX21 (Uniprot sequence Q9NR30, residues 1-783) expression plasmid carries the gene for the full-length protein in the pET-32 vector, with N-terminal Trx & Hexa-histidine tags, followed by a thrombin cleavage site. $DDX21_{\Delta N}$ (residues 186-783), $DDX21_{\Delta NC}$ (residues 186-710), $DDX21_{Core}$ (residues 186-564), $DDX21_{FI}$ monomer (residues 1-783 with the 611LAAALA616 to 611RRRRRR616 mutation), $DDX21_{\Delta N}$ monomer (residues 186-783 with the 611LAAALA616 to 611RRRRRR616 mutation) and the C-terminal basic tail (residues 704-783) were cloned in the first multi-cloning site in the pET-DuetI vector, with a N-terminal Hexa-histidine tag followed by a TEV (Tobacco Etch Virus) cleavage site. $DDX21_{Core-DD}$ (residues 186-620) was cloned with an uncleavable C-terminal hexa-histidine tag in the pET-28b(+). The coding sequences for all the variants, except FI, were optimized for *Escherichia coli* (*E. coli*) expression (Genscript). The plasmids were transformed in Rosetta (DE3) cells (Promega). Protein expression was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) when the cells reached an optical density of 0.6 and subsequent growth over night at 20°C. Cell pellets were resuspended in lysis buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 5 mM Imidazole and cOmplete™ Protease Inhibitor Cocktail (Roche)) and then lysed using sonication. Poly(ethyleneimine) solution (0.1% v/v) was added to the resulting suspensions which were then centrifuged (13000 rpm for 35 min at 4°C) and the supernatant was applied to a HisTrap HP column (GE Healthcare) previously equilibrated with lysis buffer. The proteins were eluted with a continuous gradient over 40 column volumes of elution buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 500 mM Imidazole). Subsequently, pure fractions were buffer exchanged into final buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM TCEP) using a HiPrep Desalting column (GE Healthcare) and stored over night at 4°C in final buffer. The proteins were additionally purified by Size Exclusion Chromatography (Superose^{TM} 6 Increase 10/30 GL, GE Healthcare) in final buffer, flash frozen in liquid nitrogen and stored at - 20°C. The purification of $DDX21_{FI}$ consists of an extra step where the protein is buffer exchanged in buffer C (50 mM Na Phosphate pH 7.4, 200 mM NaCl), cut by adding 8 units of thrombin (GE Healthcare) per mg of protein overnight at 4°C, and further purified by anion exchange chromatography (Resource S, GE Healthcare), using a linear gradient to 1 M NaCl, before the final gel filtration step.

RNA and R-loop preparation

All the RNA and DNA oligos were purchased HPLC purified, desalted and lyophilized from Microsynth (Switzerland). All RNA substrates were labelled with fluorescein at their 5' end (5'-FAM). The sequences are as follows: ssRNA – 15mer-5'-FAM-GUUUCGGAUCGACGC-3'; dsRNA (15/30mer – 3' overhang): 5'-FAM-GUUUCGGAUCGACGC-3' + 5'-GCGUCGAUCCGAAACUAUACUUAAUUUUAA-3' and the RNA G-quadruplex Q2, 30mer from the 3′ UTR of the PITX1 messenger RNA (McRae et al., 2017)): 5'-FAM-GUUGGGGCGGGCGUUGGGUUUGGGGGGACG-3'. Each RNA was diluted in ultra-pure DNase/RNase-free distilled water (Invitrogen) to a final concentration of 100 µM. For preparing stocks for the experiments, the RNAs were further dissolved to 10 µM in FP buffer (50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$), and annealed by heating at 95°C for 5 minutes and slowly cooling to room temperature. These stocks were aliquoted and stored at -20°C. For the dsRNA substrate, the 15mer 5'-FAM labelled oligonucleotide was mixed with equimolar amounts of the 30mer RNA oligonucleotide, resulting in a 3' overhang dsRNA. To form the R-loop substrate, the labelled 15-mer RNA oligo (5'-FAM-GUUUCGGAUCGACGC-3') was annealed with DNA oligo 1 (5′-GTA CCCG GGGATCCTCTAGAGTCGAGCGTCGATCCG AAACTT GGCACTGGCCGTCGTTTTACAAC-3′) and DNA oligo 2 (5′-GTTGTAAAACGACGGCCAGTGCCTTTTCCCAGCCTCAATCTCATCACTCTAG AGGATCCCCGGGTAC-3′) following the procedure described in (Song et al., 2017).

SEC-MALS

The molecular weights of the constructs were determined by size exclusion chromatography coupled to multi angle light scattering (SEC-MALS). The mass measurements were performed on a Dionex UltiMate3000 HPLC system equipped with a 3 angles miniDAWN TREOS static light scattering detector (Wyatt Technology). The sample volumes of 100 µl at a concentration of 5 to 10 mg/mL, were applied to a Superose 6 10/300 GL column (GE Healthcare) previously equilibrated with 20 mM HEPES pH

7.5, 300 mM NaCl, 1 mM TCEP at a flow rate of 0.5 mL/min. The data were analysed using the ASTRA 6.1 software package (Wyatt technology), using the absorbance at 280 nm and the theoretical extinction coefficient for concentration measurements.

Circular Dichroism

Spectra of the DDX21 variants (at 5 µM in 50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$) were collected with a J-815 CD Spectrometer (JASCO, Japan), (1 mm path length cuvette, 20°C, 260-200 nm). Thermal denaturation curves were acquired by heating the sample from 20 to 98°C, collecting CD data at 222 nm every two degrees. Spectra for the RNA G-quadruplex Q2 was measured at 20 µM in the same conditions except that the wavelength range was 320-220 nm.

Thermal difference spectra

UV/VIS absorbance spectra of the RNA molecules (20 µM in 50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$) were collected with a J-815 CD Spectrometer (JASCO, Japan), (1 mm path length cuvette, 320-220 nm) and background corrected against spectra of buffer alone. Thermal difference spectra (TDS) were generated by subtracting buffer-corrected spectra at 20°C from those at 98°C. For direct comparison between Q2 RNA and ssRNA, the spectra were normalised.

Dynamic Light Scattering

Dynamic Light Scattering (DLS) data were acquired using a Stunner instrument (Unchained Labs) using a standard plate, with 10x10sec acquisitions at 25°C. The RNA Q2 concentration was 10 µM in 50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$.

Mass photometry

All mass photometry measurements were executed on a Refeyn OneMP instrument. The calibration was done with a native marker protein standard mix (NativeMark Unstained Protein Standard, Thermo Scientific), which contains proteins ranging from 20 to 1200 kDa. Coverslips (24x50mm, No. 1.5H, Marienfeld) were cleaned by sequential sonication in Milli-Q water, isopropanol and Milli-Q-water, followed by drying with nitrogen. For each acquisition 2 µL of protein/RNA solution was applied to 18 µL buffer (50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$) in a gasket (CultureWellTM Reusable Gasket, Grace Bio-Labs) on a coverslip. Typical working concentrations were 15 to 30 nM. Movies were recorded at 999 Hz with an exposure time of 0.95 ms by using the AcquireMP software. All mass photometry movies were processed and analysed in the DiscoverMP software. Protein samples and R-loop only were measured in duplicates and the R-loop-protein complex in triplicate (**Figures 2 and S6**).

Crosslinking

An aliquot of the corresponding DDX21 variant was thawed and buffer exchanged using PD-10 Desalting columns (GE Healthcare UK Limited) into 20 mM sodium phosphate, 500 mM NaCl, pH 7.6 at 4°C. It was diluted to 0.4 mg/mL and was mixed with 1mM BS^3 (freshly prepared in 20 mM sodium phosphate pH 7.6) and incubated 10 min at room temperature before stopping the reaction by the addition of 50 mM Tris pH 7.6. For glutaraldehyde crosslinking, the protocol was the same, with protein concentration of 0.1 mg/mL, glutaraldehyde concentration of 0.1% and incubation time of 5 min.

Analytical Ultra Centrifugation

Sedimentation equilibrium measurements were performed using a Beckman XL-A ultracentrifuge equipped with absorbance and interference optics. Data was acquired using a sample volume of 140

μL at a concentration of 20 μM in 20 mM Hepes pH 7.5, 300 mM NaCl at 20°C and a rotor speed (An60-Ti) of 5000 rpm after equilibrium was attained. Partial specific volumes and buffer densities were estimated using SEDNTERP (Laue, T., Shaw, B. D., Ridgeway, T. M., and Pelletier, S. L, 1992) and data were analysed using SEDPHAT (Vistica et al., 2004).

Fluorescence Polarization assays for RNA binding affinity measurements

Concentrated DDX21 protein stocks were buffer exchanged into FP buffer (50 mM Tris pH 7.5, 50 mM KCl, 5 mM $MgCl_2$) and diluted to the highest concentration used in the binding assay, 5 or 10 μM, depending on the construct and the RNA substrate. This high concentrated sample was diluted in FP buffer in 1/2 series to the lowest chosen concentration. The 1/2 series with a different starting concentration was used when more data points were required. The RNA stocks were diluted to 10 nM in FP buffer. 90 μL of RNA were mixed with 90 μL of the protein dilution series directly in the 96-well plate (Greiner 96 well flat bottom black polystyrene plates (Sigma-Aldrich)) and incubated for 5 min. Samples were excited at 485 nm, and fluorescence was detected at 535 nm. FP data were collected in an Infinite® F500 (Tecan) plate reader at 22° C. Three independent titration curves were measured for each complex, which were used to extract the average FP value and the standard errors at each titration point. The Tecan I-control software calculated the FP data. These were normalized and fit to the equation "Specific binding with Hill slope" (non-linear regression analysis) in GraphPad Prism: Y = Bmax*X^h/(K_d^h + X^h), where Y is the normalised FP signal, B_{max} is the maximum signal, X is the protein concentration, K_d is the protein concentration needed to achieve a half-maximum binding at equilibrium and corresponds to the equilibrium binding constant if h = 1.

Helicase assay and RNase T1 cleavage

The dsRNA substrate was diluted to 500 nM in 25 mM Hepes pH 7.5, 2 mM $MgCl_2$. The DDX21 variants were used directly from frozen stocks in final buffer. Reaction volumes were 30 μl, with 100 nM dsRNA, 2.5 μM DDX21, 2 mM ATP and 0.025 U/μl of RNase T1 (Thermo Scientific, 1000 U/μl frozen stock) in 25 mM Hepes pH 7.5, 2 mM $MgCl_2$, 33 mM NaCl. Unwinding/cleavage reactions were performed at 37°C and samples were taken after 5, 10, 15 and 30 min of incubation after the addition of ATP. Reactions were stopped by adding equal volumes of 2x loading buffer (20% glycerol, 0.8% SDS, 10 mM EDTA, 0.04% bromophenol blue) and the samples were separated in native 12% polyacrylamide TBE gels at 4°C.

RNA G-quadruplex remodelling and RNase T1 cleavage

The RNA G-quadruplex Q2 substrate was diluted to 500 nM in 50 mM Tris pH 7,5, 50 mM KCl, 2 mM $MgCl_2$, 10% glycerol. The DDX21 variants were used directly from frozen stocks in final buffer. Reaction volumes were 30 μl, with 100 nM Q2 RNA, 2.5 μM DDX21, 2 mM ATP and 0.0125 U/μl of RNase T1 (Thermo Scientific, 1000 U/μl frozen stock) in 50 mM Tris pH 7,5, 50 mM KCl, 25 mM NaCl, 2 mM $MgCl_2$, 10% glycerol. Remodelling/cleavage reactions were performed at 30°C and samples were taken after 20 min of incubation after the addition of ATP. Reactions were stopped by adding equal volumes of 2x loading buffer (20% glycerol, 0.8% SDS, 10 mM EDTA, 0.04% bromophenol blue) and heating at 95°C for 5 min to improve the sharpness of the bands. The samples were separated in native 12% polyacrylamide TBE gels at 4°C.

Homology modelling

The human $DDX21_{FL}$ sequence was submitted to the online platform SwissModel in order to search for structural templates. Hits were evaluated for good overall sequence similarity and sequence coverage. The main hits were the bacterial DEAD-box helicases Hera (PDB IDs: 4KBF, 3I32) and CshA (PDB ID: 5IVL). For creating templates for the HC and DD domains (residues 186-621), Hera was chosen with 43% sequence identity over this range. With the individual domain models in hand, four models were created (**Figure S17**)), combining the two open conformations of the helicase core observed in Hera (PDB ID: 4KBF, chains a and b) and two positions of the helicase core with respect to the dimerization domain: one creating an "inverted V-shaped" (non-flat) structure as seen in the CshA crystal structure (PDB ID: 5IVL) and another where these domains are coplanar with the DD (flat conformation) (Core2-DD-RBD from Hera PDB ID: 3I32 and Core1 position from 4KBF). The NMR structure of the GUCT

domain (PDB ID: 2M3D) was also used to create these four models that cover a total of 524 residues (186-710). For fitting the models to the SAXS scattering curves, the N- and/or C-terminal unstructured tags were created using Pymol or I-TASSER (Yang et al., 2015). During revision, the crystal structures of the DDX21 HC were published (Chen et al., 2020) and these were used to modify the existing models. We superimposed each RecA-like domain separately into the $DDX21_{Core\text{-}DD}$ and $DDX21_{\Delta NC}$ models (RMSD of D1 (residues 188-397) and D2 (residues 410-563) RecA-like domains with those in the $DDX21_{Core\text{-}DD}$ was ~4 and with those in the $DDX21_{\Delta NC}$ was ~3.5) to place them in the optimised position fitting the SAXS data and then merge them to the DD/GUCT part of the models. Residues 529-533 were slightly remodelled to avoid clashes. Models were assessed and constructed using Coot (Emsley and Cowtan, 2004), PyMol (Schrödinger, LLC) and Chimera (Pettersen et al., 2004). Sequence alignment figures were made using Clustal Omega (Madeira et al., 2019) and Espript (Robert and Gouet, 2014).

Small-Angle X-Ray Scattering data collection and analysis

The data were collected at the European Synchrotron Radiation Facility (ESRF), beamline 29 (BM29) at a wavelength of 0.99 Å with a sample to detector distance of 2.867 m and a PILATUS 1M detector, covering a momentum transfer of $0.0025 < s > 0.5$ Å^{-1} [$s = 4\pi\sin(\theta)/\lambda$]. Measurements were made at 20°C. The proteins were analysed by size-exclusion chromatography in line with small-angle X-Ray scattering (SEC-SAXS) to avoid the signal from possible aggregates (with the exception of the $DDX21_{\Delta N}$ monomer mutant that was collected in batch mode). The samples were applied to a Superose 6 10/300 GL column (GE Healthcare) at a concentration of 9-12 mg/mL and run at a flow rate of 0.5 mL/min in 20 mM HEPES, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM TCEP. During the elution, 3000 scattering measurements were taken with 1s time-frames. The *in-house* software BsxCuBE (Biosaxs Customized Beamline Environment) connected to a data processing pipeline (EDNA) (Incardona et al., 2009) was used to control the real time data display (two dimensional and one dimensional) and to provide the first automatic data processing up to a preliminary *ab initio* model. SAXS data were analysed using the ATSAS package version 2.8.3 (Franke et al., 2017) and ScÅtter (www.bioisis.net). For each sample, using Chromixs (Panjkovich and Svergun, 2018), an elution profile was generated with the integrated intensities plotted versus the recorded frame number. Using Chromixs, ~30 buffer frames were averaged and used to (i) subtract the buffer average from each frame of the sample peak selected and (ii) calculate the corresponding Radius of Gyration (R_g). The subtracted peak region was selected in Chromixs and averaged to generate the final scattering curve used for subsequent analysis. The scattering curves were initially viewed in PRIMUS (Konarev et al., 2003) where the R_g was obtained from the slope of the Guinier plot within the region defined by $s_{min} < s < s_{max}$ where $s_{max} < 1.3/R_g$ and s_{min} is the lowest angle data point included by the program (**Figure S13**). The P(r) function, the distribution of the intra-atomic distances (r) in the particle, was generated using the indirect transform program GNOM (Svergun, 1992) using a scattering range of s_{min}/s_{max} of 0.0072/0.3792 (Fl), 0.0075/0.3787 (ΔN), 0.0117/0.3788 (ΔNC), 0.0098/0.3787 (Core-DD) and 0.0135/0.3801 (ΔN monomer). The maximum distance (D_{max}) was selected by letting the P(r) curve decay smoothly to zero (**Table S3**). The R_g was also estimated from the P(r) function, such that, unlike the Guinier R_g estimation, the P(r) R_g calculation takes a larger scattering range into account. DATPOROD, DATMOW and DATVC within the ATSAS package (Fischer et al., 2010; Franke et al., 2017; Rambo and Tainer, 2013) were used to estimate the Porod Volume (Vp) and the concentration-independent estimate of the MW for the proteins (**Table S3**). The final figures were generated using PyMOL (Schrödinger, LLC) and Chimera (Pettersen et al., 2004).

To better address the flexibility of DDX21, we performed two types of analysis: EOM 2.1 analysis (ensemble optimization method) (Bernadó et al., 2007; Tria et al., 2015) (**Figure S16**) and nonlinear Cartesian Normal Mode Analysis (NOLB NMA) (Hoffmann and Grudinin, 2017) (**Figures 5** and **S17**). In the EOM procedure, a pool of 10,000 random structures of the $DDX21_{\Delta NC}$ dimer was generated by creating flexible linkers between the domains provided. These were the DD, that was kept fixed as the dimer interface and the HC and the GUCT domains, as three independent domains. Then, 100 cycles of genetic algorithm were run using all the data range. The genetic algorithm optimized the size of the ensemble and the fraction of occupancy of each conformer in the ensemble by minimizing the discrepancy between the calculated scattering curve and the experiential data (χ^2).

In the NOLB NMA procedure, the flat and “inverted V-shaped” $DDX21_{\Delta NC}$ models described above were flexibly fitted to the SAXS data through a χ^2 –minimizing optimization procedure based on the nonlinear

Cartesian NMA method called NOLB (Hoffmann and Grudinin, 2017) and a novel SAXS profile calculator called Pepsi-SAXS (Grudinin et al., 2017). More precisely, for each initial model we performed 100 optimization iterations. Each iteration comprised the computation of 60 slowest normal modes (using the NOLB tool), nonlinear structure deformation along these modes, and choosing the deformation with the least χ^2 value to the experimental scattering profile (using Pepsi-SAXS). A steepest-descent minimization algorithm was used at the end of each iteration to keep the local topology (bonds and angles) in agreement with the initial structure. The flexible fitting method is available as a standalone executable called Pepsi-SAXS-NMA at (https://team.inria.fr/nano-d/software/pepsi-saxs/). The resulting structures converged to low values of χ^2 (**Figures 5 and S17**).

Supplemental References

Ariyo, E.O., Booy, E.P., Patel, T.R., Dzananovic, E., McRae, E.K., Meier, M., McEleney, K., Stetefeld, J., and McKenna, S.A. (2015). Biophysical Characterization of G-Quadruplex Recognition in the PITX1 mRNA by the Specificity Domain of the Helicase RHAU. PLoS ONE *10*, e0144510.

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. Journal of the American Chemical Society *129*, 5656–5664.

Chen, Z., Li, Z., Hu, X., Xie, F., Kuang, S., Zhan, B., Gao, W., Chen, X., Gao, S., Li, Y., et al. (2020). Structural Basis of Human Helicase DDX21 in RNA Binding, Unwinding, and Antiviral Signal Activation. Adv Sci (Weinh) *7*, 2000532.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389-394.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Fischer, H., Oliveira Neto, M. de, Napolitano, H.B., Polikarpov, I., and Craievich, A.F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J Appl Cryst *43*, 101–109.

Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Cryst *50*, 1212–1225.

Grudinin, S., Garkavenko, M., and Kazennov, A. (2017). Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. Acta Crystallogr D Struct Biol *73*, 449–464.

Hardin, J.W., Hu, Y.X., and McKay, D.B. (2010). Structure of the RNA binding domain of a DEAD-box helicase bound to its ribosomal RNA target reveals a novel mode of recognition by an RNA recognition motif. J. Mol. Biol. *402*, 412–427.

Hoffmann, A., and Grudinin, S. (2017). NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method. J Chem Theory Comput *13*, 2123–2134.

Huen, J., Lin, C.-L., Golzarroshan, B., Yi, W.-L., Yang, W.-Z., and Yuan, H.S. (2017). Structural Insights into a Unique Dimeric DEAD-Box Helicase CshA that Promotes RNA Decay. Structure *25*, 469–481.

Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009). EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J Synchrotron Rad, J Synchrotron Radiat *16*, 872–879.

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.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Cryst, J Appl Crystallogr *36*, 1277–1282.

Laue, T., Shaw, B. D., Ridgeway, T. M., and Pelletier, S. L (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science. (The Royal Society of Chemistry, Cambridge, U.K.), pp. 90–125.

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

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641.

McRae, E.K.S., Booy, E.P., Moya-Torres, A., Ezzati, P., Stetefeld, J., and McKenna, S.A. (2017). Human DDX21 binds and unwinds RNA guanine quadruplexes. Nucleic Acids Res. *45*, 6656–6668.

Mergny, J.-L., Li, J., Lacroix, L., Amrane, S., and Chaires, J.B. (2005). Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res. *33*, e138.

Ohnishi, S., Pääkkönen, K., Koshiba, S., Tochio, N., Sato, M., Kobayashi, N., Harada, T., Watanabe, S., Muto, Y., Güntert, P., et al. (2009). Solution structure of the GUCT domain from human RNA helicase II/Gu beta reveals the RRM fold, but implausible RNA interactions. Proteins *74*, 133–144.

Panjkovich, A., and Svergun, D.I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics *34*, 1944–1946.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605–1612.

Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477–481.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320-324.

Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev. *31*, 1370–1381.

Steimer, L., Wurm, J.P., Linden, M.H., Rudolph, M.G., Wöhnert, J., and Klostermeier, D. (2013). Recognition of two distinct elements in the RNA substrate by the RNA-binding domain of the T. thermophilus DEAD box helicase Hera. Nucleic Acids Res. *41*, 6259–6272.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Cryst, J Appl Crystallogr *25*, 495–503.

Tria, G., Mertens, H.D.T., Kachala, M., and Svergun, D.I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ *2*, 207–217.

Valdez, B.C., Perlaky, L., and Henning, D. (2002a). Expression, cellular localization, and enzymatic activities of RNA helicase II/Gu(beta). Exp. Cell Res. *276*, 249–263.

Valdez, B.C., Yang, H., Hong, E., and Sequitin, A.M. (2002b). Genomic structure of newly identified paralogue of RNA helicase II/Gu: detection of pseudogenes and multiple alternatively spliced mRNAs. Gene *284*, 53–61.

Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R.A., Rouault, T.A., and Schuck, P. (2004). Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal. Biochem. *326*, 234–256.

Xu, L., Wang, L., Peng, J., Li, F., Wu, L., Zhang, B., Lv, M., Zhang, J., Gong, Q., Zhang, R., et al. (2017). Insights into the Structure of Dimeric RNA Helicase CsdA and Indispensable Role of Its C-Terminal Regions. Structure *25*, 1795-1808.e5.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7–8.