HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 April 04.

Published in final edited form as:

Nature. 2012 October 4; 490(7418): 121–125. doi:10.1038/nature11402.

Structural basis for RNA duplex recognition and unwinding by the DEAD-box helicase Mss116p

Anna L. Mallam, **Mark Del Campo**[1], **Benjamin Gilman**, **David J. Sidote**, and **Alan M. Lambowitz**[*]

Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712, USA

Abstract

DEAD-box proteins are the largest family of nucleic acid helicases and are crucial to RNA metabolism throughout all domains of life[1,2]. They contain a conserved 'helicase core' of two RecA-like domains (domains 1 and 2; D1 and D2, respectively), which uses ATP to catalyze the unwinding of short RNA duplexes by nonprocessive, local strand separation[3]. This mode of action differs from that of translocating helicases and allows DEAD-box proteins to remodel large RNAs and RNA-protein complexes without globally disrupting RNA structure[4]. However, the structural basis for this distinctive mode of RNA-unwinding remains unclear. Here, structural, biochemical, and genetic analyses of the yeast DEAD-box protein Mss116p indicate that the helicase core domains have modular functions that enable a novel mechanism for RNA duplex recognition and unwinding. By investigating D1 and D2 individually and together, we find that D1 acts as an ATP-binding domain and D2 functions as an RNA-duplex recognition domain. D2 contains a nucleic acid-binding pocket that is formed by conserved DEAD-box protein sequence motifs and accommodates A-form but not B-form duplexes, providing a basis for RNA substrate specificity. Upon a conformational change in which the two core domains join to form a 'closed-state' with an ATPase active site, conserved motifs in D1 promote the unwinding of duplex substrates bound to D2 by excluding one RNA strand and bending the other. Our results provide a comprehensive structural model for how DEAD-box proteins recognize and unwind RNA duplexes. This model explains key features of DEAD-box protein function and affords new perspective on how the evolutionarily related cores of other RNA and DNA helicases diverged to use different mechanisms.

[*]Corresponding author: lambowitz@austin.utexas.edu.

[1]Present address: Rigaku Americas Corporation, 9009 New Trails Drive, The Woodlands, TX 77381, USA

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The PDB accession codes for coordinates and structure factors are 4DB2 (D2-dsRNA) and 4DB4 (D2-dsRNA-DNA). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature.

Author Contributions

This study was designed by A.L.M. and A.M.L. A.L.M. cloned the individual domain constructs, purified proteins, crystallized the complexes, collected X-ray crystallographic data, and performed binding assays. A.L.M., M.D.C. and D.J.S. processed and refined the X-ray diffraction data. B.G. performed genetic assays. All authors contributed to analyzing the results. A.L.M., M.D.C. and A.M.L. wrote the paper, with contributions from D.J.S. and B.G.

Mss116p is a DEAD-box RNA helicase that facilitates the folding and splicing of mitochondrial group I and group II introns primarily by acting as an RNA chaperone that unwinds RNA duplexes to disrupt stable but inactive RNA structures[5–7]. The RecA-like helicase core domains of Mss116p, which together catalyze RNA unwinding[8,9], contain conserved DEAD-box protein sequence motifs that are required for helicase function (Fig. 1a). D2 also includes a non-conserved C-terminal extension (CTE) that stabilizes the domain and extends its RNA-binding surface[10]. Small angle X-ray scattering studies show that without substrates, the helicase core of Mss116p adopts an extended 'open state' conformation, as observed for other DEAD-box proteins[11,12]. A compact 'closed state', whose X-ray crystal structure has been determined for Mss116p and other DEAD-box proteins[10,13], is formed upon binding ATP and single-stranded RNA (ssRNA) and is thought to represent a 'post-unwound' state of the enzyme.

The wide separation of D1 and D2 of Mss116p in the open state suggests that they might function independently to recognize ATP and RNA substrates. To investigate the roles of the individual helicase core domains, we compared the ATP and double-stranded RNA (dsRNA) binding properties of the full core (D1D2) and isolated D1 and D2 of Mss116p. Gel filtration and ATP-agarose binding assays show that ATP binds to D1 and the full core with similar affinities, but does not bind significantly to D2 (Fig. 1b and Supplementary Fig. 1). This result is consistent with previous studies that establish D1 of DEAD-box proteins as a conserved ATP-binding domain[14,15]. Conversely, fluorescence anisotropy and electrophoretic mobility shift assays (EMSA) show that a 14-bp dsRNA binds to D2 and the full core with similar affinities, but does not bind significantly to D1 (Fig. 1c and Supplementary Fig. 2). The preferential binding of ATP by D1 and dsRNA by D2 with similar affinities to the full core supports the hypothesis that these helicase domains function independently in initial substrate capture in the open state of Mss116p.

We next determined a crystal structure of Mss116p D2 in complex with the same 14-bp dsRNA used in the binding assays (Fig. 2a) at 3.2 Å resolution, the first structure of a DEAD-box helicase domain bound to a duplex substrate (Fig. 2c–e, Supplementary Fig. 3 and Supplementary Table 1). The crystallographic asymmetric unit contains four very similar complexes with protein molecules bound on either side of a pseudo-continuous RNA duplex (Supplementary Fig. 3a). The structure of a single complex shows that D2 contains a positively charged binding pocket for an RNA duplex of A-form geometry (Fig. 2c–e and Supplementary Fig. 3b). One duplex strand (strand 1; yellow) interacts extensively with D2 (Fig. 3a, b and Supplementary Table 2). These interactions include multiple contacts to the phosphate groups of the three centrally bound nucleotide residues (N4-N6) by DEAD-box motifs IV, IVa, V and a loop containing motif Va. The second strand (strand 2; purple) makes only a few contacts, which include H-bonds between 2′-OH groups and the CTE (Fig. 3a, c and Supplementary Table 2). No protein contacts are observed to the RNA bases of either strand (Fig. 3), consistent with the non-specific RNA binding exhibited by Mss116p and other DEAD-box proteins[1]. Except for the interaction by motif Va, the structure of D2 and its contacts to the phosphate backbone of strand 1 are the same as in the closed-state structure of Mss116p, in which D2 additionally interacts with D1 and adenosine nucleotide (Fig. 3a and Supplementary Fig. 4; r.m.s.d. = 0.46 Å)[10]. Given the similar binding affinities

observed for dsRNA by isolated D2 and the full helicase core (Fig. 1c and Supplementary Fig. 2), we propose that the D2-dsRNA structure provides a model for the initial complex of Mss116p with duplex RNA in the open state of the enzyme and a structural basis for dsRNA recognition by Mss116p.

We also determined a structure of D2 in complex with an equivalent 14-bp chimeric RNA-DNA duplex (Fig. 2b) at 3.6-Å resolution, in which each molecule of protein interacts with two A-form duplex substrates (Fig. 2f–h, Supplementary Fig. 5 and Supplementary Table 1). These chimeric duplexes make almost identical contacts with D2 as dsRNA (Supplementary Fig. 5b). Surprisingly, however, a DNA segment of strand 1 interacts with the DEAD-box motifs in the main "RNA"-binding tract of D2, while an RNA segment binds to the CTE (Fig. 2f–h and Supplementary Fig. 5). This orientation is likely favored because D2 interacts primarily with nucleic acid substrate phosphate groups, while the non-conserved CTE makes H-bond contacts with the 2′-OHs of RNA (see above). The finding that A-form DNA and RNA interact in a similar manner in the conserved RNA-binding tract of D2 suggests that the substrate specificity of DEAD-box proteins for RNA duplexes is dictated primarily by phosphate backbone geometry. Consistent with this idea, modeling demonstrates that the binding pocket of D2 is not shaped to recognize a B-form DNA duplex, the predominant conformation of dsDNA (Supplementary Fig. 6). Additionally, a genetic assay that is stringently dependent upon Mss116p function indicates that the side chains of conserved residues of D2 that interact with phosphate groups (R415, motif IVa and T433, motif V) are critical for Mss116p function *in vivo*[9], while the side chains of residues in the CTE that interact with 2′-OH groups of the RNA (S535, R538) can be mutated without detectable loss-of-function (Supplementary Fig. 7). Nucleic acid recognition based upon duplex geometry can explain why DEAD-box proteins can unwind chimeric RNA-DNA duplexes with as few as two centrally located ribonucleotides[3], as structural studies indicate that chimeric duplexes with only one ribonucleotide can adopt A-form geometry[16,17].

Our D2 structures provide insight into several other DEAD-box protein activities. DEAD-box proteins function on a wide variety of RNA substrates[1,2]. A comparison of the D2-dsRNA and D2-dsRNA-DNA structures, and different complexes within their asymmetric units, shows that the binding orientation of the distal regions of duplex substrates can vary, while the contacts between the conserved motifs and centrally bound nucleotide residues are maintained (Supplementary Fig. 8). Although these differences in substrate binding orientation could be influenced by crystal-packing, the observed flexibility in nucleic-acid binding away from the basic binding tract of D2 may be advantageous for the loading and unwinding of diverse physiological RNA substrates, such as group I and group II introns, and could contribute towards the general RNA-chaperone activity of Mss116p[5,6]. The presence of a dsRNA binding pocket in D2 also raises the possibility that this domain could play a role in strand annealing in the absence of ATP, an activity observed for Mss116p and other DEAD-box proteins[1], by orienting two ssRNAs in a position to pair in the duplex binding pocket. The additional RNA interactions with the CTE of Mss116p may explain the relatively high strand annealing activity of Mss116p compared to other DEAD-box proteins[18].

Collectively, our results indicate that RNA unwinding by Mss116p begins with the helicase core domains functioning to independently bind ATP and RNA substrates (Fig. 4a). This previously unobserved mechanism for substrate recognition by a helicase is consistent with the wide separation of the two core domains (~ 50 Å for their centers-of-mass) in the solution structure of the open state[11]. Subsequent interactions of the exposed regions of these substrates with the remainder of their binding sites in the opposite domain would result in cooperative tight binding coupled to core closure, RNA-strand separation, and formation of the ATPase active site (Fig. 4a). Notably, the loop in D2 that contains part of motif Va and interacts with duplex RNA in the D2-dsRNA structure (see above) shifts markedly upon formation of the closed state and helps form the ATPase active site (Fig. 4b). This conformational change may be part of a switch that triggers ATP hydrolysis upon core closure. Following ATP hydrolysis, dissociation of P_i and ADP cause reopening of the core, release of the bound strand, and regeneration of the enzyme[8,19].

Comparison of the D2-dsRNA structure to the previously reported closed-state structure of the helicase core of Mss116p bound to adenosine nucleotide and ssRNA (Fig. 4c) indicates that D2 functions as a stationary platform that positions dsRNA for unwinding by the incursion of D1 (Fig. 4d,e and Supplementary Fig. 9). In the closed-state of Mss116p, RNA-binding motifs Ia, Ib, Ic, and the post-II region of D1 are sterically incompatible with dsRNA bound in the duplex binding site of D2 (Fig. 4d). This suggests that D1 promotes RNA unwinding in two ways (Fig. 4e and Supplementary Fig. 9): first, the conserved post-II region of D1 interrupts the centrally bound base pairs of the RNA duplex to displace strand 2. This displacement is presumably facilitated by the minimal interaction of strand 2 with the protein and could occur actively or during ‘breathing’ of the duplex[20]. Second, core closure introduces two bends in strand 1, one by interactions with the conserved ‘wedge helix’ in D1 (motif Ic) and the other by interactions with the CTE[9,10]. DEAD-box proteins that unwind RNA but lack the CTE introduce only the first bend using the same conserved D1 wedge helix[13]. Upon core closure, the buried solvent accessible surface area of strand 1 increases due to additional interactions with D1 (1256 $Å^2$ compared to 544 $Å^2$ in the D2-dsRNA structure; Supplementary Table 2), contributing to the driving force for RNA unwinding. Further, the two bends induced in strand 1 (Supplementary Fig. 9b) impede its reannealing to strand 2. Additional conformational changes that occur upon ATP hydrolysis, dissociation of P_i, and/or reopening of the core may also contribute to RNA unwinding.

The model we propose here for RNA duplex recognition and unwinding by Mss116p readily explains the previously reported requirement for ATP binding, but not hydrolysis, for RNA unwinding by DEAD-box proteins[21,22]. This is because substrate binding drives core closure and strand separation (Fig. 4a), while ATP hydrolysis to regenerate the enzyme cannot occur until formation of the ATPase active site in the closed-state core. The model also elucidates why RNA-unwinding by DEAD-box proteins is nonprocessive and can initiate directly from a double-stranded region of substrate[1,2], as RNA duplexes are bound directly by D2. Additionally, the differences in RNA-unwinding activity observed for different ATP analogs[22] could reflect differences in their binding affinity for the closed state.

Because the RNA-unwinding mechanism described here for Mss116p depends primarily upon conserved DEAD-box protein structures and motifs with only an ancillary role for the

CTE, we hypothesize that its major features are used by all DEAD-box proteins. All DEAD-box proteins rely on their helicase core for RNA unwinding and use appended domains for auxiliary functions, such as interactions with partner proteins or to target the helicase core to specific RNA substrates[1]. Structural studies of other DEAD-box proteins show that D1 can by itself bind adenosine nucleotide in a binding pocket formed by the Q-motif, which recognizes the adenine base, motif I (the phosphate-binding or P-loop) and motif II[14,15]. By contrast, structures of Mss116p and other DEAD-box proteins in the closed state show that D2 interacts minimally with the adenine base[1,10,13], in agreement with our observation that D2 does not by itself bind specifically to ATP (Fig. 1b and Supplementary Fig. 1). Likewise, all DEAD-box proteins contain a conserved RNA-binding track in D2 that contains motifs IV, IVa, and V[1] and could recognize dsRNA similarly to Mss116p. By contrast, dsRNA is sterically incompatible with the motif Ic wedge helix and the post-II region in the RNA-binding track of D1 of other DEAD-box proteins[10,13,23], in agreement with our finding that D1 of Mss116p cannot by itself bind an RNA duplex (Fig. 1c, Supplementary Fig. 2). The post-II region, which displaces strand 2 of a bound duplex in the unwinding mechanism proposed for Mss116p (Fig. 4d, e), is conserved and positioned to play the same role in other DEAD-box proteins[9,23], as is motif Va[1,15], which forms part of loop that contributes to initial binding of the duplex RNA and rearranges to help form the ATPase active site in the closed state (Fig. 4b).

Non-ring forming helicases with structurally conserved cores of RecA-like domains D1 and D2 are classified into two superfamilies (SFs), SF1 and SF2, with DEAD-box proteins comprising the largest family of SF2[24,25]. These helicases are thought to have evolved from a common ancestor[26], but have diverged to possess a variety of accessory domains, to have different specificities for RNA or DNA substrates, and to operate by distinct mechanisms[24,25]. These include processive and non-processive duplex unwinding and translocation without unwinding. The modular substrate binding functions of D1 and D2, duplex binding by D2, and substrate specificity based on nucleic acid geometry found here for Mss116p may be features that underlie these diverse mechanisms. Interestingly, two recent crystal structures of the closed state of the pathogen recognition receptor RIG-I, an SF2 helicase closely related to DEAD-box proteins, show dsRNA bound in the RNA-binding track of D2 of RIG-I in the same orientation as D2 of Mss116p and interacting with conserved SF2 helicase RNA-binding motifs including IV, IVa and V (Supplementary Fig. 10a, c)[27,28]. However, the orientation of D1 in the RIG-I complex differs from that of Mss116p in the closed state (Supplementary Fig. 10b), enabling RIG-I to bind and translocate on a duplex substrate without RNA unwinding[27,28]. RIG-I also contains ancillary domains that contribute to dsRNA binding and may influence substrate orientation to favor duplex binding over unwinding[27–29]. Other SF1 and SF2 helicases could have evolved similarly to promote different closed-state conformations of their helicase cores that give rise to distinct mechanisms of action.

Full Methods

Oligonucleotides

The RNA and RNA-DNA oligonucleotides rGrGrGrCrGrGrGrCrCrCrGrCrCrC and rGrGrGrCrGrGrGdCdCdCdGdCdCdC (Integrated DNA Technologies) were annealed to form 14-bp RNA or chimeric RNA-DNA duplexes by heating 6 mM solutions in 100 mM potassium acetate, 30 mM HEPES (pH 7.5) at 94°C for 1 min and then slowly cooling to room temperature over 1 h.

Protein expression and purification

pMAL-Mss116p contains the coding sequence for Mss116p (codons 37–664) with an in-frame N-terminal MalE fusion cloned downstream of a *tac* promoter in the expression vector pMAL-c2t (a derivative of pMAL-c2x; New England Biolabs)[18]. Mss116p/D1D2 is a derivative of pMAL-Mss116p that expresses the active helicase core of Mss116p (residues 88–597) with deletions of an unstructured N-terminal extension and C-terminal tail[11]. Expression vectors for the Mss116p constructs D1 (residues 88–330) and D2 (residues 342–597) were created by PCR of pMAL-Mss116p with primers that introduce BamHI and HindIII sites at the 5′ and 3′ end of the desired gene segment, and then cloning the PCR product between the corresponding sites of pMAL-c2t to link the protein-coding sequence to that of the MalE tag.

Mss116p D1D2, D1 and D2 were expressed as N-terminal MalE fusions in *Escherichia coli* Rosetta 2 (EMD Biosciences), grown in ZYP-5052 auto-inducing medium for 24 h at 22 °C, and purified at 4 °C, as described[10,11]. Purification steps included: (i) removal of nucleic acids by polyethyleneimine precipitation; (ii) isolation of MBP-Mss116p by amylose-affinity chromatography (New England Biolabs); (iii) removal of the MBP tag by digestion with tobacco etch virus protease; (iv) isolation of Mss116p by heparin-Sepharose chromatography (GE Healthcare); and (v) purification and buffer exchange by gel filtration chromatography using a Superdex S200 column (GE Healthcare). MBP-tagged proteins were purified in an identical manner but without step (iii). Proteins were concentrated for crystallization by using a 10-kDa MWCO concentrator (Millipore), and protein concentrations were determined by Bradford Assay (Bio-Rad). Crystallization and storage buffers were 20 mM Tris-HCl (pH 7.5), 200 mM KCl, 10% glycerol, 1 mM dithiothreitol (DTT).

Crystallization

For the D2-dsRNA complex, the protein (~500μM) was incubated with dsRNA (650 μM duplex) and $MgCl_2$ (2 mM) for 30 min on the desktop. Hanging drops were assembled using 1 μl of complex and 1 μl of a well solution of 8% tacsimate (pH 6.0; Hampton Research), 20% PEG 3350. Drops were stored at 22 °C and plate-like crystals appeared within 1 week. Crystals were stabilized in a cryoprotectant containing the crystallization solution plus 20% glycerol before flash cooling in liquid N_2. Crystals of D2-dsRNA-DNA were obtained similarly with hanging drops assembled from 1 μl of dsRNA-DNA complex and 1 μl of a well solution of 6% tacsimate (pH 5.0; Hampton Research), 20% PEG 3350 and were cryoprotected as above.

Structure Determination

X-ray diffraction data were collected either on our in-house system (Rigaku MicroMax-007 HF generator with VariMax HF optics and an R-AXIS IV++ imaging plate detector; wavelength = 1.54178 Å) or at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory (mail-in service on beamlines 5.0.2, or 5.0.3; wavelength = 1.00003 Å). Details of data collection and refinement are in Supplementary Table 1. Diffraction intensities were indexed, integrated, and scaled with HKL-2000[31] or autoPROC[32]. For diffraction data processed with HKL-2000, additional statistics were calculated from unmerged data using d*TREK[33]. Initial space groups were determined by using Pointless[34] and confirmed by decreases in both R_{work} and R_{free} after refinement of molecular replacement solutions. Molecular replacement was performed with Phaser[35], using the previously determined structure of Mss116p D2 in the closed state (PDB 3I5X) as a search model. Composite omit maps were calculated to determine that there was no model bias[36]. Structures were completed with cycles of manual model building in Coot[37] and refinement in Phenix[36]. Validation of protein and nucleic acid models and their contacts was done by using MolProbity[38] and indicated that at least 96% of residues are located in the most favorable region of the Ramachandran plot. Structural figures were prepared by using the PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC. SASA calculations and interface analyses were performed using PDBePISA[39].

ATP-binding assays

Equilibrium binding of ATP to D1, D2 and D1D2 was measured by gel filtration and ATP-agarose binding assays. Because the helicase core of Mss116p does not contain tryptophan residues, its calculated extinction coefficient is small (ε_{280} = 18,255 M^{-1} cm^{-1}; ExPASy Proteomics Server ProtParam tool[40]). The binding of ATP (ε_{260} ~ 15,400 M^{-1} cm^{-1}) therefore gives rise to a large change in A_{260} compared to A_{280}. Protein samples (10 μM) were incubated at 22 °C for 30 min in increasing concentrations of ATP-Mg^{2+} (0–200 μM) and loaded onto a HiTrap desalting column (GE Healthcare) pre-equilibrated in a buffer containing the same amount of ATP-Mg^{2+} and 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$. The absorbance of the eluted protein above the background signal of the buffer was measured at 260 and 280 nm, and the change in A_{260}/A_{280} signal at increasing concentrations of ATP was fit to a simple one-site ligand-binding model to calculate a K_d for the protein-ATP complex.

ATP-agarose binding assays were performed in 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$ by incubating AP-ATP-agarose (150 μl; Jena Bioscience) with ~50 μg of protein overnight at 4 °C with agitation. The beads were pelleted by centrifugation (1,000 x g, 1 min), the supernatant was removed, and the pellet was resuspended in binding buffer (150 μl). The load, supernatant, and ATP-agarose pellet were analyzed by SDS-PAGE, and the gels were stained with SYPRO Ruby protein gel stain (Invitrogen) to detect and quantify bound and unbound Mss1116p.

RNA-binding assays

Equilibrium binding of dsRNA to D1, D2 and D1D2 was measured by fluorescence anisotropy and electrophoretic mobility shift assays (EMSA). Fluorescence anisotropy

measurements were performed using MBP-tagged proteins to increase the change in anisotropy signal upon binding. A 3′ FAM-labeled 14-bp RNA duplex (10 nM; IDT; Fig. 2a) was incubated with increasing concentrations of protein (0 to 4 μM) at 22 °C for at least 1 h in a reaction medium containing 20 mM Tris-HCl (pH 7.5), 100 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$ and 0.1 mg/ml of bovine serum albumin to stabilize the protein at low concentrations. The observed fluorescence anisotropy, r_{obs}, of FAM-dsRNA at increasing concentrations of protein was measured by using an EnVision Microplate Reader (Perkin Elmer) and was fit to the equation:

$$r_{\mathrm{obs}}=r_b f_b+r_f(1-f_b)$$

where f_b is the fraction of protein-bound dsRNA and r_f and r_b are the anisotropy of the free and protein-bound dsRNA, respectively[41]. f_b was defined as $f_b = [RP]/[P_t]$, where $[RP]$ and $[P_t]$ are the concentration of protein-RNA complex and the total protein concentration, respectively. The dissociation constant, K_d, is defined by the quadratic equation[41]:

$$[RP]=\frac{(K_d+[P_t]+[R_t])-\sqrt{(K_d+[P_t]+[R_t])^2-4[P_t][R_t]}}{2}$$

where $[R_t]$ is the total concentration of dsRNA.

EMSA measurements were performed on both MBP-tagged proteins (to increase protein solubility under the experimental conditions) and untagged proteins. A 3′ FAM-labeled 14-bp RNA duplex (100 nM; IDT; Fig. 2a) was incubated with increasing concentrations of protein (0–3 μM) at 22 °C as described above. Samples were then analyzed in a nondenaturing 6% polyacrylamide gel run at 4 °C for 60 min, and the fluorescence signal of the bound duplex substrate was quantified on a Typhoon imager (GE Healthcare). Control gels verified that the fluorescence signal of the unbound substrate increased linearly with concentration (Supplementary Fig. 2e).

Genetic assays

Yeast genetic selections of functional Mss116p variants from centromere-containing (CEN) plasmid libraries in which specified codons were randomized and glycerol growth tests of individual variants were performed as described[9].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Art Monzingo (University of Texas at Austin Macromolecular Crystallography Facility) for help with X-ray diffraction data collection, and Rick Russell (University of Texas at Austin) and Eckhard Jankowsky (Case Western Reserve University) for comments on the manuscript. X-ray diffraction data were collected at the Berkeley Center for Structural Biology, which is supported in part by the NIH, NIGMS, and HHMI. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of

Energy under Contract No. DE-AC02–05CH11231. A.L.M. is the recipient of an EMBO Long-term Fellowship (ALTF 389–2010). This work was supported by NIH grant GM037951.

References

1. Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. Nat Rev Mol Cell Biol. 2011; 12:505–516. [PubMed: 21779027]
2. Jarmoskaite I, Russell R. DEAD box proteins as RNA helicases and chaperones. WIREs: RNA. 2011; 2:135–152. [PubMed: 21297876]
3. Yang Q, Del Campo M, Lambowitz AM, Jankowsky E. DEAD-box proteins unwind duplexes by local strand separation. Mol Cell. 2007; 28:253–263. [PubMed: 17964264]
4. Pan C, Russell R. Roles of DEAD-box proteins in RNA and RNP folding. RNA Biol. 2010; 7:667–676. [PubMed: 21045543]
5. Huang HR, et al. The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. Proc Natl Acad Sci USA. 2005; 102:163–168. [PubMed: 15618406]
6. Del Campo M, et al. Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. J Mol Biol. 2009; 389:674–693. [PubMed: 19393667]
7. Potratz JP, Del Campo M, Wolf RZ, Lambowitz AM, Russell R. ATP-dependent roles of the DEAD-box protein Mss116p in group II intron splicing *in vitro* and *in vivo*. J Mol Biol. 2011; 411:661–679. [PubMed: 21679717]
8. Cao W, et al. Mechanism of Mss116 ATPase reveals functional diversity of DEAD-box proteins. J Mol Biol. 2011; 409:399–414. [PubMed: 21501623]
9. Mohr G, et al. High-throughput genetic identification of functionally important regions of the yeast DEAD-box protein Mss116p. J Mol Biol. 2011; 413:952–972. [PubMed: 21945532]
10. Del Campo M, Lambowitz AM. Structure of the yeast DEAD-box protein Mss116p reveals two wedges that crimp RNA. Mol Cell. 2009; 35:598–609. [PubMed: 19748356]
11. Mallam AL, et al. Solution structures of DEAD-box RNA chaperones reveal conformational changes and nucleic acid tethering by a basic tail. Proc Natl Acad Sci USA. 2011; 108:12254–12259. [PubMed: 21746911]
12. Wang S, Overgaard MT, Hu Y, McKay DB. The *Bacillus subtilis* RNA helicase YxiN is distended in solution. Biophys J. 2008; 94:L01–03. [PubMed: 17951299]
13. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. Cell. 2006; 125:287–300. [PubMed: 16630817]
14. Rudolph MG, Heissmann R, Wittmann JG, Klostermeier D. Crystal structure and nucleotide binding of the *Thermus thermophilus* RNA helicase Hera N-terminal domain. J Mol Biol. 2006; 361:731–743. [PubMed: 16890241]
15. Schutz P, et al. Comparative structural analysis of human DEAD-box RNA helicases. PLoS One. 2010; 5:e12791. [PubMed: 20941364]
16. Egli M, Usman N, Rich A. Conformational influence of the ribose 2′-hydroxyl group - crystal structures of DNA-RNA chimeric duplexes. Biochemistry. 1993; 32:3221–3237. [PubMed: 7681688]
17. Wahl MC, Sundaralingam M. B-form to A-form conversion by a 3 ′-terminal ribose: crystal structure of the chimera d(CCACTAGTG)r(G). Nucleic Acids Res. 2000; 28:4356–4363. [PubMed: 11058136]
18. Halls C, et al. Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. J Mol Biol. 2007; 365:835–855. [PubMed: 17081564]
19. Henn A, et al. Pathway of ATP utilization and duplex rRNA unwinding by the DEAD-box helicase, DbpA. Proc Natl Acad Sci USA. 2010; 107:4046–4050. [PubMed: 20160110]
20. Rocak S, Linder P. DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol. 2004; 5:232–241. [PubMed: 14991003]

21. Chen Y, et al. DEAD-box proteins can completely separate an RNA duplex using a single ATP. Proc Natl Acad Sci USA. 2008; 105:20203–20208. [PubMed: 19088196]
22. Liu F, Putnam A, Jankowsky E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. Proc Natl Acad Sci USA. 2008; 105:20209–20214. [PubMed: 19088201]
23. Andersen CBF, et al. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science. 2006; 29:1968–1972. [PubMed: 16931718]
24. Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. Curr Opin Struct Biol. 2010; 20:313–324. [PubMed: 20456941]
25. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007; 76:23–50. [PubMed: 17506634]
26. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 1989; 17:4713–4730. [PubMed: 2546125]
27. Kowalinski E, et al. Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. Cell. 2011; 147:423–435. [PubMed: 22000019]
28. Jiang F, et al. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature. 2011; 479:423–427. [PubMed: 21947008]
29. Luo D, et al. Structural insights into RNA recognition by RIG-I. Cell. 2011; 147:409–422. [PubMed: 22000018]
30. Ozalp VC, Pedersen TR, Nielsen LJ, Olsen LF. Time-resolved measurements of intracellular ATP in the yeast *Saccharomyces cerevisiae* using a new type of nanobiosensor. J Biol Chem. 2010; 285:37579–37588. [PubMed: 20880841]
31. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276:307–326.
32. Vonrhein C, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr. 2011; 67:293–302. [PubMed: 21460447]
33. Pflugrath JW. The finer things in X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr. 1999; 55:1718–1725. [PubMed: 10531521]
34. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006; 62:72–82. [PubMed: 16369096]
35. McCoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
36. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
37. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
38. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66:12–21. [PubMed: 20057044]
39. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007; 372:774–797. [PubMed: 17681537]
40. Wilkins MR, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999; 112:531–552. [PubMed: 10027275]
41. Tang GQ, Bandwar RP, Patel SS. Extended upstream A-T sequence increases T7 promoter strength. J Biol Chem. 2005; 280:40707–40713. [PubMed: 16215231]

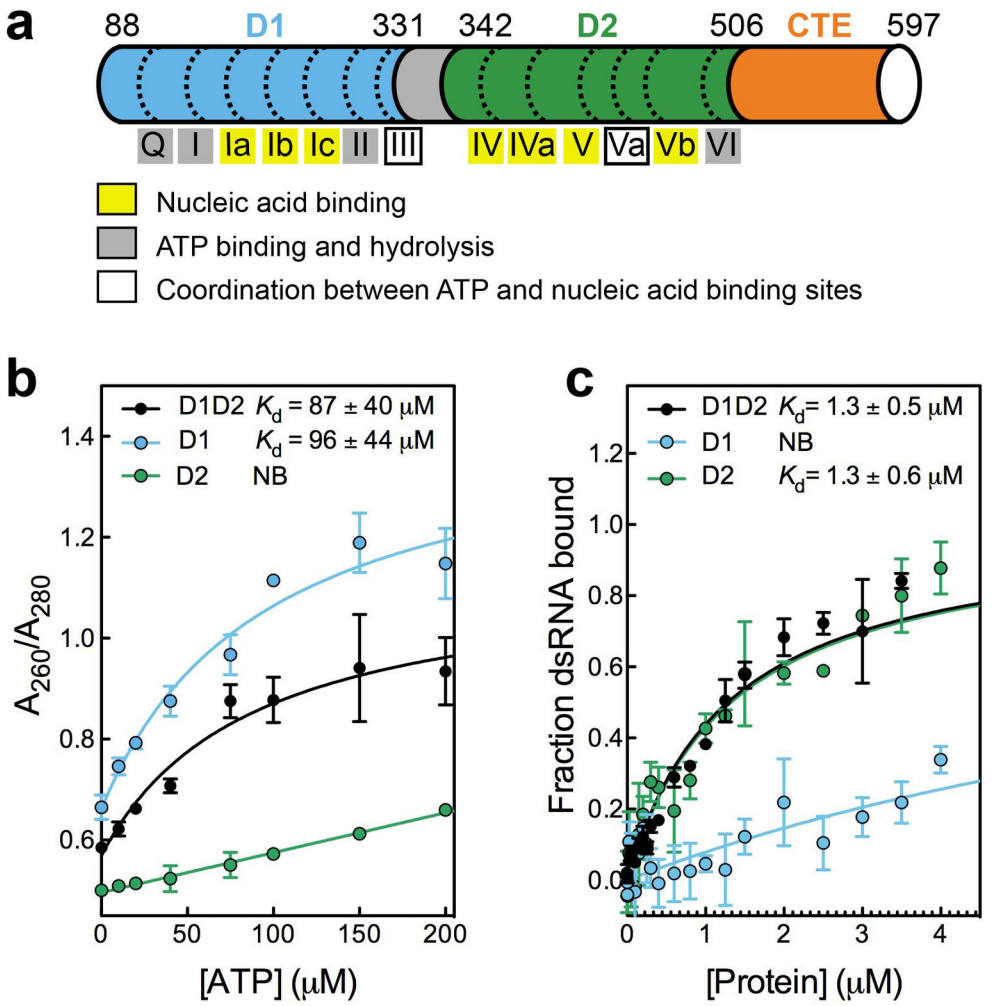

Figure 1. The distinct substrate binding characteristics of the helicase core domains of Mss116p

a, Schematic of the domain architecture of the helicase core of Mss116p (D1, blue; D2, green; C-terminal extension of D2 (CTE), orange), indicating conserved DEAD-box sequence motifs defined according to Fairman-Williams et al. (2010)[24]. Full-length Mss116p contains additional unstructured N-terminal (residues 37–87) and C-terminal (residues 598–664) regions that are not required for helicase activity[8,9]. **b,** Affinity of ATP for D1, D2, and the full helicase core (D1D2) measured by gel filtration chromatography under equilibrium conditions. ATP binding was also assessed by an ATP-agarose binding assay (Supplementary Fig. 1). **c,** Affinity of FAM-dsRNA (Fig. 2a) for MBP-tagged D1, D2 and D1D2 determined by fluorescence anisotropy measurements. Similar results for dsRNA binding were obtained by EMSA (Supplementary Fig. 2). Error bars in **b** and **c** represent the standard error for at least three independent measurements, and the error in the K_d represents the standard error of the non-linear regression (see Full Methods). NB indicates no significant binding.

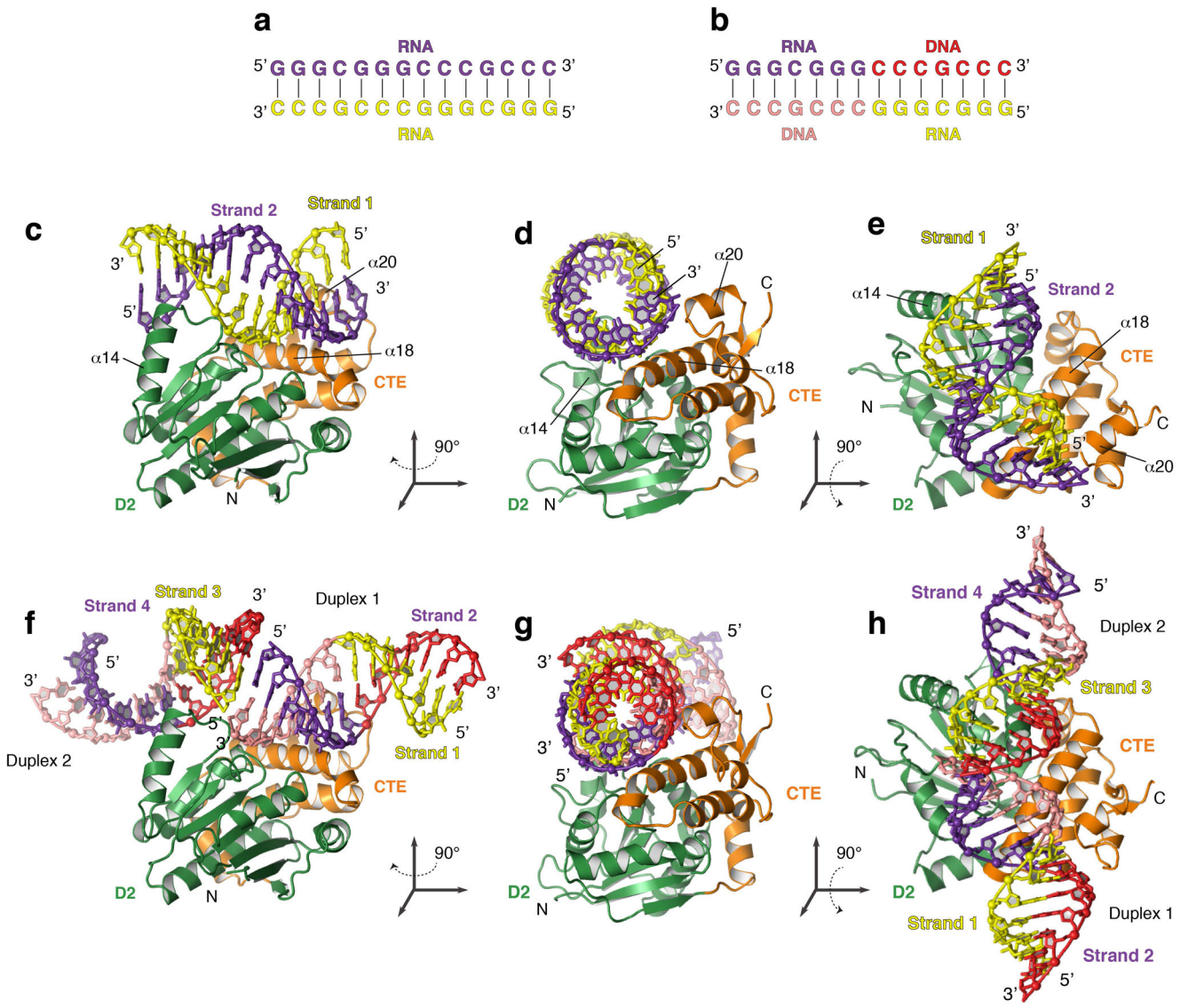

Figure 2. Crystal structures of Mss116p D2 bound to A-form duplexes

a, The 14-bp self-complementary GC-rich RNA duplex substrate. **b,** The 14-bp GC-rich chimeric RNA-DNA duplex substrate. **c**-**e**, Orthogonal views of the D2-dsRNA complex colored as in Fig. 1a and Fig. 2a. Helix α 14 of D2, which contains motif IVa, faces the major groove of the dsRNA, and α 18 and α 20 of the CTE face the minor groove of the dsRNA. **f**-**h**, Orthogonal views of the D2-dsRNA-DNA complex, colored as in Fig. 1a and Fig. 2b, in which D2 is bound to two stacked 14-bp chimeric RNA-DNA duplexes.

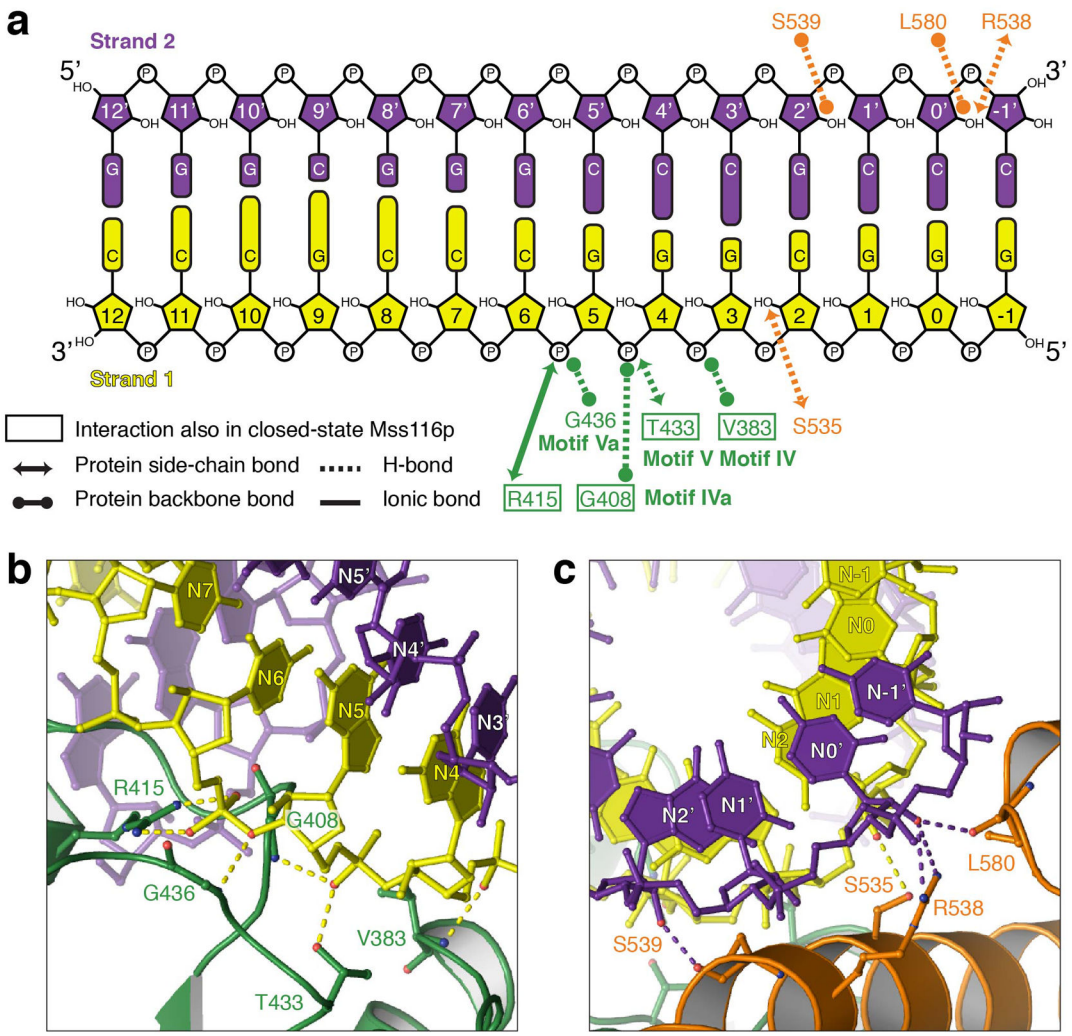

Figure 3. Interactions between Mss116p D2 and duplex RNA

a, Schematic of RNA-protein interactions observed in the D2-dsRNA structure. The dsRNA interacts with conserved DEAD-box motifs IV-Va of D2 (green) and the CTE of D2 (orange). The box indicates that the interaction is maintained in closed-state Mss116p[10]. RNA bases are numbered according to their position relative to ssRNA in closed-state Mss116p (see Supplementary Fig. 4d). Similar nucleic acid-protein interactions were observed in the D2-dsRNA-DNA structure (Supplementary Fig. 5). **b**, Interactions between strand 1 (yellow) of the duplex RNA and D2 (green). **c**, Interactions between duplex RNA and the CTE of D2 (orange).

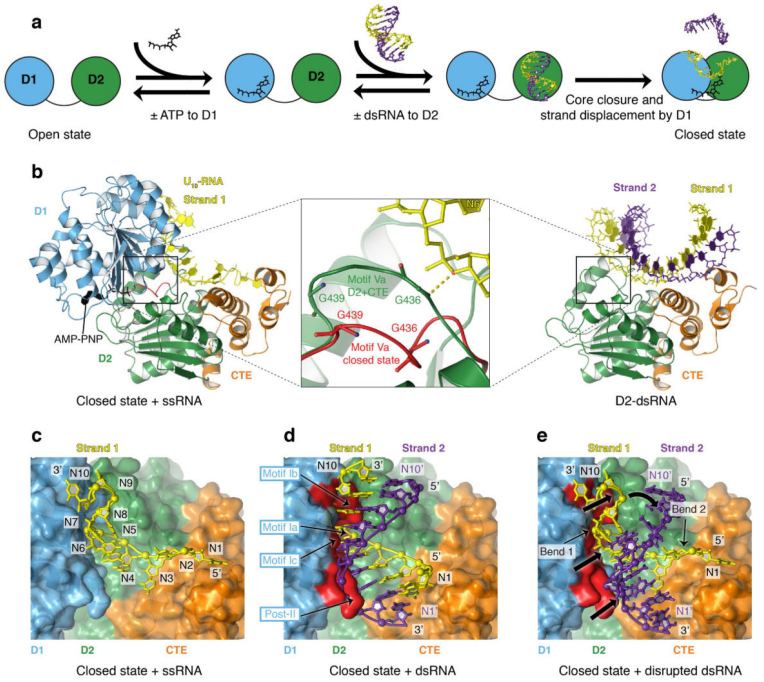

Figure 4. RNA duplex binding and unwinding by Mss116p

a, Model for the modular roles of the helicase core domains of Mss116p during RNA recognition and unwinding. Although ATP and duplex RNA could bind in either order, the binding of ATP to D1 is shown as a first step because Mss116p-ATP complexes are likely pre-populated at physiological concentrations of ATP (>1 mM[30]). **b**. Comparison of the position and interactions of the flexible motif Va loop (residues 435–440) in the D2-dsRNA and closed-state structures of Mss116p. The closed-state helicase core of Mss116p (PDB = 3I5X)[10] bound to ssRNA (U_{10}-RNA; yellow) and adenosine nucleotide (AMP-PNP; black) is shown with domains colored as in Fig. 1a. In the D2-dsRNA structure, the motif Va loop (green) interacts with strand 1 of dsRNA (yellow), whereas in the closed-state structure, the loop (red) shifts to a different position where motif Va helps to form the ATP-binding site. **c**, Surface representation of closed-state Mss116p with N1–N10 of U_{10}-RNA (yellow) indicated. **d**, Surface representation of closed-state Mss116p with dsRNA modeled in the duplex RNA-binding pocket of D2. Sterically incompatible regions of D1 are highlighted in red. **e**, Change in trajectory of strand 1 and predicted displacement of strand 2 of bound dsRNA by D1 upon core closure of Mss116p with arrows indicating regions of the substrate that are displaced.