Published in final edited form as:

J Mol Biol. 2008 February 1; 375(5): 1344–1364.

Function of the C-terminal Domain of the DEAD-Box Protein

Mss116p Analyzed *In Vivo* and *In Vitro*

Georg Mohr¹, Mark Del Campo¹, Sabine Mohr¹, Quansheng Yang², Huijue Jia², Eckhard Jankowsky², and Alan M. Lambowitz^{1,*}

¹Institute 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

²Department of Biochemistry and Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH 44106

Abstract

The DEAD-box proteins CYT-19 in *Neurospora crassa* and Mss116p in *Saccharomyces cerevisiae* are general RNA chaperones that function in splicing mitochondrial group I and group II introns and in translational activation. Both proteins consist of a conserved ATP-dependent RNA helicase core region linked to N- and C-terminal domains, the latter with a basic tail similar to many other DEAD-box proteins. In CYT-19, this basic tail was shown to contribute to non-specific RNA binding that helps tether the core helicase region to structured RNA substrates. Here, multiple sequence alignments and secondary structure predictions indicate that CYT-19 and Mss116p belong to distinct subgroups of DEAD-box proteins, whose C-terminal domains have a defining extended α -helical region preceding the basic tail. We find that mutations or C-terminal truncations in the predicted α -helical region of Mss116p strongly inhibit RNA-dependent ATPase activity, leading to loss of function in both translational activation and RNA splicing. These findings suggest that the α -helical region may stabilize and/or regulate the activity of the RNA helicase core. By contrast, a truncation that removes only the basic tail leaves high RNA-dependent ATPase activity and causes only a modest reduction in translation and RNA splicing efficiency *in vivo* and *in vitro*. Biochemical analysis shows that deletion of the basic tail leads to weaker non-specific binding of group I and group II intron RNAs, and surprisingly, also impairs RNA-unwinding at saturating protein concentrations and nucleotide-dependent tight binding of single-stranded RNAs by the RNA helicase core. Together, our results indicate that the two subregions of Mss116p's C-terminal domain act in different ways to support and modulate activities of the core helicase region, whose RNA-unwinding activity is critical for both the translation and RNA splicing functions.

Keywords

catalytic RNA; ribozyme; RNA chaperone; RNA-protein interaction; translation initiation

*Corresponding author: Phone: 512-232-3418, Fax: 512-232-3420, e-mail: lambowitz@mail.utexas.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

DExH/D-box proteins are an ubiquitous protein family, members of which mediate RNA structural rearrangements in a variety of cellular processes, including translation, RNA splicing, ribosome assembly, RNA degradation, and nuclear transport.¹ All DExH/D-box proteins have a conserved ATP-dependent RNA helicase core comprised of two tandem RecA-like domains.¹⁻³ This core region contains a series of at least eight conserved sequence motifs, with three protein subfamilies distinguished by the sequence of motif II (DEAD, DEAH, or DExH).^{4,5} Some viral DExH-box proteins are processive RNA helicases,⁶⁻⁸ while others, particularly members of the DEAD-box subfamily, are not processive and disrupt only short RNA duplexes.⁹ The latter proteins may function as RNA chaperones or RNP remodeling enzymes.⁹⁻¹⁵ In many DExH/D-box proteins, the core helicase region is linked to distinct N- and/or C-terminal domains, some of which have been shown to target the proteins to their sites of action via protein-RNA or protein-protein interactions.^{1,4,16-19} In most cases, however, little is known about the structure and function of these domains or how they interact with the core helicase region to contribute to the function of DExH/D-box proteins on their natural substrates.

The DEAD-box proteins CYT-19 in *Neurospora crassa* and Mss116p in *Saccharomyces cerevisiae* are general RNA chaperones that promote the splicing of mitochondrial (mt) group I and group II introns, other types of RNA processing reactions, and translation of mt mRNAs.^{14,20,21} As general RNA chaperones, these proteins bind multiple RNA or RNP substrates with little or no specificity and resolve stable non-native or intermediate structures that are kinetic traps in RNA folding. This general RNA chaperone function is suggested by the pleiotropic defects in mt RNA synthesis in *cyt-19* and *mss116* mutants,^{21,22} by the interchangeability of CYT-19 and Mss116p with each other and with other DEAD-box proteins, such as Ded1p,²³⁻²⁵ and by biochemical analysis of how CYT-19 and Mss116p affect group I and group II intron RNA splicing.^{14,15,23,24,26}

The RNA chaperone function of DEAD-box proteins correlates closely with their ATP-dependent RNA-unwinding activity.^{15,24} DEAD-box proteins may also unwind short helices (*e.g.*, 6-bp) in the absence of ATP, likely by binding selectively to transiently single-stranded regions, and such ATP hydrolysis-independent strand separation may contribute to RNA chaperoning, similar to the RNA chaperone activity of many non-specific RNA-binding proteins.¹⁵ Additionally, DEAD-box proteins have other activities, including strand annealing, strand exchange, non-specific RNA binding, and RNP remodeling, which may act in concert with their RNA-unwinding activity to promote reactions on natural substrates.^{12,24} Although it has been proposed that Mss116p can stimulate the splicing of the yeast *al5 γ* group II intron without unwinding RNA,²⁵ recent findings suggest this is not the case.¹⁵

The N- and/or C-terminal domains of CYT-19 and Mss116p likely contribute to their RNA chaperone function at least in part by helping to mediate interactions with RNA and RNP substrates. In the case of CYT-19, a C-terminal truncation of 49-amino acid residues had little effect on the RNA-unwinding activity of the protein, but weakened its non-specific binding to RNA substrates and reduced its ability to stimulate CYT-18-dependent splicing of the *N. crassa* mt LSU- Δ ORF group I intron by ~15-fold.²⁷ These findings suggested that CYT-19's C-terminal domain contributes to a non-specific RNA-binding site that helps tether the core helicase region to RNA substrates, enabling multiple rounds of RNA chaperoning without dissociating from the RNA. It was possible, however, that the C-terminal domains of CYT-19 and Mss116p have more complex functions. In Mss116p, some mutations in this domain completely inactivate both the mt translation and RNA splicing functions,²⁸ and it has been suggested that strand annealing and/or RNA binding by the C-terminal domain rather than

RNA-unwinding by the helicase core region may be critically important for RNA splicing activity.²⁵

Here, we used multiple sequence alignments and secondary structure predictions to delineate subregions of the C-terminal domains of CYT-19 and Mss116p and carried out mutational analysis of Mss116p to analyze the function of these subregions *in vivo* and *in vitro*. We find that the C-terminal domains of both Mss116p and CYT-19 consist of an upstream region with predicted α -helices and a downstream basic tail, with little predicted structure. Truncations or mutations within Mss116p's predicted α -helical region strongly inhibit RNA-dependent ATPase activity, accounting for previous findings that mutations in this region result in loss of function in both mt translation and RNA splicing. Mss116p's basic tail contributes to the non-specific binding of group I and group II intron RNAs and is required for optimal activity of the core helicase region, but a C-terminally truncated protein that lacks the entire basic tail retains high RNA-dependent ATPase activity and functions with only moderately reduced efficiency in mt translation and RNA splicing *in vivo* and *in vitro*. Our results reveal no activities of Mss116p's C-terminal domain uniquely relevant to its mt translation or RNA splicing functions and suggest instead that it contributes by enhancing and modulating the RNA-unwinding activity of the core helicase region, which is primarily responsible for both functions.

Results

Sequence alignments of Mss116p and CYT-19 with Ded1p and Vasa

Figure 1 shows the amino acid sequences of Mss116p and CYT-19 aligned with those of two other well-characterized DEAD-box proteins: the *S. cerevisiae* DEAD-box protein Ded1p, which normally functions in translation, pre-mRNA splicing, and possibly other processes in the cytosol and nucleus,²⁹⁻³¹ but can also stimulate group I and group II intron splicing (M.D. and A.M.L., unpublished data),^{24,25} and the *Drosophila melanogaster* DEAD-box protein Vasa, which functions in translation and whose X-ray crystal structure with a bound RNA oligonucleotide has been determined.^{32,33} All four proteins have very similar core helicase regions, each spanning ~300 amino acid residues, with the conserved helicase motifs of Mss116p and CYT-19 matching closely those of Vasa and Ded1p. A secondary structure prediction program JPRED^{34,35} correctly predicted most of the experimentally determined secondary structure of Vasa and predicted very similar secondary structures for the core helicase regions of Ded1p, CYT-19, and Mss116p (predicted α -helices and β -strands indicated by amino acid residues in red and gold lettering in Figure 1).

Excluding the N-terminal mt import sequences (highlighted yellow), the N-terminal domains of Mss116p and CYT-19, defined as the regions upstream of the conserved motif Q, are 91 and 42 amino acid residues, respectively, and show little sequence similarity with the N-terminal domains of Vasa or Ded1p, or with each other, except for a few amino acid residues in predicted α -helix 2 (α 2).

The C-terminal domains of Mss116p and CYT-19, defined as the region downstream of helicase motif VI, are 193 and 176 amino acid residues, respectively. In both proteins, JPRED predicts structural elements corresponding to β -strand 14 and α -helices η 6 and α 15 of Vasa and Ded1p, followed by a region that is not present in the latter proteins and contains two predicted α -helices (Ca1 and Ca2; red), which are conserved in length and position between CYT-19 and Mss116p. This predicted α -helical region is followed by a hydrophilic tail (blue).

Hydropathy plots show that a hydrophilic C-terminal tail is readily discernible in all four DEAD-box proteins (Figure 2). In Mss116p, CYT-19, and Ded1p, the hydrophilic tails are 71-85 amino acid residues long and contain a preponderance of basic amino acid residues,

while in Vasa, the hydrophilic tail is shorter (~30 amino acid residues) and is acidic rather than basic. The basic tails of Ded1p, CYT-19, and Mss116p are largely devoid of predicted secondary structure, except for one or two short predicted α -helices in Ded1p and Mss116p. In Ded1p, the basic tail is 77 amino acid residues long and has a calculated pI of >11, with the most abundant amino acid residues being S (24%), G (22%), R (12%), and N (10%) and two RG sequences, a motif found in the C-terminal domains of some DEAD-box proteins.¹² In CYT-19, the basic tail is 71-amino acid residues long and has a calculated pI of >12, with R being the most abundant amino acid residue (21%), but no RG sequences. In Mss116p, the basic tail is 85-amino acid residues long and has a calculated pI of 9.5, with the most abundant amino acid residues being N (18%), S (18%), R (14%), and D (12%), and one RG sequence. Thus, while Ded1p, CYT-19, and Mss116p all have a basic hydrophilic C-terminal tail, its characteristics differ somewhat in each of the three proteins, and only in the case of CYT-19 is R the most abundant amino acid residue. Notably, the *cyt-19-1* mutations (G548V, L549S), which impair CYT-19 function,¹⁴ and the Mss116p mutation S557P, which inactivates the protein (see below), lie in similar positions outside the basic tail and within or just downstream of the predicted α -helix Ca2.

Mss116p and CYT-19 belong to distinct subfamilies of DEAD-box proteins

BLAST searches using the entire Mss116p sequence found as the highest matches several related proteins from yeasts, while BLAST searches using the CYT-19 sequence found a different set of >50 proteins from fungi, plants, and animals. Figure 3 shows multiple sequence alignments of the C-terminal domains of representative proteins of the Mss116p subfamily in yeasts (top grouping) and the CYT-19 subfamily in fungi and plants (middle grouping). Most of these proteins are predicted to have N-terminal mt targeting sequences by the TargetP 1.1 server (<http://www.cbs.dtu.dk/services/TargetP/>)³⁶ and are presumed to be mitochondrial. Exceptions in the figure are four proteins in the CYT-19 (middle) group (*Ustilago maydis* (U.may_82943, **EAK82943**), *Magnaporthe grisea* (M.gri_MG10312, **MG10312.4**), *Phaeosphaeria nodorum* (P.nod_09314, **SNOG_09314**), and *Arabidopsis thaliana* (A.tha_196479, **AtNP_196479**)), for which no intracellular localization could be predicted.

Figure 3 shows that the C-terminal domains of all members of the CYT-19 and Mss116p subfamilies have similar predicted secondary structures in which the region following α -helices $\eta 6/\alpha 15$ differs from those in Vasa and Ded1p by having two extended α -helices, denoted Ca1 and Ca2, preceding the basic C-terminal tail. In some proteins, JPRED predicts a single long α -helix instead of Ca1 and Ca2 and/or an additional short α -helix (Ca3) at the beginning of the basic tail.

Although predicted to have similar secondary structures, the C-terminal domains of proteins in the Mss116p and CYT-19 subfamilies can be distinguished from each other by several short conserved sequence motifs (highlighted in dark blue and green, respectively) and by the characteristics of the basic tail. Most Mss116p homologs have a C-terminal tail with high frequencies of SNRD/Q residues, while most CYT-19 homologs have a tail with a high frequency of R residues (e.g., *Aspergillus oryzae*, (A.ory_55384, 27%), *N. crassa* (N.cra_CYT-19, 21%), *C. immitis* (C.imm_03340, 21%), *C. globosum* (C.glo_04693, 20%)), and in some cases, also G residues (e.g., *P. nodorum*, (P.nod_06545, 59%), *M. grisea* (M.gri_367174, 45%), *Gibberella zeae* (G.zea_FG05108, 34%), *N. crassa* (N.cra_DEAD, 28%), *Chaetomium globosum* (C.glo_04693, 27%)). The amino acid residues mutated in *cyt-19-1* (G548 and L549; asterisks in Figure 3) are part of a conserved motif oGxxxxPPxo ("o" and "x" are hydrophobic and any amino acid residue, respectively) found just upstream of the basic tail in proteins of the CYT-19 but not Mss116p subfamily. By contrast, Mss116p S557, the site of the inactivating S557P mutation (see below), is not strongly conserved in the Mss116p subfamily, suggesting that the deleterious effect of the mutation may be due to

disrupting higher-order structure. Analysis of the Mss116p and CYT-19 N-terminal domains showed no conserved sequences or predicted secondary structural features within or between the two subfamilies (not shown).

The BLAST searches also identified a third group of potentially related DEAD-box proteins, which are not mitochondrial, but are also characterized by C-terminal domains with predicted Ca1 and Ca2 helices preceding a basic tail. This group includes the human DEAD-box protein DDX18 protein and the *S. cerevisiae* DEAD-box protein Has1p, which functions in the release of snoRNAs from rRNA during ribosome biogenesis.^{37,38} Like proteins in the CYT-19 subgroup, the proteins in the DDX18/Has1p subgroup contain the motif oGxxxxPPxo just upstream of the basic tail. However, in phylogenetic analysis based on full-length protein sequences, the DDX18/Has1p proteins comprise a separate clade (not shown) and are further distinguished by their basic tails, which are generally shorter than those of Mss116p and CYT-19 (Figure 3). Together, the bioinformatic analysis shows that CYT-19 and Mss116p belong to distinct subfamilies of mt DEAD-box proteins with some relationship to at least one group of non-mitochondrial DEAD-box proteins in higher organisms.

A yeast genetic assay for Mss116p function and unigenic evolution analysis

To analyze the functions of Mss116p's C-terminal domain, we used a yeast genetic assay developed by Perlman and co-workers in which an *mss116Δ* strain is complemented by Mss116p expressed from a centromere-containing (CEN) plasmid (Figure 4(a)).²⁸ Because Mss116p function is required for the synthesis of mt respiratory components, complementation by Mss116p variants is scored readily by the ability to support respiratory growth of *mss116Δ* mutants on media containing the non-fermentable carbon source glycerol (Gly⁺ phenotype). Mss116p's translational activation function is assayed by complementation of an *mss116Δ-I⁰* strain lacking mtDNA introns at low temperature (20–24°C), which accentuates the dependence of mt translation on Mss116p, while its RNA splicing function is assayed by complementation of *mss116Δ* strains containing different mtDNA introns at 30°C, a temperature at which mt translation is not dependent upon Mss116p function.²¹

By using this genetic assay, Perlman and coworkers found that Mss116p mutants having N-terminal truncations of 88 amino acid residues or C-terminal truncations of 139 and 178 amino acid residues (Mss116p/Δ526–664 and Mss116p/Δ487–664, respectively) have no detectable translational activation or RNA splicing activity, while Mss116p/Δ598–664 with a C-terminal truncation of 67 amino acid residues functions similarly to the wild-type protein in both translational activation and RNA splicing.²⁸ The genetic assay was also used to screen random PCR-induced mutations in the C-terminal domain, leading to isolation of the mutant S557P, which causes loss of function in RNA splicing.²⁸

To systematically identify important features of Mss116p's C-terminal domain, we used this genetic assay in conjunction with a method termed unigenic evolution analysis (Figure 4(b)).³⁹ This method involves isolating a collection of functional variants from a protein library containing random PCR-induced mutations, and then statistically analyzing ratios of missense to silent mutations to assess the degree of constraint on different protein regions. The latter is expressed by a parameter termed the mutability value (M), with negative and positive M values indicating hypomutable and hypermutable regions, respectively (see Materials and Methods).

To apply this method to Mss116p, we constructed a CEN-plasmid library of Mss116p variants containing random PCR-induced mutations in the C-terminal domain (amino acid residues 517–664) (Figure 4(a)). The upstream boundary was chosen to take advantage of a convenient restriction enzyme site introduced by Perlman and coworkers for screening of PCR-induced mutations.²⁸ The library was transformed into *S. cerevisiae* strain *mss116Δ-1⁺2⁺* containing mtDNA with twelve group I and II introns, and the transformants were plated on glycerol

medium at 30°C to select Gly⁺ colonies that express functional Mss116p variants capable of splicing these introns. After replating to confirm the Gly⁺ phenotype, the plasmids in these strains were sequenced via colony PCR to identify the mutations in their C-terminal domains. In total, we sequenced the C-terminal domains of 111 functional Mss116p variants containing 703 mutations of which 468 were missense mutations and 235 were silent mutations.

Figure 4(b) shows the amino acid sequence of the C-terminal domain of wild-type Mss116p with the missense mutations at each position listed below, and Figure 4(c) shows plots of M values as a function of amino acid position calculated for a 20-amino acid sliding window. Because it was possible that the function of the basic tail might depend on charge as well as amino acid identity, we calculated M values both by the conventional method in which all missense mutations are treated as equally negative (black line) and by a modified method in which conservative substitutions (defined according to the BLOSUM62 matrix⁴⁰) are treated as silent mutations (gray line). Both plots are similar in showing that the most constrained (hypomutable) region of the C-terminal domain is that containing the predicted α -helices, particularly Ca2 (negative M values significant at $p < 0.003$), while the region encompassing the basic tail (amino acid residues 578-664) is also hypomutable, but considerably less so than the predicted α -helical region.

Notably, the functional variants identified by unigenic evolution include ten with premature termination codons (asterisks in Figure 4(b)), which presumably result in truncated proteins. The farthest upstream of these is at position 569, deleting 96-amino acid residues, including the entire basic tail. Six amino acid residues were invariant in all functional proteins, including S557, which was identified previously as the site of an inactivating point mutation (see above). Four of the invariant amino acid residues are upstream of position 569 (Y544, P553, S557, and L562), while the other two are R residues downstream of position 569 (R584 and R639). Among the invariant residues in Mss116p, only Y544 and L562 appear significantly conserved (5/6) in Mss116p homologs from other yeasts, while the remainder are not conserved (Figure 3), suggesting that their hypomutability in Mss116p may reflect constraints on higher-order structure. Together, these findings indicate the predicted α -helical region is the most critical part of Mss116p's C-terminal domain for function *in vivo* and that the basic tail may also contribute, but is not essential *in vivo*.

***In vivo* analysis of Mss116p C-terminal truncations**

To further analyze the functions of the two C-terminal domain subregions, we compared directly the *in vivo* activities of wild-type Mss116p and two C-terminal truncations: Mss116p/ Δ 569-664, which has a 96-amino acid residue truncation that deletes the entire basic tail and corresponds to the longest functional truncation identified in the unigenic evolution analysis, and Mss116p/ Δ 551-664, which has a slightly longer truncation that extends into Ca2. In these experiments, the Mss116p wild-type and mutant proteins were expressed from a CEN plasmid with a C-terminal myc tag to monitor expression levels, and Mss116p's mt translation and splicing functions were assessed by complementation of *mss116 Δ* mutants lacking or containing mtDNA introns at different temperatures, as described above.

Figure 5 shows plating assays comparing the ability of Mss116p wild type (WT), the two C-terminal truncations, and a vector control to support glycerol growth of *mss116 Δ* strains. The assays show that Mss116p/ Δ 569-664, which completely lacks the basic tail, was as efficient as wild-type Mss116p in complementing both the mt translation defect in *mss116 Δ -I⁰* at 22°C and the splicing defects in *mss116 Δ* strains containing different combinations of mtDNA introns at 30°C. By contrast, Mss116p/ Δ 551-664, with a slightly longer C-terminal truncation that extends into the predicted α -helical region, was unable to complement these defects, with growth on glycerol not significantly different from that of the vector control.

Northern hybridizations confirmed that Mss116p/ Δ 569-664 supports efficient group I and group II intron splicing, while Mss116p/ Δ 551-664 has no detectable splicing activity above background (Figure 6(a) and (b)). The Mss116p wild-type and Δ 569-664 proteins were indistinguishable in their ability to support splicing of the group II introns *al5 γ* or *bI1* in *mss116 Δ* strains containing only those introns (Figure 6(a) and (b), right panels). However, in the *mss116 Δ -1⁺2⁺* strain, which contains multiple *COX1* and *COB* introns, Mss116p/ Δ 569-664 appears somewhat less efficient than wild-type Mss116p in supporting the production of spliced *COX1* and *COB* mRNA, possibly due to the cumulative effect of small decreases in the splicing efficiency of individual group I and/or group II introns (Figure 6(a) and (b)). In *mss116 Δ -1⁺2⁺* containing Mss116p/ Δ 551-664 or the vector control, neither precursor RNAs nor spliced mRNA were detectable as discrete bands, presumably reflecting a complete block in producing spliced mRNA coupled with instability of the multiple unspliced precursor RNAs containing different introns (Figure 6(a) and (b)). Immunoblots of whole-cell protein probed with an anti-myc antibody confirmed that both the Mss116p/ Δ 551-664 and Δ 569-664 truncations are expressed at high levels relative to the wild-type protein in all strains (Figure 6(c)). We note that the Mss116p wild-type and mutant proteins were expressed from their natural promoter from genes cloned on a low copy number CEN-plasmid and are of the size predicted for the mature proteins after processing of the mt targeting sequence. Thus, the relative levels of proteins in the immunoblots are likely to reflect those in mitochondria, with little or no contribution from inactive or insoluble precursors in the cytosol. Together, these findings show that the predicted α -helical region of the C-terminal domain is critical for both the mt translation and RNA splicing functions of Mss116p, while the basic tail may contribute but is not essential for either of these functions *in vivo*.

RNA-dependent ATPase activity of wild-type and mutant Mss116p proteins

To investigate further how the mutations in different regions of the C-terminal domain affect Mss116p function, we characterized the biochemical activities of the wild-type and mutant proteins *in vitro*. First, we compared the RNA-dependent ATPase activity of the Mss116p wild-type and Δ 569-664 proteins in the presence of saturating concentrations of different RNA substrates (Table 1). As published previously, for wild-type Mss116p, k_{cat} values with group I or group II intron RNAs ranged from 168-180 min⁻¹, with no detectable ATPase activity above background in the absence of RNA,¹⁵ and the protein also had similarly high ATPase activity with a 38-bp duplex RNA (k_{cat} = 190 min⁻¹).

Mss116p/ Δ 569-664, which lacks the entire basic tail, nevertheless retained high RNA-dependent ATPase activity similar to that of the wild-type protein with the different RNAs (k_{cat} with group I and group II intron RNAs = 195-208 min⁻¹, and k_{cat} with double-stranded RNA = 154 min⁻¹). By contrast, Mss116p/ Δ 551-664, with a slightly longer truncation extending into *Ca2*, had greatly decreased ATPase activity (undetectable under standard conditions and 15-20-fold lower than wild type under modified conditions for detecting low ATPase activity; not shown; see Materials and Methods). Two other mutant proteins (Mss116p/ Δ 503-664 and Mss116p/ Δ 508-664), with truncations that remove the entire predicted α -helical region, had no detectable ATPase activity with any of the RNAs tested (< 0.01 min⁻¹; not shown). Because the truncated proteins bind RNA more weakly than the wild-type protein (see below), RNA substrate saturation was confirmed separately in each case. The mutant S557P, which had been found previously to be inactive *in vivo*,²⁸ also had no detectable RNA-dependent ATPase activity above background with the group II intron *al5 γ* RNA (not shown). Together, these findings show that the basic tail region is not required for RNA-dependent ATPase activity, while truncations or point mutations in the predicted α -helical region strongly inhibit RNA-dependent ATPase activity. This strong inhibition of RNA-dependent ATPase activity explains the loss-of-function caused by mutations in the predicted α -helical region *in vivo*.

Splicing of group I and group II intron RNAs by wild-type and mutant Mss116p *in vitro*

Although the basic tail region is not essential for Mss116p function *in vivo*, both the unigenic evolution and complementation assays show that mutations or truncations within this region cause the protein to act suboptimally. To assess how the basic tail contributes to RNA splicing activity, we compared the ability of the purified Mss116p wild-type and $\Delta 569-664$ proteins to stimulate group I and group II intron splicing *in vitro*, using previously developed assays based on CYT-18-dependent splicing of the *N. crassa* mt LSU- Δ ORF group I intron^{14,24} and splicing of the *S. cerevisiae* aI5 γ and bI1 group II introns at low Mg^{2+} concentrations.^{14,24,26} Figure 7 shows gel analysis of 120-min splicing reactions in the presence or absence of ATP, ADP, or AMP-PNP, and Figure 8 shows time courses of precursor RNA disappearance in ATP-stimulated splicing reactions with different concentrations of wild-type and mutant proteins.

The results show that compared to the wild-type protein, Mss116p/ $\Delta 569-664$ has only moderately reduced activity in stimulating both group I and group II intron splicing. For the group II introns, the splicing stimulation by both the wild-type and mutant proteins was completely ATP-dependent, with no detectable stimulation above background in the absence of nucleotide or in the presence of hexokinase-treated AMP-PNP or ADP (Figure 7(b) and (c)). For the *N. crassa* mt LSU- Δ ORF group I intron, stimulation of CYT-18-dependent splicing was highest with ATP, but there was detectable stimulation above the level of CYT-18 alone without nucleotide or in the presence of AMP-PNP or ADP (Figure 7(a), and data not shown). ATP-independent stimulation of this group I intron splicing reaction by wild-type Mss116p was reported previously, most strongly in the presence of ADP, and may reflect ATP-independent RNA chaperoning via single-strand capture among other possibilities (see Introduction).^{15,24}

The time courses enabled us to compare the splicing efficiency of the wild-type and mutant proteins with each intron by calculating k_{cat}/K_m values from plots of the observed rate constants (k_{obs}) as a function of protein concentration (Figure 8). A complication is that while plots of k_{obs} versus protein concentration are best fit by a hyperbolic function for both the wild-type and mutant proteins with the *N. crassa* mt LSU- Δ ORF intron and for the wild-type protein with the bI1 intron, giving a direct measure of k_{cat}/K_m , the plots in the remaining cases are sigmoidal, suggesting the participation of multiple protomers. In these cases, we estimated an efficiency value that could be compared to k_{cat}/K_m from linear fits of the data up to saturation (dashed lines; see Figure Legend). Compared in this way, Mss116p/ $\Delta 569-664$ was about 1.6-fold less efficient than the wild-type protein in splicing the *N. crassa* mt LSU- Δ ORF group I intron (Figure 8(g)), 1.9-fold less efficient in splicing the *S. cerevisiae* aI5 γ group II intron (Figure 8(h)), and about 5.4-fold less efficient in splicing the *S. cerevisiae* bI1 intron (Figure 8(i)). The sigmoidal concentration dependences may reflect that some introns are more difficult to splice, requiring the cooperative action of multiple protomers, and that Mss116p mutations may have complex and countervailing effects - *e.g.*, Mss116p mutations that decrease the efficiency of unwinding may require increased cooperativity, but may also affect the ability of the protomers to function cooperatively (see also ref. ¹⁵). As found previously, high concentrations of Mss116p can inhibit the splicing reactions (*e.g.*, see data for wild-type protein in Figure 8(i)). Together, these results show that Mss116p's basic tail is required for optimal splicing of both group I and group II introns. However, the contribution of the basic tail to RNA splicing appears smaller for Mss116p than for CYT-19, where a truncation within the basic tail decreased the efficiency of CYT-18-dependent splicing of the *N. crassa* mt LSU- Δ ORF intron by ~15-fold.²⁷

Duplex-unwinding and strand-annealing activities of wild-type and mutant Mss116p

In principle, the suboptimal RNA splicing activity of Mss116p/ Δ 569-664 could reflect decreased unwinding activity, decreased RNA-binding affinity, or a combination of the two. To assess these possibilities, we assayed duplex unwinding and strand-annealing activities under both group I and group II intron splicing conditions (100 mM KCl and 5 or 8 mM Mg^{2+} , respectively). To accentuate differences between the wild-type and mutant proteins, the assays were done with RNA oligonucleotide substrates containing a relatively long 13-bp duplex region with or without a 25-nt 3' overhang.¹⁵

As shown in Figure 9, Mss116p/ Δ 569-664 unwound the RNA duplexes with and without the overhang under both group I and group II intron splicing conditions, but its unwinding activity was consistently lower than that of the wild-type protein. For Mss116p/ Δ 569-664 with the 3' overhang substrate, the unwinding rate constant at enzyme saturation (k_{unw}^{max}) was decreased about 13-fold compared to the wild-type protein in reaction medium containing 5 mM Mg^{2+} (group I intron splicing conditions; Figure 9(a)) and about 30-fold in reaction medium containing 8 mM Mg^{2+} (group II intron splicing conditions; Figure 9(b)). In addition, the functional binding constants ($K_{1/2}$) for the mutant protein were about three-fold lower than that of the wild-type protein under both conditions (Figure 9(a) and (b)). Both the wild-type and mutant proteins displayed cooperative behavior in functional binding ($n > 2$) under both conditions (Figure 9(a) and (b)).

For the blunt-end substrate, saturation could be approached only for the wild-type protein in reaction medium containing 5 mM Mg^{2+} (Figure 9(c) and (d)). While this precludes definitive conclusions, the plots and $K_{1/2}$ values calculated by assuming k_{unw}^{max} is the same as that for the overhang substrate (see Figure Legend for justification)^{9,24} show that deletion of the basic tail also inhibits unwinding of the blunt-end substrate and the degree of inhibition appears similar to that for the overhang substrate (Figure 9(c) and (d)). These findings suggest that the C-terminal tail does not selectively bind the single-stranded overhang region during unwinding. Rather, the deletion of the C-terminal tail seems to result in a "general" non-specific RNA binding defect, as discussed in more detail below (see Discussion).

For the annealing activity, neither the wild-type and nor mutant protein could saturate the substrates at experimentally accessible concentrations, consistent with previous observations for Ded1p.¹² Comparison of annealing rate constants (k_{ann}) at the same enzyme concentration showed that Mss116p/ Δ 569-664 consistently had lower annealing activity than wild-type Mss116p for both the overhang and blunt-end substrates at both 5 and 8 Mg^{2+} (activity differences ranged from two- to seven-fold; Table 2). Thus, the C-terminal tail is also required for maximal strand-annealing activity. We also note that at comparable enzyme concentrations strand annealing was consistently slower for the blunt-end than the overhang substrate, consistent with results for Ded1p showing an increase in the annealing activity with substrate length.¹²

Taken together, the above data show that deletion of the basic C-terminal tail affects both the duplex-unwinding and strand-annealing activities of Mss116p, but does not appear to differentially affect the use of blunt or overhang substrates. The latter finding suggests that the basic tail does not have a strong preference for interaction with single- or double-stranded regions and are most consistent with a model in which the basic tail interacts with both duplex and single-stranded RNA primarily via non-specific electrostatic interactions.

Binding of wild-type and mutant Mss116p to group I and group II intron RNAs

To assess directly how Mss116p's C-terminal domain contributes to RNA binding, we carried out equilibrium-binding assays with Mss116p wild-type and C-terminal truncation mutant

proteins (Table 3). First, we tested the binding of the *N. crassa* mt LSU-ΔORF group I intron and the *S. cerevisiae* a15γ and b11 group II introns. In agreement with previous results,²⁴ wild-type Mss116p bound all three RNAs similarly with K_d s of 1.6-2.3 nM in the presence or absence of 1 mM ATP and ADP (compared to 3-11 nM in the previous work), and the K_d values decreased little if at all in the presence of 1 mM AMP-PNP. The latter behavior differs from that of CYT-19, where the K_d values for group I and group II intron RNAs in the absence of nucleotide or presence of ATP or ADP were higher (10-40 nM), but decreased substantially to 0.5-1 nM in the presence of AMP-PNP.^{14,26} Other DEAD-box proteins also show enhanced RNA substrate binding in the presence of AMP-PNP, which is thought to reflect locking of the core helicase region in a closed conformation with a high RNA-binding affinity.^{33,41}

The C-terminal truncation mutant Mss116p/Δ569-664, which lacks only the basic tail, bound the group I and II intron RNAs with K_d s of 16-20 nM in the presence or absence of ATP or ADP, 7- to 13-fold higher than those for the wild-type protein. Notably, the truncation increased the K_d values for all three intron RNAs proportionately within error, suggesting that the basic tail contributes to the non-specific binding of these intron RNAs. As for wild-type Mss116p, the K_d values for the binding of group I and II intron RNAs by Mss116p/Δ569-664 did not decrease significantly in the presence of 1 mM AMP-PNP. These findings show that the basic tail region is required for tight binding of Mss116p to group I and group II intron RNAs (K_d s = 0.8-2.3 nM for the wild-type protein compared to 15-20 nM for the mutant lacking the basic tail). This weaker non-specific binding of the Mss116p/Δ569-664 to group I and II intron RNAs presumably contributes to its suboptimal RNA splicing activity.

The C-terminal truncation Mss116p/Δ551-664, which extends into Ca2, further weakened non-specific binding and also resulted in a change in the binding curves from hyperbolic to sigmoidal (see Table 3, Legend). The $K_{1/2}$ values for the binding of Mss116p/Δ551-664 to all three introns RNAs with or without 1 mM ATP, ADP, or AMP-PNP were 140-243 nM, 61-304-fold higher than for the wild-type protein. The $K_{1/2}$ values for binding of the group I and II intron RNAs by Mss116p/Δ503-664, which completely lacks the predicted α-helical region, were also > 100 nM in the presence or absence of ATP, ADP, or AMP-PNP, but complete saturation for this poorly expressed mutant protein was not possible at accessible protein concentrations (not shown). The findings for the longer truncation mutants indicate that the predicted α-helical region also contributes to non-specific RNA binding, either directly or indirectly by stabilizing the RNA-binding structure of the core helicase region.

Binding of wild-type and mutant Mss116p to single- and double-stranded RNA oligonucleotides

Finally, we gained additional insight into how the C-terminal truncations affect the activity of Mss116p by comparing the binding of wild-type and mutant proteins to single- and double-stranded RNAs oligonucleotides (Table 4). Wild-type Mss116p bound a 38-nt single-strand RNA oligonucleotide with a K_d of 19-28 nM in the absence of nucleotide or presence of ATP and ADP, but this K_d decreased substantially to 2.0 nM in the presence of AMP-PNP. By contrast Mss116p/Δ569-664 bound the single-stranded RNA oligonucleotide with a K_d of 13 nM, which was not significantly affected by ATP, ADP, or AMP-PNP (K_d s = 11-15 nM). The wild-type and Mss116p/Δ569-664 mutant proteins bound the double-stranded RNA oligonucleotide with K_d s of 24 and 9 nM, respectively, and for neither protein, did the K_d decrease differentially in the presence of AMP-PNP compared to other nucleotides. Thus, the deletion of the basic tail specifically impairs the tighter binding of the single-stranded RNA oligonucleotide by the core helicase region in the presence of AMP-PNP. This effect, which may be related to the decrease in k_{unw} at saturating protein concentration, could reflect that the basic tail contributes additional RNA contacts and/or to a nucleotide-dependent protein

conformational change required for the tight binding of single-stranded RNA by the core helicase region.

Except for the tighter binding of the single-stranded RNA oligonucleotide in presence of AMP-PNP, the K_{ds} for the binding of single- and double-stranded RNA oligonucleotides by the wild-type protein are about 10-fold higher than those for the binding of group I and group II intron RNAs (Tables 3 and 4). By contrast, for Mss116p/ Δ 569-664, which lacks the basic tail, the K_{ds} for the binding of RNA oligonucleotides are in the same range as the higher K_{ds} for binding the group I and group II intron RNAs (Tables 3 and 4). These findings are consistent with a model in which the core helicase region interacts with short duplex-containing regions of large RNAs for unwinding, while the basic C-terminal tail interacts with separate sites, enabling tighter binding and tethering the protein for multiple rounds of unwinding.

Discussion

Here we find that the DEAD-box proteins *S. cerevisiae* Mss116p and *N. crassa* CYT-19, which function in mt translation and group I and II intron splicing, have C-terminal domains that consist of two distinct subregions: an upstream region with predicted α -helices and a basic C-terminal tail with little predicted secondary structure. The predicted α -helical region is absent in the well-studied DEAD-box proteins Ded1p and Vasa and is a defining characteristic of Mss116p, CYT-19, and related proteins. *In vivo* and *in vitro* assays with Mss116p show that mutations in the predicted α -helical region strongly inhibit the RNA-dependent ATPase activity of the RNA helicase core and lead to loss of function in mt translation and RNA splicing. The basic tail helps tether the core helicase region to large structured RNA substrates and also influences its RNA-binding and unwinding activities. Our results reveal no activities of Mss116p's C-terminal domain uniquely relevant to its mt translation or RNA splicing functions.

Multiple sequence alignments show that although Mss116p and CYT-19 have related C-terminal domains, they nevertheless belong to two distinct DEAD-box protein subfamilies. These two subfamilies comprise different clades based on phylogenetic analysis of the full-length proteins (not shown), and their C-terminal domains are distinguished from each other by different conserved sequence motifs and by features of the basic tail. Both subfamilies consist largely if not entirely of putative mitochondrial proteins, with the CYT-19 subfamily found in fungi and plants, and the Mss116p subfamily found in yeasts. A third more distantly related DEAD-box protein subfamily with an analogous C-terminal domain is not mitochondrial and includes *S. cerevisiae* Has1p, which functions in ribosome assembly, and its human ortholog, DDX18. We note that the human mt DEAD-box protein DDX28 and three other human DEXH/D-box proteins with predicted mt targeting sequences (DDX30, PEO1, and SUV3) differ from Mss116p and CYT-19 in having C-terminal domains lacking an extended α -helical region. These phylogenetic differences may reflect that DEAD-box proteins of the Mss116p and CYT-19 subfamilies are particularly well-suited for specialized functions in fungal and plant organelles, such as splicing of group I and group II introns, which are not present in the mitochondria of metazoan animals. Alternatively, DEAD-box proteins may be tuned by coevolution to function optimally on their native substrates, making it difficult to replace them by other DEAD-box proteins without incurring a selective disadvantage. We found previously that although Ded1p could replace CYT-19 or Mss116p in splicing mt group II introns *in vitro*, it acts suboptimally on some introns.²⁴

The unigenic evolution analysis of Mss116p showed that the C-terminal α -helical region, particularly C α 2, is the most highly constrained part of the C-terminal domain. This region is the site of previously isolated missense mutations that inhibit the splicing activity of both Mss116p and CYT-19,^{14,28} and mutations or truncations within this region in Mss116p cause

loss of function in both mt translation and RNA splicing. Biochemical assays show that the loss-of-function mutations in Mss116p's predicted α -helical region strongly inhibit RNA-dependent ATPase activity, suggesting that this region is required to stabilize the structure of the RNA helicase core. The same appears to be true for CYT-19, where the *cyt-19-1* mutations in the predicted α -helical region likewise inhibits RNA-dependent ATPase activity (S.M. and A.M.L., unpublished data). The unigenic evolution analysis identifies several highly conserved or invariant amino acid residues in the predicted α -helical region that may provide a starting point for more detailed structure-function analysis. It is also noteworthy that truncations within Mss116p's predicted α -helical region further decrease non-specific RNA-binding affinity beyond the decrease for truncations that remove only the basic tail. This finding suggests that the predicted α -helical region also contributes to RNA binding, either directly or indirectly by affecting the RNA-binding structure of the core helicase region.

Although the C-terminal α -helical region is critically important in Mss116p and likely in related DEAD-box proteins, it is not present in Ded1p, where the RNA helicase core is followed directly by the basic tail. Because Ded1p has RNA-dependent ATPase and RNA-unwinding activities similar to Mss116p and CYT-19 and can also promote group I and group II intron splicing, we infer that the C-terminal α -helical region is neither an essential part of the RNA helicase core, nor a special adaptation required for RNA splicing activity. Rather, it seems likely that the RNA helicase core in Mss116p and CYT-19 has evolved to become dependent upon interaction with the predicted α -helical region for structural stabilization. This presumed interaction could also be used to regulate the activity of the core helicase region, perhaps providing a means of communication between distinct RNA-binding sites in different domains.

As found previously for CYT-19,²⁷ our results suggest that the basic C-terminal tail of Mss116p contributes to the non-specific binding of large RNA substrates. In the unigenic evolution analysis, the basic tail appears hypomutable, albeit less so than the predicted α -helical region, and the Mss116p/ Δ 569-664 protein, which lacks the basic tail, functions suboptimally in group I and group II intron splicing *in vivo* and *in vitro*. Biochemical assays show that deletion of the basic tail leaves high RNA-dependent ATPase activity, but decreases RNA-unwinding activity at saturating protein concentrations by 13- to 30-fold, and increases the K_d s for binding of both group I and II intron RNAs from 1-3 nM to 15-20 nM. These findings suggest that Mss116p's basic tail, like that of CYT-19, is or contributes to a second, non-specific RNA-binding site that helps tether the core helicase region to structured RNAs for multiple rounds of RNA chaperoning. However, Mss116p's basic tail appears to make a relatively smaller contribution to the binding of group I and group II introns than does CYT-19's, where a C-terminal truncation that deletes part of the basic tail increased the K_d for a group I intron ribozyme substrate from 30 to 200 nM and decreased the efficiency of CYT-18-dependent splicing of the *N. crassa* mt LSU- Δ ORF intron by ~15-fold.²⁷ As for CYT-19 and Mss116p, the basic tails of Ded1p and other DEAD-box RNA chaperones with multiple substrates may also contribute to non-specific RNA binding.

A function of Mss116p's basic tail in non-specific RNA binding is also evident from the results of the unwinding and annealing assays (Figure 9 and Table 2). The decrease in functional binding constants for the unwinding reactions and in the observed annealing rate constants for the mutant lacking the basic tail (Figure 9 and Table 2, respectively) are straightforward to interpret as binding defects. Importantly, however, the 13- to 30-fold decrease in the unwinding rate constant at saturating protein concentration indicates that the basic tail additionally contributes to a step that becomes rate-limiting for the mutant protein after RNA binding. One possibility is that the basic tail helps activate the core helicase region by affecting its structure or position on the RNA substrate. The basic tail could also act directly in strand separation, for example by helping to sequester the emerging single strand, thereby driving the reaction forward. The finding that deletion of the basic tail impedes the nucleotide-dependent tight

binding of single-stranded RNA by the core helicase region (Table 4) provides additional evidence for a significant effect on an important step in the unwinding cycle. The deletion of the basic tail does not appear to differentially affect the unwinding of duplexes with or without single-stranded overhangs, consistent with largely non-specific RNA-binding with a strong electrostatic component.

Importantly, we find that all mutations in both subregions of Mss116p's C-terminal domain have parallel effects on mt translation and RNA splicing. Mutations in the predicted α -helical region lead to loss of function in both processes by inhibiting RNA-dependent ATPase activity, and truncation of the basic tail has no effect on mt translation and only small effects on group I and II intron splicing due to suboptimal RNA binding and unwinding. The previously characterized Mss116p mutants GKTm in the ATP-binding site (GTGKT in helicase motif I changed to ATANA) and SAT/AAA in helicase motif III also have parallel effects on mt translation and RNA splicing, the former leading to complete loss of function in both processes, and the latter leading to leaky defects in both processes, due to impaired ability to unwind longer RNA duplexes.^{15,28} Although our results do not exclude the possibility that Mss116p's C-terminal domain is involved in specific interactions with as yet to be tested RNA substrates, the results thus far suggest that it contributes to general RNA chaperone activity by tethering the core helicase domain to large RNA substrates via non-specific binding and by modulating and regulating its RNA-unwinding activity, which is critically important for both the mt translation and RNA splicing functions.

Materials and Methods

Computational methods

For multiple sequence alignments, Mss116p and CYT-19 homologs were identified by BLAST searches of non-redundant protein sequences at NCBI, supplemented by searches of genome sequencing projects. Sequences were aligned with CLUSTALX,⁴² and alignments were refined manually based on secondary-structure predictions. Similar amino acids were defined according to the matrix of Henikoff and Henikoff.⁴⁰ Secondary structures were predicted using the JPred server (<http://www.compbio.dundee.ac.uk/~www-jpred/submit.html>).^{34,35} Hydrophobicity plots were done by the method of Kyte and Doolittle⁴³ with a 20-amino-acid residue sliding window (<http://www.vivo.colostate.edu/molkit/hydropathy/>).

S. cerevisiae strains and growth media

S. cerevisiae *mss116Δ* strains have the nuclear background of strain 161-U7 MATa *ade1 lys1 ura3* (denoted 161) with the *MSS116* gene replaced by a *kan^r* disruption cassette.²¹ Derivatives of the *mss116Δ* strain containing different mtDNA introns were constructed previously by cytoduction.²¹ *mss116Δ-1⁺2⁺* contains eight group I introns (aI3 α , aI4 α , aI5 α , aI5 β , bI2, bI3, bI4, bI5) and four group II introns (aI1, aI2, aI5 γ and bI1); *mss116Δ-GII5 γ* contains four group I introns (aI3 α , aI4 α , bI4, bI5) and the group II intron aI5 γ ; and *mss116Δ-aI5 γ* and *mss116Δ-bI1* contain the group II introns aI5 γ and bI1, respectively.

S. cerevisiae strains were grown in yeast peptone medium (1% yeast extract, 2% peptone) supplemented with 2% glucose (YPD) or 2% glycerol (YPG). *ura3* transformants were selected on synthetic medium (YNB; 0.17% yeast nitrogen base, 1% casamino acids, and 0.5% ammonium sulfate), supplemented with lysine (30 mg/l), tryptophan (30 mg/l), adenine (20 mg/l) and either 2% glucose (YNBD) or 2% raffinose (YNBR). Solid media contained 2% bacto agar (DIFCO, Sparks, MD).

Recombinant plasmids

Mss116p was expressed in *S. cerevisiae* from derivatives of the CEN plasmid pRS416.⁴⁴ pHRH108 contains the *MSS116* gene from wild-type *S. cerevisiae* strain 161 cloned as a 3.0-kb HindIII fragment in pRS416, and pHRH110 is a derivative of pHRH108, which expresses Mss116p with a C-terminal myc-tag.²¹ pHRH197, used for mutagenesis, has an *MSS116* ORF with two silent mutations (A222G and C1546T), which destroy and create an XbaI site, respectively, and an added SacII site just downstream of the *MSS116* stop codon.²⁸

Derivatives of pHRH197, which express C-terminally truncated Mss116p/Δ569-664 and Mss116p/Δ551-664 with a C-terminal myc-tag, were made by PCR of pHRH197, using an upstream primer that corresponds to a sequence just in front of the XbaI site (MSS116-1440s, 5'-CAGAAGTGGGAAAGAAGGAT), and a downstream primer that introduces a SpeI site at the site of the truncation (551tgamyc, 5' AAAACTAGTGCCTTTTCAGAAAACCTGTATTC; 569tagmyc, 5' AAAACTAGTTAATTGAGGATCATTTAACAG). The PCR products were digested with XbaI and SpeI, gel-purified, and three-way ligated with a gel-purified 3.8-kb SpeI + NcoI fragment of pHRH178²¹ (myc-tag, *MSS116* 3'-UTR, and part of the vector), and a gel-purified 3.7-kb XbaI + NcoI fragment of pHRH197 (N-terminal 516 codons of *MSS116*, the *MSS116* 5'-UTR, and another part of the vector).

pMAL-Mss116p, used for expression and purification of wild-type Mss116p from *E. coli*, contains the Mss116p coding sequence (codons 37-664) with an in-frame N-terminal MalE fusion cloned downstream of a tac promoter in the expression vector pMAL-c2t.²⁴ The latter is a derivative of pMAL-c2x (New England Biolabs; Ipswich, MA) in which the factor Xa protease cleavage site between MalE and the expressed protein is replaced by a tobacco etch virus protease-cleavage site.⁴⁵ Derivatives of pMAL-Mss116p expressing the C-terminal truncations Mss116p/Δ551-664 and Mss116p/Δ569-664 were made by PCR of pHRH197 using primer MSS116-1440s (see above) and a downstream primer that introduces a stop codon followed by a SacII site at the site of the truncation (551tga, 5' CATCCCGCGGTTATCAGCGCCTTTTCAGAAAACCTGTATTC; 569tag, CATCCCGCGGTTACTATAATTGAGGATCATTTAACAG). The 450-bp PCR products were digested with SacII, filled in with Klenow DNA polymerase, digested with XbaI, gel-purified, and three-way ligated with two gel-purified restriction fragments, one a 1.2-kb BsrGI + XbaI fragment of pHRH197 containing *MSS116* codons 132-516, and the other a 7.4-kb fragment containing *MSS116* codons 1-132 and the expression vector (HindIII fragment of pMAL-Mss116p filled in with Klenow polymerase and digested with BsrGI).

Derivatives of pMAL-Mss116p, which express C-terminal truncations Mss116p/Δ503-664 and Mss116p/Δ508-664, were made by PCR of pHRH197 using the upstream primer MSS116-BsrGs (5' CCTGTACAGCAAAAGACTATC) and a downstream primer that introduces a stop codon followed by a HindIII site at the site of the truncation (MSS116del503a, 5' TTTAAGCTTCTATTGTTTGGCAATGACTATAT; MSS116del508a, 5' TTTAAGCTTCTACTCATATTTTCTTGTGTTGTC). The PCR products were digested with HindIII and BsrGI, gel-purified, and ligated to a gel-purified 7.4-kb HindIII + BsrGI fragment of pMAL-Mss116p containing the N-terminal 132 codons of *MSS116* and the expression vector.

The Mss116p/S557P mutant was made by PCR amplifying two overlapping segments of the *MSS116* ORF from pHRH197 with primers that introduce the mutation at the overlap. *MSS116* codons 516 to 557 were amplified using primers MSS116-BsrGs (5' CCTGTACAGCAAAAGACTATC) and MSS116-S557Pa (5' ACCATAAGTGGgAGCAATTTC), and codons 557-664 were amplified using primers MSS116-S557Ps (5' GAAATTGCTcCCACTTATGGT) and MSS116-3'UTR_HIIIa (5'

TTTAAGCTTCAACCCACATCTCTTGGTCCT) (lower case letter indicates mutant nucleotide residue). The two PCR products were gel-purified, mixed, and used as template for a second PCR with primers MSS116-BsrGs and MSS116-3'UTR_HIIIa (see above). The resulting PCR product was digested with HindIII and BsrGI, gel-purified, and cloned into the vector, as described above for the Mss116p/Δ503-664 and Mss116p/Δ508-664.

All cloning steps were done in *E. coli* DH5α (Invitrogen; Carlsbad, CA) grown in LB medium (0.5% yeast extract, 1% peptone, and 1% NaCl, pH 7), with 2% Difco agar for solid media, and ampicillin (100 μg/ml) and kanamycin (40 μg/ml) added as required. All constructs were sequenced through the region amplified by PCR to insure that the cloning was correct and that no adventitious mutations had been introduced by PCR.

Unigenic evolution analysis

For unigenic evolution analysis, a library of Mss116p variants with random mutations in the C-terminal domain (amino acid residues 517-664) was constructed by error-prone PCR of the *MSS116* ORF in pHRH108.^{39,46} The PCR was done in 100 μl of reaction medium containing 0.06 pmol of pHRH108, 50 pmol each of primers Mss116-1536Xba(s) 5'-ATAAAATCTGAGGTTCTAGACGC and Mss116-3001sacII(as) 5'-AACCCACATCCCGCGGTCCTCCT, 0.39 mM dATP, 0.15 mM dCTP, 1.17 mM dGTP, 3.85 mM dTTP, 11 mM MgCl₂, 0.5 mM MnCl₂, and 5 units Taq DNA polymerase (Invitrogen). The reaction conditions were initial denaturation at 94°C for 5 min, followed by 16 cycles of 91°C for 1 min, 51°C for 1 min, and 72°C for 10 min, plus a final extension at 72°C for 15 min. The 0.5-kb PCR product containing the mutagenized segment of the *MSS116* ORF was gel-purified, digested with XbaI and SacII, and swapped for a *kan^R* stuffer fragment that had been inserted at the corresponding position of pHRH197. The use of an intermediate with a *kan^R* stuffer prevents contamination of the library by the wild-type *MSS116* sequence. The mutation frequency, determined by sequencing 36 randomly selected clones from the library, was 0.78%, corresponding to ~10 amino acid substitutions per protein. The library was transformed into *mss116Δ-1⁺2⁺*, and plated on YPG at 30°C to identify splicing-competent variants, which were picked to YPG and further purified by restreaking on the same medium. The mutated region of Mss116p was then amplified by colony PCR (<http://www.imbf.ku.dk/LisbyHolmberg/colonyPCR.htm>), using primers MSS116-1440s (see above) and MSS116-3'UTRa 5'-TTAAAATACAACCCACATCTC. The PCR products were purified by using a Qiagen PCR clean-up kit (Qiagen, Valencia, CA) and sequenced using primer MSS116-1440s.

Mutational data were analyzed as described.^{39,46,47} Briefly, the mutability value (*M*) in a 20-amino acid residue sliding window was calculated by using the formula:

$$M = \frac{fOmis}{fEmis} - 1$$

where *fOmis* is the observed frequency of missense mutations and *fEmis* is the expected frequency. The latter was calculated for each codon based on the probability of a single nucleotide change producing a missense mutation, correcting for the transition/transversion ratio of 3.4 for the variants selected from the library. Negative mutability values indicate hypomutability, with a minimum value of -1 indicating no missense mutations in a given window. Positive values indicate hypermutability and were normalized by using the formula:

$$\frac{1}{fEmis} - 1$$

so that all mutability values fall between +1 and -1. χ^2 was calculated for each hypomutable region by using the equation:

$$\chi^2 = \frac{(\text{abs}(O_{\text{mis}} - E_{\text{mis}}) - 0.5)^2}{E_{\text{mis}}} + \frac{(\text{abs}(O_{\text{sil}} - E_{\text{sil}}) - 0.5)^2}{E_{\text{sil}}}$$

where abs indicates the absolute value; O_{mis} and O_{sil} are numbers of missense and silent mutations found in a given region; and E_{mis} and E_{sil} are expected numbers of missense and silent mutations in that region.

***S. cerevisiae* genetic assay of Mss116p function**

S. cerevisiae mss116Δ strains were transformed with CEN plasmids expressing Mss116p wild-type or mutant proteins with a 10-amino acid residue C-terminal myc-tag. Transformants were selected by complementation of the *ura3* mutation on YNBD plates with the supplements indicated above, then grown up in YNBR with the same supplements to O.D.₆₀₀ = ~1.0. Cells (5×10^7 per ml) were then transferred to a microtiter plate, serially diluted, and stamped onto YPD and YPG plates.

Northern hybridizations and immunoblotting

S. cerevisiae strains were grown at 30°C in YNBR to O.D.₆₀₀ = 1.0-2.0. For Northern hybridizations, RNA was isolated as described,⁴⁸ extracted with phenol-chloroform:isoamyl alcohol (25:24:1), and ethanol precipitated. Portions (~3 O.D.₂₆₀ units) were treated with 20% glyoxal at 60°C for 15 min and run in a 1.4% agarose gel containing 10 mM phosphate buffer pH 7 at 4°C with the buffer recirculated. The gel was blotted to a nylon membrane (Hybond-XL, GE Healthcare, Piscataway, NJ), and the blots were hybridized with 5'-end labeled DNA oligonucleotide probes (*COB* exon 6 probe: 5' AAGGTACTTCTACATGGCATGCT; *COX1* exon 6 probe: 5' ATTTTCATCTGCGAAAGCATCAGGAT; *COX2* exon probe: 5' AGGTAATGATACTGCTTCGATC).

For immunoblotting, proteins were isolated by trichloroacetic acid precipitation (David Bowtell University of Melbourne, Australia, <http://www.bio.com/protocolstools/protocol.jhtml?id=p1248>). Protein pellets were washed once with 0.5 M Tris and resuspended in 150 μl SDS-PAGE sample buffer. Portions (~30 μg protein) were run in a 0.1% SDS-7.5% polyacrylamide gel (BioRad, Hercules, CA) and transferred to nitrocellulose membrane by using a semi-dry electro-blotting apparatus (Hoefer SemiPhor, San Francisco, CA). The blots were probed with anti-myc and anti-porin primary antibodies (1:1,000 dilutions), developed using an ECL Plus Western Blotting-kit (GE Healthcare), and imaged using a FluoroImager (Typhoon Trio, Amersham Bioscience) or on Kodak Biomax XAR film.

Protein purification and biochemical assays

Wild-type and mutant Mss116 proteins were expressed in *E. coli* Rosetta 2 (Novagen/EMD Biosciences, Madison, WI) from pMAL-plasmids with N-terminal MalE fusions and purified as described.^{15,24} CYT-18 was also expressed and purified as described.⁴⁹ Proteins were dialyzed into 500 mM KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, and 50% glycerol, flash-frozen, and stored at -80°C.

ATPase activity was assayed by polyethyleneimine-cellulose thin-layer chromatography, as described.²⁴ Assays were done at 30°C in 20 μl of reaction medium containing 100 mM KCl, 5 or 8 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 10% glycerol, 50 nM Mss116p, 10 μCi [γ -³²P] ATP (10 Ci/mmol; Perkin Elmer, Wellesley, MA), and saturating (3 mM) unlabeled ATP. Group I or group II intron RNAs or a double-stranded RNA oligonucleotide (duplex of R38B: 5' UCUUUACGGUGCUAUUUUGUUUUUGUUUUUGUUUUUGUUUUU and R-38C: 5' AAAACAAAACAAAACAAAACAAAUAGCACCGUAAAGA) were added at the concentrations indicated in Table 1. Mutant proteins that showed no detectable ATPase activity

under the above conditions were reassayed in reactions with 100 μ M instead of 3 mM unlabeled ATP to increase the sensitivity of detection of ATP hydrolysis.

Group I and group II intron splicing reactions were as described.²⁴ For group I introns, reactions were with the indicated concentrations of wild-type or mutant Mss116p, 20 nM CYT-18 dimer, and 10 nM ³²P-labeled precursor RNA at 25°C in 100 μ l of reaction medium containing 100 mM KCl, 5 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 10% glycerol, 1 mM GTP-Mg²⁺. For group II introns, reactions were with the indicated concentrations of wild-type or mutant Mss116p, 10 nM ³²P-labeled precursor RNA at 30°C in 100 μ l of the same reaction medium containing 8 mM MgCl₂. 1 mM ATP-Mg²⁺, AMP-PNP-Mg²⁺, or ADP-Mg²⁺ were added as indicated for individual experiments, using ADP and AMP-PNP stocks that had been treated with hexokinase to remove residual ATP.¹⁵ Reactions were initiated by adding RNA, and the data for precursor RNA disappearance were fit to single exponential equations to obtain k_{obs} .

Duplex-unwinding and strand-annealing assays were done at 30°C in buffers used for the *in vitro* splicing reactions, as described.²⁴ The substrates were comprised of the following oligonucleotides: 3'-overhang bottom strand: 3' (A₄C)₄AAAAUUCGUGGCAUUUCU; blunt-end bottom strand: 3' UCGUGGCAUUUCU; top strand for both substrates: 5' AGCACCGUAAAGA. Reactions were initiated by adding 2 mM ATP-Mg²⁺. Samples were analyzed by electrophoresis in non-denaturing 15% polyacrylamide gels, which were visualized and quantified with a phosphorimager, using ImageQuant 5.2. software (Molecular Dynamics, Sunnyvale, CA).¹² The fractions of single-stranded and duplex RNAs were determined from the relative amount of radioactivity in the respective bands, and unwinding and annealing rate constants were calculated as described.¹²

Equilibrium-binding assays were done by incubating ³²P-labeled RNA (5 pM) with increasing concentrations of wild-type or mutant Mss116p at 30°C in reaction medium containing 100 mM KCl, 5 or 8 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, and 0.1 mg/ml of bovine serum albumin to stabilize Mss116p at low concentrations.²⁴ The group I and group II intron RNAs used for binding assays were prepared as described.²⁴ The single-stranded RNA oligonucleotide was 5'-labeled R-38C, and the double-stranded RNA oligonucleotide was 5' labeled R-38C annealed to the complementary strand R-38B (see above). Both the single- and double-stranded RNA oligonucleotides were gel-purified. After binding, the solutions were filtered through nitrocellulose (Bio-Rad) backed by nylon (Hybond-N; GE Healthcare Biosciences Corp.), and the radioactivity retained on the filters was quantified with a phosphorimager.

Acknowledgements

We thank Rick Russell (UT Austin) for comments on the manuscript. This work was supported by NIH R01 grants GM037951 to A.M.L. and GM06770 to E.J. M.D. is recipient of NIH fellowship F32 GM76961.

Abbreviations used

CEN plasmid, centromere-containing plasmid; Gly, glycerol; M value, mutability value; mt, mitochondrial; TCA, trichloroacetic acid.

References

1. Tanner NK, Linder P. DEXD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell* 2001;8:251–262. [PubMed: 11545728]
2. Caruthers JM, McKay DB. Helicase structure and mechanism. *Curr. Opin. Struct. Biol* 2002;12:123–133. [PubMed: 11839499]
3. Cordin O, Banroques J, Tanner NK, Linder P. The DEAD-box protein family of RNA helicases. *Gene* 2006;367:17–37. [PubMed: 16337753]

4. Linder P. Dead-box proteins: a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res* 2006;34:4168–4180. [PubMed: 16936318]
5. Jankowsky E, Fairman ME. RNA helicases--one fold for many functions. *Curr. Opin. Struct. Biol* 2007;17:316–324. [PubMed: 17574830]
6. Jankowsky E, Gross CH, Shuman S, Pyle AM. The DExH protein NPH-II is a processive and directional motor for unwinding RNA. *Nature* 2000;403:447–451. [PubMed: 10667799]
7. Pang PS, Jankowsky E, Planet PJ, Pyle AM. The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. *EMBO J* 2002;21:1168–1176. [PubMed: 11867545]
8. Serebrov V, Pyle AM. Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature* 2004;430:476–480. [PubMed: 15269774]
9. 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]2007
10. Bizebard T, Ferlenghi I, Iost I, Dreyfus M. Studies on three *E. coli* DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* 2004;43:7857–7866. [PubMed: 15196029]
11. Rogers GW Jr, Richter NJ, Merrick WC. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J. Biol. Chem* 1999;274:12236–12244. [PubMed: 10212190]
12. Yang Q, Jankowsky E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* 2005;44:13591–13601. [PubMed: 16216083]
13. Yang Q, Jankowsky E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat. Struct. Mol. Biol* 2006;13:981–986. [PubMed: 17072313]
14. Mohr S, Stryker JM, Lambowitz AM. A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing. *Cell* 2002;109:769–779. [PubMed: 12086675]
15. Del Campo M, Tijerina P, Bhaskaran H, Mohr S, Yang Q, Jankowsky E, Russell R, Lambowitz AM. Do DEAD-box proteins promote group II intron splicing without unwinding RNA? *Mol. Cell* 2007;28:159–166. [PubMed: 17936712]
16. Schneider S, Schwer B. Functional domains of the yeast splicing factor Prp22p. *J. Biol. Chem* 2001;276:21184–21191. [PubMed: 11283007]
17. Wang Y, Guthrie C. PRP16, a DEAH-box RNA helicase, is recruited to the spliceosome primarily via its nonconserved N-terminal domain. *RNA* 1998;4:1216–1229. [PubMed: 9769096]
18. Kossen K, Karginov FV, Uhlenbeck OC. The carboxy-terminal domain of the DExD/H protein YxiN is sufficient to confer specificity for 23 S rRNA. *J. Mol. Biol* 2002;324:625–636. [PubMed: 12460566]
19. Pugh GE, Nicol SM, Fuller-Pace FV. Interaction of the *Escherichia coli* DEAD box protein DbpA with 23 S ribosomal RNA. *J. Mol. Biol* 1999;292:771–778. [PubMed: 10525403]
20. Bertrand H, Bridge P, Collins RA, Garriga G, Lambowitz AM. RNA splicing in *Neurospora* mitochondria. Characterization of new nuclear mutants with defects in splicing the mitochondrial large rRNA. *Cell* 1982;29:517–526. [PubMed: 7116448]
21. Huang H-R, Rowe CE, Mohr S, Jiang Y, Lambowitz AM, Perlman PS. 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]
22. Séraphin B, Simon M, Boulet A, Faye G. Mitochondrial splicing requires a protein from a novel helicase family. *Nature* 1989;337:84–87. [PubMed: 2535893]
23. Tijerina P, Bhaskaran H, Russell R. Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc. Natl. Acad. Sci. USA* 2006;103:16698–16703. [PubMed: 17075070]
24. Halls C, Mohr S, Del Campo M, Yang Q, Jankowsky E, Lambowitz AM. Involvement of DEAD-box proteins in group I and 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]
25. Solem A, Zingler N, Pyle AM. A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol. Cell* 2006;24:611–617. [PubMed: 17188036]

26. Mohr S, Matsuura M, Perlman PS, Lambowitz AM. A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. *Proc. Natl. Acad. Sci. USA* 2006;103:3569–3574. [PubMed: 16505350]
27. Grohman JK, Del Campo M, Bhaskaran H, Tijerina P, Lambowitz AM, Russell R. Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* 2007;46:3013–3022. [PubMed: 17311413]
28. Huang, H-R. Functional studies of intron- and nuclear-encoded splicing factors in the mitochondria of *Saccharomyces cerevisiae*. The University of Texas Southwestern Medical Center; Dallas TX: 2004. Ph. D. thesis
29. Chuang RY, Weaver PL, Liu Z, Chang TH. Requirement of the DEAD-box protein Ded1p for messenger RNA translation. *Science* 1997;275:1468–1471. [PubMed: 9045610]
30. de la Cruz J, Iost I, Kressler D, Linder P. The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 1997;94:5201–5206. [PubMed: 9144215]
31. Linder P. Yeast RNA helicases of the DEAD-box family involved in translation initiation. *Biol. Cell* 2003;95:157–167. [PubMed: 12867080]
32. Carrera P, Johnstone O, Nakamura A, Casanova J, Jackle H, Lasko P. VASA mediates translation through interaction with a *Drosophila* yIF2 homolog. *Mol. Cell* 2000;5:181–187. [PubMed: 10678180]
33. 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]
34. Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* 2000;40:502–511. [PubMed: 10861942]
35. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. *Bioinformatics* 1998;14:892–893. [PubMed: 9927721]
36. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol* 2000;300:1005–1016. [PubMed: 10891285]
37. Emery B, de la Cruz J, Rocak S, Deloche O, Linder P. Has1p, a member of the DEAD-box family, is required for 40S ribosomal subunit biogenesis in *Saccharomyces cerevisiae*. *Mol. Microbiol* 2004;52:141–158. [PubMed: 15049817]
38. Liang XH, Fournier MJ. The helicase Has1p is required for snoRNA release from pre-rRNA. *Mol. Cell. Biol* 2006;26:7437–7450. [PubMed: 16908538]
39. Deminoff SJ, Tornow J, Santangelo GM. Unigenic evolution: a novel genetic method localizes a putative leucine zipper that mediates dimerization of the *Saccharomyces cerevisiae* regulator Gcr1p. *Genetics* 1995;141:1263–1274. [PubMed: 8601472]
40. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 1992;89:10915–10919. [PubMed: 1438297]
41. Cordin O, Tanner NK, Doère M, Linder P, Banroques J. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* 2004;23:2478–2487. [PubMed: 15201868]
42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
43. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol* 1982;157:105–132. [PubMed: 7108955]
44. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 1989;122:19–27. [PubMed: 2659436]
45. Kristelly R, Earnest BT, Krishnamoorthy L, Tesmer JJ. Preliminary structure analysis of the DH/PH domains of leukemia-associated RhoGEF. *Acta Crystallog. sect. D* 2003;59:1859–1862.
46. Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. A group II intron-encoded maturase functions preferentially *in cis* and requires both the reverse transcriptase and X domains to promote RNA splicing. *J. Mol. Biol* 2004;340:211–231. [PubMed: 15201048]

47. Behrsin CD, Brandl CJ, Litchfield DW, Shilton BH, Wahl LM. Development of an unbiased statistical method for the analysis of unigenic evolution. *BMC Bioinformatics* 2006;7:150. [PubMed: 16545116]
48. Schmitt ME, Brown TA, Trumpower BL. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res* 1990;18:3091–3092. [PubMed: 2190191]
49. Paukstelis PJ, Coon R, Madabusi L, Nowakowski J, Monzingo A, Robertus J, Lambowitz AM. A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface. *Mol. Cell* 2005;17:417–428. [PubMed: 15694342]

FIGURE 1.

Multiple sequence alignments and secondary structure predictions for the DEAD-box proteins Vasa, Ded1p, CYT-19, and Mss116p. CYT-19's predicted and Mss116p's experimentally determined mt targeting sequences, which are proteolytically cleaved and not present in the mature protein, are highlighted in yellow; α-helices and β-strands predicted by JPRED^{34,35} are indicated by amino acid residues in red and gold lettering, respectively; and hydrophilic C-terminal tails, as identified by hydropathy plots (see Figure 2), are highlighted in blue. Secondary structural features in the X-ray crystal structure of Vasa are shown above, and conserved DEAD-box protein motifs are boxed, with the consensus sequence indicated below. Predicted α-helices Cα1 and Cα2 in CYT-19 and Mss116p are labeled below. The location of the *cyt-19-1* mutations (G548V/L549S) and the S557P mutation in Mss116p are indicated by arrows and highlighted in green, and sites of truncation are indicated by arrows.

FIGURE 2.

Hydropathy plots for Mss116p, CYT-19, Ded1p, and Vasa. Hydrophobicity was calculated over a 20-amino acid residue sliding window using the Kyte-Doolittle algorithm.⁴³ Schematics for each protein are shown above the plot, with conserved sequence motifs indicated by black bars, and mt targeting sequences highlighted in yellow. The hydrophilic tail is highlighted in blue.

FIGURE 3.

Multiple sequence alignment of the C-terminal domains of DEAD-box proteins related to Mss116p and CYT-19. α -helices and β -strands predicted by JPRED^{34,35} are indicated by amino acid residues in red and gold lettering, respectively. Gray shading indicates conserved amino acid residues in all proteins (dark > 90% similarity, light > 65% similarity), and dark blue and green shading indicate conserved amino acid residues in Mss116p and CYT-19 homologs, respectively (>65% similarity). Abbreviations and accession numbers are: *Vasa*: *Drosophila melanogaster* Vasa (**CAA31405**); *S. cer_MSS116*: *Saccharomyces cerevisiae* MSS116 (**P15424**); *K. wal_MSS116*: *Cluyveromyces waltii* (**KLLA0DCHR**); *C. gla_CBS138*: *Candida glabrata* (**CAG59137**); *A. gos_AGL112C*: *Ashbya gossypii* (**NP_986555**); *K. lac_0DCHR*: *Cluyveromyces lactis* (**CAH00984**); *C. alb_SC5314*: *Candida albicans* (**EAL04727**); *CYT-19*: *Neurospora crassa* CYT-19 (**AAM62413**); *C. glo_04693*: *Chaetomium globosum* (**CHGG_04693**); *G. zea_FG05108*: *Gibberella zeae* (**FG05108.1**); *M. gri_367174*: *Magnaporthe grisea* (**XP_367174**); *P. nod_06545*: *Phaeosphaeria nodorum* (**SNOG_06545**); *A. ory_55384*: *Aspergillus oryzae* (**BAE55384**); *C. imm_03340*: *Coccidioides immitis* (**CIMG_03340**); *U. may_82943*: *Ustilago maydis* (**EAK82943**); *M. gri_MG10312*: *Magnaporthe grisea* (**MG10312.4**); *N. cra_DEAD*: *Neurospora crassa* (**XP_961716**); *P. nod_09314*: *Phaeosphaeria nodorum* (**SNOG_09314**); *A. tha_196479*: *Arabidopsis thaliana* (**AtNP_196479**); *S. cer_has1*: *Saccharomyces cerevisiae* has1 (**Q03532**); *A. tha_heli51*: *Arabidopsis thaliana* (**Q9LIH9**); *T. the_1027904*: *Tetrahymena thermophila* (**XP_001027904**); *D. mel_pitchoun*: *Drosophila melanogaster* pitchoun (**NP_732694**); *X. lae_68907*: *Xenopus laevis* (**XP_001027904**); *H. sap_DDX18*: *Homo sapiens* DDX18 (**NP_006764.3**).

FIGURE 4.

S. cerevisiae genetic assay for Mss116p function and unigenic evolution analysis of Mss116p's C-terminal domain. (a) Schematic of the genetic assay. Constructs are derivatives of pHRH197, a yeast CEN plasmid, which expresses Mss116p. The region of Mss116p subject to mutagenic PCR for unigenic evolution analysis is in gray, and X and S are the XbaI and SacII sites used in library constructions. A flow diagram of the genetic assay is shown below the plasmid map, with glycerol growth phenotypes expected for active or inactive Mss116 proteins expressed in *mss116* Δ strains indicated at the bottom. *mss116* Δ -I⁰, whose mtDNA lacks introns, is tested for growth on glycerol at 22°C to assay Mss116p's translational activation activity, and *mss116* Δ -1+2⁺, whose mtDNA contains 12 group I and II introns, is tested for growth on glycerol at 30°C to assay group I and group II intron splicing activity (see Materials and Methods). Active proteins support growth on glycerol (Gly⁺ phenotype), while inactive proteins do not (Gly⁻ phenotype). (b) Unigenic evolution analysis. The figure shows the amino acid sequence of Mss116p's C-terminal domain with α -helices predicted by JPRED shown above. Below the sequence are 468 amino acid substitutions resulting from missense mutations found in 111 functional variants that support group I and group II intron splicing, as judged by Gly⁺ phenotype when expressed in the *mss116* Δ -1+2⁺ strain. The functional variants were selected using the genetic assay from a library of Mss116p variants containing random PCR-induced mutations in the C-terminal domain. Asterisks indicate premature stop codons. (c) Plots of mutability values calculated as described in Materials and Methods for a 20-aa sliding window, based on 235 silent and 468 missense mutations. The plot in black was calculated with all missense mutations treated equally, and the plot in gray was calculated with conservative changes (R=K, D=E, L=I=V=F, F=Y, S=T)⁴⁰ treated as silent. Negative mutability values indicate hypomutable regions. The p value was calculated as described in Materials and Methods.

FIGURE 5.

S. cerevisiae complementation assay for translational activation and RNA splicing activity of Mss116p wild-type and C-terminally truncated proteins. CEN plasmids pHRH110 expressing myc-tagged wild-type (WT) Mss116p, Mss116p/ Δ 569-664, or Mss116p/ Δ 551-664, or the vector pRS416 were transformed into *S. cerevisiae mss116Δ* strains containing mtDNAs with different group I and group II introns: mtDNA 1⁺2⁺ contains eight group I and four group II introns; I⁰ lacks mtDNA introns; GII-5 γ contains four group I introns and the group II intron α I5 γ ; α I5 γ and bI1 contain only the group II intron α I5 γ or bI1, respectively. After growing the transformants in YNB medium at 30°C with raffinose, serial dilutions were stamped onto plates containing glycerol (YPG) or glucose (YPD) medium and incubated at the indicated temperatures.

FIGURE 6.

Northern hybridizations and correlated immunoblots comparing the ability of wild-type Mss116p and C-terminal truncation mutants to support group I and group II intron splicing *in vivo*. (a) and (b) Northern hybridizations. *S. cerevisiae mss116Δ* strains containing mtDNAs with different group I and group II introns were transformed with CEN plasmids expressing myc-tagged wild-type (WT) Mss116p, Mss116p/Δ569-664, or Mss116p/Δ551-664, or with the vector pRS416. Northern blots of RNA isolated from the indicated strains were hybridized with *COX1* (a) or *COB* (b) exon probes, as described in Materials and Methods. The blots were stripped and rehybridized with a *COX2* exon probe to confirm equal loading. Bands are labeled to the right of the blots. (c) Immunoblots of total TCA-precipitable protein isolated from the strains in (a) and (b). One gel was probed with an anti-myc antibody (Ab) preparation to detect myc-tagged Mss116p, and a parallel gel was probed with an anti-porin antibody preparation to detect the outer membrane protein porin, used as a loading control.

FIGURE 7.

Stimulation of group I and group II intron splicing reactions by the Mss116p wild-type and Δ569-664 proteins. (a) Splicing reactions with the *N. crassa* mt LSU-ΔORF group I intron. Reactions were at 25°C with 10 nM ³²P-labeled precursor RNA, 20 nM CYT-18, and 40 nM Mss116p in reaction medium containing 100 mM KCl, 5 mM Mg²⁺ and 1 mM GTP-Mg²⁺. (b) and (c) Splicing reactions with the *S. cerevisiae* ai5γ and bi1 group II introns, respectively. Reactions were at 30°C with 10 nM ³²P-labeled precursor RNA and the indicated concentrations of Mss116p in reaction medium containing 100 mM KCl and 8 mM Mg²⁺. For both group I and group II introns, 1 mM ATP-Mg²⁺, AMP-PNP-Mg²⁺, or ADP-Mg²⁺ were added as indicated. The reactions were incubated for 120 min, and the products were analyzed in a denaturing 4% polyacrylamide gel, which was dried and scanned with a phosphorimager. For each intron, all samples were analyzed on the same gel, and lane 1 is the RNA prior to incubation. In (b) and (c), the data for wild-type Mss116p were obtained in assays done in parallel with Mss116p mutants Δ569-664 and SAT/AAA and were published previously with the assays for the latter mutant.¹⁵

FIGURE 8.

Stimulation of group I and group II intron splicing reactions as a function of protein concentration for the Mss116p wild-type and Δ 569-664 mutant proteins. (a)-(f) Splicing time courses at different protein concentrations for Mss116p wild type (WT) ((a)-(c)) and Δ 569-664 ((d)-(f)). Splicing reactions were as described in Figure 7 and Materials and Methods. The plots show the disappearance of precursor RNA as a function of time, with the data fit to a single exponential equation. (g)-(i) Dependence of the observed rate constants (k_{obs}) on concentrations of Mss116p wild type (circles) and Δ 569-664 (squares). Closed and open symbols show rate constants calculated from the splicing time courses in (a)-(f) and from a repeat experiment using a different RNA transcript, respectively. For the *N. crassa* (Nc) mt

LSU-ΔORF intron (g), hyperbolic fits to the wild-type and Mss116p/Δ569-664 data gave the $k_{\text{cat}}/K_{\text{m}}$ values indicated. For the *S. cerevisiae* (Sc) aI5γ intron (h), the data for both the wild-type and mutant protein were best fit to the Hill equation (wild-type Mss116p, $k_{\text{obs}}^{\text{max}} = 0.037 \text{ min}^{-1}$, $K_{1/2} = 21 \text{ nM}$, $n = 2.7$; Mss116p/Δ569-664, $k_{\text{obs}}^{\text{max}} = 0.041 \text{ min}^{-1}$, $K_{1/2} = 43 \text{ nM}$, $n = 3.1$). Linear fits to the wild-type data from 0-40 nM and the Mss116p/Δ569-664 data from 0-80 nM (dashed lines) gave the indicated efficiency values that could be compared to $k_{\text{cat}}/K_{\text{m}}$. For the Sc bI1 intron (i), the plot shows a hyperbolic fit to the wild-type data from 0-60 nM and a sigmoidal fit to the Mss116p/Δ569-664 data ($k_{\text{obs}}^{\text{max}} = 0.030 \text{ min}^{-1}$, $K_{1/2} = 19 \text{ nM}$, $n = 2.2$). The hyperbolic fit to the wild-type data gave the indicated $k_{\text{cat}}/K_{\text{m}}$ value and a linear fit to the Mss116p/Δ569-664 data from 0-40 nM (dashed line) gave the comparable efficiency value for that protein. The data for wild-type Mss116p in (b), (c), (h), and (i) were obtained from assays done in parallel with Mss116p mutants Δ569-664 and SAT/AAA and were published previously with the data for the latter mutant.¹⁵

FIGURE 9.

RNA-unwinding assays for the Mss116p wild-type and $\Delta 569-664$ proteins. The wild-type (WT) and mutant proteins were incubated with RNA-oligonucleotide substrates containing a 13-bp duplex with ((a) and (b)) or without ((c) and (d)) a 25-nt 3'-overhang in reaction media containing 5 mM Mg²⁺ (group I splicing conditions; (a) and (c)) or 8 mM Mg²⁺ (group II splicing conditions (b) and (d)). Unwinding rate constants at the indicated protein concentrations were obtained from unwinding time courses as described¹² and plotted as a function of protein concentrations. Data points are the average for multiple independent measurements, with error bars indicating one standard deviation. For the 3'-overhang substrate, the resulting functional binding curves were fit to a sigmoidal binding isotherm. Parameters are given below the plots, with k_{unw}^{max} representing the unwinding rate constant at enzyme saturation, $K_{1/2}$ the observed functional binding constant, and n the Hill coefficient. For unwinding assays with the blunt-end substrate, for which saturation could not be obtained, $K_{1/2}$ and n values were calculated by fitting the data to a sigmoidal binding isotherm, with k_{unw}^{max} fixed at the value obtained with the overhang substrate, as found for wild-type Mss116p at lower Mg²⁺ concentrations.^{9,24} The data for wild-type Mss116p in (b) were obtained from assays done in parallel with Mss116p mutants $\Delta 569-664$ and SAT/AAA and were published previously with the data for the latter mutant.¹⁵

TABLE 1
RNA-dependent ATPase activity for the Mss116p wild-type and $\Delta 569-664$ proteins

RNA	k_{cat} (min^{-1})	
	WT	$\Delta 569-664$
No RNA	n.d. ^a	n.d. ^a
Nc mt LSU- Δ ORF	180 $\pm$ 24	195 $\pm$ 68
Sc aI5 γ	168 $\pm$ 10	208 $\pm$ 16
Sc bII	172 $\pm$ 29	208 $\pm$ 64
dsRNA	190 $\pm$ 53	154 $\pm$ 37

ATPase activity of the Mss116p wild-type (WT) and $\Delta 569-664$ proteins was assayed in reaction medium containing 100 mM KCl and 8 mM MgCl₂ at 30°C with 50 nM protein in the presence of saturating (3 mM) unlabeled ATP and RNA (300 nM for intron RNAs and 1 μ M for the double-stranded (ds) RNA oligonucleotide). k_{cat} values were determined by measuring the initial rate (r_0), corresponding to <1% of final product in a 90-min time course at saturating ATP and RNA concentrations. All reactions were linear for >10 min. Saturation was confirmed by measuring r_0 at multiple RNA concentrations. Values are the mean $\pm$ the standard deviation for at least three repeats in each case. The data for wild-type Mss116p with group I and II intron RNAs were obtained from assays done in parallel with Mss116p mutants $\Delta 569-664$ and SAT/AAA and were published previously with the data for the latter mutant.¹⁵ Similar relative activities of wild-type and mutant protein were found for the *N. crassa* mt LSU- Δ ORF group I intron in reaction medium containing 100 mM KCl and 5 mM Mg²⁺ (group I intron splicing conditions).

^a n.d., no detectable ATPase activity, $k_{\text{cat}} < 1 \text{ min}^{-1}$.

TABLE 2
Annealing-rate constants for the Mss116p wild-type and $\Delta 569$ -664 proteins

k_{ann} [$\text{M}^{-1} \text{min}^{-1}$]				
5 mM Mg^{2+}		8 mM Mg^{2+}		
Protein	3' Overhang	Blunt	3' Overhang	Blunt
WT	$16 \pm 1.7 \times 10^9$	$5.2 \pm 0.8 \times 10^9$	$9.6 \pm 1.7 \times 10^9$	$0.47 \pm 0.05 \times 10^9$
$\Delta 569$ -664	$8.4 \pm 2.7 \times 10^9$	$0.7 \pm 0.6 \times 10^9$	$1.7 \pm 0.4 \times 10^9$	$0.16 \pm 0.02 \times 10^9$

The table summarizes bimolecular rate constants for strand-annealing of RNA-oligonucleotide substrates by the Mss116p wild-type and $\Delta 569$ -664 proteins in reaction media containing 5 mM Mg^{2+} (group I splicing conditions) or 8 mM Mg^{2+} (group II splicing conditions). The rate constants were calculated from unwinding time courses at 400 nM protein as described.¹² Values are the mean $\pm$ the standard deviation for multiple repeats.

TABLE 3

Binding of Mss116p wild-type and mutant proteins to group I and group II intron RNAs

Protein	K_d or $K_{1/2}$, nM		Nc mt LSU-AORF	Sc al5 γ		Sc bII
WT						
- NTP	1.6 $\pm$ 0.6			2.2 $\pm$ 1.6		2.0 $\pm$ 1.0
+ATP	2.0 $\pm$ 1.0			2.2 $\pm$ 1.6		1.6 $\pm$ 0.6
+ADP	2.0 $\pm$ 1.0			2.3 $\pm$ 1.5		2.3 $\pm$ 1.5
+AMP-PNP	1.8 $\pm$ 1.6			0.8 $\pm$ 0.6		1.6 $\pm$ 0.6
Δ 569-664						
- NTP	17 $\pm$ 8			19 $\pm$ 3		16 $\pm$ 2
+ATP	18 $\pm$ 9			19 $\pm$ 4		17 $\pm$ 6
+ADP	20 $\pm$ 0			20 $\pm$ 0		20 $\pm$ 9
+AMP-PNP	20 $\pm$ 0			18 $\pm$ 3		15 $\pm$ 5
Δ 551-664					n	n
- NTP	215 $\pm$ 47		3.3 $\pm$ 1.9	150 $\pm$ 3	2.6 $\pm$ 0.5	211 $\pm$ 11
+ATP	193 $\pm$ 52		2.1 $\pm$ 0.8	243 $\pm$ 17	2.0 $\pm$ 0.8	231 $\pm$ 70
+ADP	178 $\pm$ 15		2.9 $\pm$ 0.9	210 $\pm$ 10	2.4 $\pm$ 0.8	231 $\pm$ 59
+AMP-PNP	192 $\pm$ 4		2.9 $\pm$ 0.7	140 $\pm$ 1	2.8 $\pm$ 0.6	177 $\pm$ 13
						3.4 $\pm$ 1.4
						2.9 $\pm$ 0.1
						3.4 $\pm$ 1.6
						3.0 $\pm$ 1.3

For equilibrium nitrocellulose-filter binding assays, 32 P-labeled RNAs (5 pM) were incubated with increasing concentrations of Mss116p wild-type or mutant proteins in reaction medium containing 100 mM KCl and 8 mM MgCl₂ for 90 min at 30°C without nucleotide (-NTP), or with 1 mM AMP-PNP-Mg²⁺, ADP-Mg²⁺, or ATP-Mg²⁺. RNAs were *in vitro* transcripts containing the *N. crassa* (Nc) mt LSU-AORF group I intron or the *S. cerevisiae* (Sc) al5 γ or bII group II introns. The percent input RNA bound was plotted as a function of protein concentration, and the plots were best-fit by a hyperbola (wild-type and the Δ 569-664 mutant protein) or the Hill equation (Δ 551-664 mutant protein) to obtain K_d or $K_{1/2}$ values, respectively. Values are the mean $\pm$ the standard deviation for at least three repeats in each case. For each condition for the wild-type protein and in representative cases for the mutant proteins, we confirmed that essentially identical binding curves were obtained when the RNA and protein were incubated for 10 or 120 min before filtration through nitrocellulose, or in reaction medium containing 100 mM KCl and 5 mM MgCl₂ (data not shown).

TABLE 4

Binding of the Mss116p wild-type and $\Delta 569-664$ proteins to single-stranded and double-stranded RNA oligonucleotides

Protein	K_d [nM]	
	Ss RNA	ds RNA
WT		
- NTP	28 ± 7	24 ± 5
+ATP	19 ± 9	17 ± 4
+ADP	22 ± 7	21 ± 12
+AMP-PNP	2.0 ± 0.6	21 ± 11
$\Delta 569-664$		
- NTP	13 ± 3	9 ± 1
+ATP	12 ± 3	20 ± 14
+ADP	15 ± 5	11 ± 1
+AMP-PNP	11 ± 1	11 ± 1

Equilibrium nitrocellulose-filter binding assays were done as described in the legend of Table 3. RNAs were single-stranded (ss) and double-stranded (ds) RNA oligonucleotides (see Materials and Methods). K_d values were obtained from plots of the percent input RNA bound as a function of protein concentration fit by a hyperbola. Values are the mean $\pm$ the standard deviation for at least three repeats in each case.