Electronic Supplementary Materials

DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA

Experimental Methods

Yeast strains, plasmids, oligonucleotides, and recombinant protein purification

The *S. cerevisiae* strain YPH499 was purchased from ATCC (ATCC204679). The pET28a-DBP2 plasmid was a gift from Dr. Elizabeth Tran¹. Oligonucleotides (Supplementary Table S1) were used to amplify *MSS116* using purified yeast genomic DNA from strain S288C as the template. The PCR products were purified and subjected to standard molecular cloning protocols as previously described² to construct the pSUMO-MSS116 plasmid (which encodes the mature mitochondrial version of Mss116, amino acids 37-664). All plasmids were confirmed by DNA sequencing at the UAMS DNA Sequencing core facility.

DNA and RNA oligonucleotides (Supplementary Table S1) were purchased from Integrated DNA Technologies, and Dharmacon, respectively. All G4 forming oligonucleotides were purified as described³, and G4DNA formation was confirmed by circular dichroism³ in the presence of 100mM KCI. All other oligonucleotides were purified as described⁴. Oligonucleotides were quantified by absorbance at 260 nm using calculated extinction coefficients⁵.

Recombinant Dbp2¹ and Ded1⁶ were purified as previously described.

Recombinant Mss116 Protein Purification --- Purified pSUMO-MSS116 plasmid was transformed into Rosetta[™] 2 (DE3) Competent *E. coli* (EMD Millipore) and plated on LB agar plates containing 50 µg/ml kanamycin and 25 µg/ml chloramphenicol for selection at 37 °C overnight. Clones were inoculated into 125 ml LB media containing 50 µg/ml kanamycin and 25 µg/ml chloramphenicol, and grown at 37 °C to an OD₆₀₀ around 1. Cultures were transferred into 6 L of LB media containing 50 µg/ml kanamycin and 25 μ g/ml chloramphenicol and grown at 37 °C to OD₆₀₀ = 0.8. IPTG was added to a final concentration of 1 mM to induce recombinant protein expression at 20°C for 19 hours. Cells were harvested by centrifugation at 3000x g for 30 minutes at 4 °C. All subsequent steps in this protein purification were performed at 4°C. Cell pellets were re-suspended and homogenized in lysis buffer (50 mM NaH₂PO₄, 500 mM KCl, 10 mM imidazole, 10% glycerol, 0.1 g/ml lysozyme, 1 mM PMSF, 1 μM pepstatin, 1 mM β-mercaptoethanol, pH 7.5) at a volume of 10 ml per gram of cells, passed through a Microfluidizer (Microfluidics Inc.), followed by sonication. Cell lysates were pelleted by ultra-centrifugation at 148,000x g for 90 minutes. The supernant was loaded onto a Ni-NTA Agarose (MCLAB) column, washed with buffer (50 mM NaH₂PO₄, 500 mM KCl, 20 mM imidazole, 10% glycerol, 1 mM β-mercaptoethanol, pH 7.5), and eluted using imidazole buffer (50 mM NaH₂PO₄, 500 mM KCl, 250 mM imidazole, 10% glycerol, 1 mM β-mercaptoethanol, pH 7.5). Fractions containing SUMO-Mss116 were pooled and dialyzed to 50 mM NaH₂PO₄, 500 mM KCl, 10 mM imidazole, 10% glycerol, 1 mM β-mercaptoethanol, pH 7.5. During dialysis, Ulp1 protease was added to cleave the SUMO tag. Complete Ulp1 digestion was confirmed by SDS-PAGE. The dialyzed samples were loaded onto a Ni-NTA agarose column, and the flow-through was collected in fractions. Fractions containing

purified recombinant Mss116 were pooled and concentrated using Amicon Ultracel Centrifugal Filters (Millipore), followed by dialysis into storage buffer (25 mM HEPES, 500 mM KCl, 20% glycerol, 1 mM β-mercaptoethanol, pH 7.5).

Dynabead pull-down for identifying G-quadruplex binding proteins

Conjugation of DNA to Dynabeads, affinity pull-down of interacting proteins from yeast whole cell lysate, LC-MS/MS analysis, and identification of G4-interacting proteins using spectral counting, were performed as previously described². Briefly, 3'-biotinylated G4DNA or ssDNA conjugated to Streptavidin M-280 Dynabeads[®] was incubated with YPH499 yeast cell lysates. After extensive washing, bound proteins were eluted in Laemmli Sample Buffer, separated by SDS-PAGE, and subjected to LC-MS/MS analysis. G4DNA interacting proteins were determined by spectral counting^{7,8} and G-test of independence⁹. A value of 0 in spectral counting was replaced with 0.1 in each replicate to allow the log of the value to be calculated in subsequent calculations. Proteins that were significant at the 5% significance level (after Benjamini-Hochberg correction for multiple testing) in each of the two replicates were classified as G4DNA interacting proteins.

MEME motif finder tool analysis

G4DNA binding motif was identified using the MEME motif finder tool (<u>http://meme-suite.org/tools/meme</u>)^{10,11}. Command string and parameters: "meme string_protein_sequences.fa -protein -oc . -nostatus -time 18000 -maxsize 60000 -mod anr -nmotifs 4 -minw 6 -maxw 50".

Fluorescence anisotropy binding assay

Fluorescein labeled DNA or RNA (Supplementary Table S1), at the concentrations indicated in the figure legends, was incubated with varying concentrations of protein at room temperature for 30 minutes in Dbp2 buffer [40 mM Tris, pH 8.0, 50 mM KCl, 2 mM β-mercaptoethanol (BME), 0.1 mM EDTA, 0.5 mg/ml bovine serum albumin (BSA)], Ded1 buffer [40 mM Tris pH 8.0, 50 mM KCl, 2 mM DTT, 0.1 mM EDTA, 0.05% NP-40], or Mss116 buffer [10 mM Hepes pH 7,5, 200 mM KCl, 1 mM EDTA, 1 mM DTT, 0.01 mg/ml BSA]. Polarization was measured at 535 nm using a Perkin Elmer 1420 Victor³V Multi-label Counter with excitation at 485 nm. Data was converted to anisotropy and fit to the quadratic equation to determine the K_d value using KaleidaGraph software.

G-quadruplex DNA unfolding

DNA (Supplementary Table S1) was radiolabeled as described⁴. All concentrations listed are final. G4DNA (2 nM) was pre-incubated with helicase (400 nM) in reaction buffer [Dbp2 (40 mM Tris, pH 8.0, 50 mM KCl, 2 mM BME, 0.1 mM EDTA, 0.5 mg/ml BSA), Ded1 (40mM Tris pH 8.0, 50 mM KCl, 2 mM DTT, 0.1 mM EDTA, 0.01% NP-40), Mss116 (20 mM Hepes pH 7.5, 100 mM KCl, 1 mM DTT, 0.1 mM EDTA, 0.1 mg/ml BSA, 10% glycerol), or Pif1 (25 mM Hepes, pH 7.5, 50 mM KCl, 2 mM β -mercaptoethanol, 0.1 mM EDTA, 0.1 mg/ml BSA). Reactions were initiated by addition of 5 mM ATP, 10 mM MgCl₂, and DNA trap complementary to the G4-forming

region (60 nM) for trapping the unfolded DNA as described³. Reactions were performed at 25 °C. Aliquots were removed and quenched at increasing times with 100 mM EDTA, 0.3% SDS 0.1 mM (in nucleotides) T_{50} protein trap, 0.9 µM unlabeled G4DNA substrate, 0.1% bromophenol blue, 0.1% xylene cyanol, 6% glycerol. Samples were separated by 20% native PAGE, visualized using a Typhoon Trio PhosphorImager, and quantitated using ImageQuant software.

Bioinformatic analysis for identifying G4RNAs in yeast transcriptome

Potential quadruplex forming sequences in the *S. cerevisiae* transcriptome were identified using EuQuad (http://quadbase.igib.res.in/EuQuad). Search parameters were: G4 motif, G₂N₁₋₇G₂N₁₋₇G₂N₁₋₇G₂; algorithm, non-greedy; strands, + strand; bulge size, 0.

G-quadruplex RNA unfolding

RNA (Supplementary Table S1) was radiolabeled as described⁴. All concentrations listed are final. G4RNA was folded by heating to 95°C for 10 minutes in 100mM KCl and slowly cooling to room temperature. G4RNA (2 nM) was pre-incubated with 5 mM ATP and 10 mM MgCl₂, in the same reaction buffers used for G4DNA unfolding experiments. Reactions were initiated by addition of helicase (400 nM) and DNA trap (Q-trap) complementary to the G4-forming region (15 nM) for trapping the unfolded RNA as described³. Reactions were performed at 25 °C. Aliquots were removed and quenched at various times with 50 mM EDTA, 0.5% SDS, and 0.9 μM C-

trap (unlabeled G4DNA). Samples were separated by 20% native PAGE, visualized using a Typhoon Trio PhosphorImager, and quantitated using ImageQuant software.

RNA:DNA hybrid duplex unwinding

RNA (Supplementary Table 1) was radiolabeled as described. All concentrations listed are final. RNA:DNA hybrid duplex was formed by mixing the RNA strand with the complementary DNA strand at 1:1.2 ratio, heating to 95°C for 5 minutes, and slowly cooling to room temperature. The hybrid duplex (2nM) was pre-incubated with 5 mM ATP and 10 mM MgCl₂, in the reaction buffer that corresponded to the same buffer used in the G4DNA unfolding experiments. Reactions were initiated by addition of helicase (400 nM) and DNA trap (60 nM) complementary to the displaced DNA strand. Reactions were performed at 25 °C. Aliquots were removed and quenched at various times with 50 mM EDTA, 0.5% SDS. Samples were separated by 20% native PAGE, visualized using a Typhoon Trio PhosphorImager, and quantitated using ImageQuant software.

References in Supplementary Materials

- 1 S. C. Cloutier, W. K. Ma, L. T. Nguyen and E. J. Tran, *J. Biol. Chem.*, 2012, **287**, 26155–26166.
- 2 J. Gao, B. L. Zybailov, A. K. Byrd, W. C. Griffin, S. Chib, S. G. Mackintosh, A. J. Tackett and K. D. Raney, *Chem. Commun. (Camb.)*, 2015, **51**, 7242–7244.
- 3 A. K. Byrd and K. D. Raney, *J. Biol. Chem.*, 2015, **290**, 6482–6494.
- 4 P. D. Morris, A. J. Tackett, K. Babb, B. Nanduri, C. Chick, J. Scott and K. D. Raney, *J. Biol. Chem.*, 2001, **276**, 19691–19698.
- 5 S. S. Carroll, F. Benseler and D. B. Olsen, *Meth. Enzymol.*, 1996, **275**, 365–382.
- 6 Q. Yang and E. Jankowsky, *Biochemistry*, 2005, **44**, 13591–13601.
- 7 B. Zybailov, M. K. Coleman, L. Florens and M. P. Washburn, *Anal. Chem.*, 2005, **77**, 6218–6224.
- 8 B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens and M. P. Washburn, *J. Proteome Res.*, 2006, **5**, 2339–2347.

- 9 B. Zybailov, G. Friso, J. Kim, A. Rudella, V. R. Rodríguez, Y. Asakura, Q. Sun and K. J. van Wijk, Mol. Cell Proteomics, 2009, 8, 1789–1810.
- 10 T. L. Bailey and M. Gribskov, *Bioinformatics*, 1998, 14, 48–54.
 11 T. L. Bailey, M. Bodén, T. Whitington and P. Machanick, *BMC Bioinformatics*, 2010, **11**, 179.

| Name | Sequence | |
|---------------------------------|---|--|
| | | |
| PCR | | |
| Mss116 For | 5 ′ – ATTGCACGTCTCAAGGTTTATACAATGATGGAAACAGAGATCAA | |
| | AGAAATTTTGGTAGG-3' | |
| Mss116 Rev | 5 ' -ACAACGCGTCTCCTCGAGTTACTAATATATGTTGCTGTTTCTAC | |
| | TGGAG-3 ' | |
| | | |
| G4 Binding | | |
| Tailless cMYC | 5′-T <u>GGG</u> T <u>GGG</u> TA <u>GGG</u> T <u>GGG</u> TTT-6FAM-3′ | |
| 5'-tailed-cMYC | 5′-6FAM-TTTTTTTTT <u>GGG</u> T <u>GGG</u> TA <u>GGG</u> T-3′ | |
| Tailless hTEL | 5′-A <u>GGG</u> TTA <u>GGG</u> TTA <u>GGG</u> TT-6FAM-3′ | |
| Tailless ScTEL | 5′-T <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TT-6FAM-3′ | |
| Blunt duplex | 5'-CGCTGATGTCGCCTGG-3' | |
| · | 3'-6FAM-GCGACTACAGCGGACC-5' | |
| ssDNA | 5′-6FAM-T ₂₀ -3′ | |
| YNL098C | 5'-GUUGGUGGUGGUGU-6FAM-3' | |
| ssRNA | 5′-U ₂₀ -6FAM-3′ | |
| | | |
| G4DNA unfolding | | |
| 5'-tailed G4DNA | 5'-T ₁₅ GAGGGTGGGTAGGGTGGGTAA-3' | |
| 3'-tailed G4DNA | 5'-TGAGGGTGGGTAGGGTGGGTAAT14-3' | |
| DNA Annealing Trap | 5'-TTACCCACCCTACCCACCCTCA-3' | |
| | | |
| G4RNA unfolding | | |
| Tailless G4RNA | 5'- GUUGGUGGUGGUGGUGU-3' | |
| 3'-tailed G4RNA | 5′–GUUGGUGGUGGUGU13–3′ | |
| Otrap DNA | 5'- ACCACCACCA-3' | |
| Otrap DNA-1 | 5' - ACACCACCACCAAC-3' | |
| Ctrap DNA | 5' - GTTGGTGGTGGTGGTGT-3' | |
| | | |
| RNA DNA hybrid duplex unwinding | | |
| RNA | $5' - GUUGGUGGUGGUGGUGGUGU_2 - 3'$ | |
| Complementary DNA | 3' - CAACCACCACCACCA-5' | |
| DNA trap | 5'-GTTGGTGGTGGTGGTGT-3' | |
| Вникар | | |
| G4RNA CD | | |
| | 5' - GUUGGUGGUGGUGGUGU -3' | |
| YEL066C | 3' - IIGIGGCAGUACAGGCCCCACCGGCUAUGGUAA - 5' | |
| VPR204W | 5' - 100000000000000000000000000000000000 | |
| | 5 - 5 | |
| Cuppings involved in tet | and formation are underlined | |

Supplementary Table S1. Sequences of oligonucleotides.

Guanines involved in tetrad formation are underlined.

Duplex forming regions are italicized.

Supplementary Table S2. Full list of the mass spectrometry excel table.

Supplementary Table S3. Most frequently occurring potential two-tetrad G4RNA sequences in yeast.

| Sequence | Number of occurrences |
|---|-----------------------|
| <u>GGUGGUGG</u> UGG | 23 |
| <u>GG</u> CAGUACA <u>GG</u> CCCCACC <u>GG</u> GCUAU <u>GG</u> | 19 |
| <u>GG</u> U <u>GG</u> UAU <u>GG</u> AUACAC <u>GG</u> | 18 |
| <u>GG</u> CA <u>GG</u> UGA <u>GG</u> C <u>GG</u> | 16 |
| <u>GG</u> UGCAAU <u>GG</u> AUA <u>GG</u> AAAA <u>GG</u> | 15 |
| <u>GG</u> A <u>GG</u> AU <u>GG</u> CUU <u>GG</u> | 15 |
| <u>GG</u> AAA <u>GG</u> UAAUU <u>GG</u> A <u>GG</u> | 14 |
| <u>GG</u> UCCCCGU <u>GG</u> U <u>GG</u> CGCU <u>G</u> G | 12 |
| <u>GG</u> AU <u>GG</u> AUGU <u>GG</u> U <u>GG</u> | 11 |
| <u>GG</u> AGUC <u>GG</u> UACUUUC <u>GG</u> U <u>GG</u> | 10 |
| <u>GGUGGUACGGAGG</u> | 10 |

Guanines involved in tetrad formation are underlined.

Sequences used in this study are highlighted.

Supplementary Figure Legends

Supplementary Figure S1. The RGG motif is one of potential G-quadruplex binding motifs.

A. Diagram of three top scoring motifs found in the MEME analysis of the proteins listed in Table 1. MEME-1 is the RGG motif. MEME-2 and MEME-3 are DEAD-box helicase motifs. **B.** Block diagram of the locations of these motifs in the proteins.

Supplementary Figure S2. Circular dichroism analysis of G4DNA or G4RNA substrates.

A. Circular dichroism spectra of the folded G4DNA substrates. **B.** Circular dichroism spectra of the folded two-tetrad G4RNAs. A minimum near 240 nm and a maxima near 260 nm indicates formation of a parallel G4. A minimum near 240 nm and a maxima near 290 nm indicates formation of a G4 structure that is a hybrid of parallel and antiparallel structures. **C.** Circular dichroism spectra of the two RNA substrates used in the subsequent binding assays and unfolding assays.

Supplementary Figure S3. Dbp2, Ded1 and Mss116 do not unfold G4DNA.

A. Helicases were pre-incubated with G4DNA. ATP and Mg²⁺ were added to initiate the reaction. At increasing times, the reactions were quenched by adding excess unlabeled substrate (Ctrap) with 100mM EDTA and 1% SDS. **B.** No unfolding activity of a 5'-tailed cMYC G4DNA by Dbp2, Ded1, or Mss116 was observed. *S. cerevisiae* Pif1 helicase served as a positive control. **C.** No unfolding activity of a 3'-tailed cMYC G4DNA by Dbp2, Ded1, or Mss116 was observed.

Supplementary Figure S4. Mss116 binding to single-stranded RNA (ssRNA).

Binding of Mss116 to ssRNA at higher protein concentration. Data were fit to the Hill equation. Error bars represent the standard deviation of three independent experiments.

Supplementary Figure S5. Dbp2, Ded1, and Mss116 destabilize a tailless G4RNA structure.

The representative gel images of G4RNA unfolding assays by Dbp2 (**A**), Ded1 (**B**), and Mss116 (**C**).

Supplementary Figure S6. Dbp2, Ded1 and Mss116 destabilize a 3'-tailed G4RNA structure.

A. Diagram illustrating the experiments. ³²P-labelled 3'-tailed G4RNA (2 nM) was preincubated with Mg²⁺ in the presence or absence of ATP. Helicases (400 nM) were added to initiate the reaction, along with 15 nM Qtrap DNA (or Qtrap DNA-1 in Ded1 reactions). The destabilized G4RNA will be single-stranded and rapidly complemented with Qtrap to form a stable RNA:DNA duplex (blue and black duplex). At increasing times, the reactions were quenched by adding excess Ctrap (0.9 µM) with 100mM EDTA and 1% SDS. The Ctrap will form a duplex with the leftover Qtrap (blue and green duplex). (**B-D**) 3'-tailed G4RNA was unfolded by Dbp2 (**B**), Ded1 (**C**) or Mss116 (**D**). Representative gel images are in the upper panel. The lower panel is the graph of the reaction progress curves. Data were fit to a single exponential. The rate constants represent the average and standard deviation from three independent experiments.

Supplementary Figure S7. Dbp2, Ded1 and Mss116 unwind a RNA:DNA hybrid duplex in an ATP-dependent manner.

A. Diagram illustrating the experiments. ³²P-labelled 3'-tailed RNA:DNA hybrid duplexes (2 nM) were pre-incubated with Mg²⁺, in the presence or absence of ATP. Helicases (400 nM) were added to initiate the reaction along with a DNA trap (60 nM). At increasing times, the reactions were quenched by adding 100mM EDTA and 1% SDS. (B-D) RNA:DNA hybrid duplex was unfolded by Dbp2 (B), Ded1 (C) or Mss116 (D). Representative gel images are in the left panel. The right panel is the kinetic graphics of the reactions. Data were fit to a single exponential. The rate constant represents the average and standard deviation from three independent experiments.



В



MEME-3 (GCDLLVATPGRLNDLLERGKVSLANVKYLVLDEADRMLDMGFEPQIRHIV)

Supplementary Figure S1. The RGG motif is one of potential Gquadruplex binding motifs.

A. Diagram of three top scoring motifs found in the MEME analysis by the amino acid sequences of proteins listed in Table 1. MEME-1 is the RGG motif. MEME-2 and MEME-3 are DEAD-box helicase motifs. **B**. Block diagram of the locations of these motifs in the proteins.



Supplementary Figure S2. Circular dichroism analysis of G4DNA or G4RNA substrates.

A. Circular dichroism spectra of the folded G4DNA substrates. **B.** Circular dichroism spectra of the folded two-tetrad G4RNAs. A minimum near 240 nm and a maxima near 260 nm indicates formation of a parallel G4. A minimum near 240 nm and a maxima near 290 nm indicates formation of a G4 structure that is a hybrid of parallel and anti-parallel structures. **C.** Circular dichroism spectra of the two RNA substrates used in the subsequent binding assays and unfolding assays.

Supplementary Figure S3



B 5'-tailed G4DNA



C 3'-tailed G4DNA



Supplementary Figure S3. Dbp2, Ded1 and Mss116 do not unfold G4DNA.

A. Helicases were pre-incubated with G4DNA. ATP and Mg²⁺ were added to initiate the reaction. At increasing times, the reactions were quenched by adding excess unlabeled substrate (Ctrap) with 100mM EDTA and 1% SDS. **B.** No unfolding activity of a 5'-tailed cMYC G4DNA by Dbp2, Ded1, or Mss116 was observed. *S. cerevisiae* Pif1 helicase served as a positive control. **C.** No unfolding activity of a 3'-tailed cMYC G4DNA by Dbp2, Ded1, or Mss116 was observed.



Supplementary Figure S4. Mss116 binding to single-stranded RNA (ssRNA).

Binding of Mss116 to ssRNA at higher protein concentration. Data were fit to the Hill equation. Error bars represent the standard deviation of three independent experiments.

Supplementary Figure S5

B. tailless G4RNA unfolded by Ded1



C. tailless G4RNA unfolded by Mss116



Supplementary Figure S5. Dbp2, Ded1, and Mss116 destabilize a tailless G4RNA structure.

The representative gel images of G4RNA unfolding assays by Dbp2 (A), Ded1 (B), and Mss116 (C).



Supplementary Figure S6 (continue)



D 3'-tailed G4RNA unfolded by Mss116

Supplementary Figure S6. Dbp2, Ded1 and Mss116 destabilize a 3'-tailed G4RNA structure.

A. Diagram illustrating the experiments. ³²P-labelled 3'-tailed G4RNA (2 nM) was pre-incubated with Mg²⁺ in the presence or absence of ATP. Helicases (400 nM) were added to initiate the reaction, along with 15 nM Qtrap DNA (or Qtrap DNA-1 in Ded1 reactions). The destabilized G4RNA will be single-stranded and rapidly complemented with Qtrap to form a stable RNA:DNA duplex (blue and black duplex). At increasing times, the reactions were quenched by adding excess Ctrap (0.9 μ M) with 100mM EDTA and 1% SDS. The Ctrap will form a duplex with the leftover Qtrap (blue and green duplex). (**B-D**) 3'-tailed G4RNA was unfolded by Dbp2 (**B**), Ded1 (**C**) or Mss116 (**D**). Representative gel images are in the upper panel. The lower panel is the graph of the reaction progress curves. Data were fit to a single exponential. The rate constants represent the average and standard deviation from three independent experiments.

Supplementary Figure S7



B 3'-tailed RNA/DNA hybrid duplex unwound by Dbp2



C 3'-tailed RNA/DNA hybrid duplex unfolded by Ded1



D 3'-tailed RNA/DNA hybrid duplex unwound by Mss116



Supplementary Figure S7. Dbp2, Ded1 and Mss116 unwind a RNA:DNA hybrid duplex in an ATP-dependent manner.

A. Diagram illustrating the experiments. ³²P-labelled 3'-tailed RNA:DNA hybrid duplexes (2 nM) were pre-incubated with Mg²⁺, in the presence or absence of ATP. Helicases (400 nM) were added to initiate the reaction along with a DNA trap (60 nM). At increasing times, the reactions were quenched by adding 100mM EDTA and 1% SDS. (B-D) RNA:DNA hybrid duplex was unfolded by Dbp2 (B), Ded1 (C) or Mss116 (D). Representative gel images are in the left panel. The right panel is the kinetic graphics of the reactions. Data were fit to a single exponential. The rate constant represents the average and standard deviation from three independent experiments.