Supporting Information

Mallam et al. 10.1073/pnas.1109566108

SI Methods

Protein Expression and Purification. Full-length Mss116p [excluding its mitochondrial (mt) targeting sequence (residues 1–36)] was cloned in pMAL-c2t (a derivative of pMAL-c2x; New England Biolabs) and expressed as an N-terminal fusion to maltose-binding protein (MBP), as described (1, 2). Derivatives of pMAL-Mss116p that express the deletion proteins Mss116p/ΔNTE (deletion of N-terminal extension; residues 37–87), Mss116p/ΔC-tail (deletion of C-terminal tail; residues 597–664), and Mss116p/ΔNTE + ΔC-tail (deletion of both the N-terminal extension and C-terminal tail) were created as described (2). The proteins were expressed in *Escherichia coli* Rosetta 2 (EMD Biosciences) grown in ZYP-5052 auto-inducing medium for 24 h at 22 °C and purified at 4 °C, as described (2). Purification steps included (*i*) removal of nucleic acids by polyethyleneimine precipitation; (*ii*) isolation of MBP-Mss116p by amylose affinity chromatography (New England Biolabs); (*iii*) removal of the MBP tag by digestion with tobacco etch virus (TEV) protease; and (*iv*) isolation of Mss116p by affinity chromatography, using a heparin-Sepharose column (GE Healthcare). Proteins were dialyzed into small-angle X-ray scattering (SAXS) buffer [20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT)]. Protein concentrations were determined by Bradford Assay (BioRad Laboratories) and by absorbance at 280 nm. Extinction coefficients were calculated from the protein sequence by using the ExPASy Proteomics Server ProtParam tool (3).

Full-length CYT-19 and CYT-19/ΔC-tail (deletion of residues 578–626) were expressed and purified as MalE fusions using a similar strategy, as described (4). Prior to SAXS experiments, CYT-19 and CYT-19/ΔC-tail were concentrated in a 30-kDa molecular weight cut off (MWCO) concentrator (Millipore) and rediluted into a buffer suitable for SAXS measurements (20 mM Tris-HCl (pH 7.5), 200 mM KCl, 5 mM $MgCl_2$, 20% glycerol, 0.4 mM EDTA, 0.2 mM DTT).

SAXS Sample Preparation and Characterization. The oligonucleotides U_{10}-RNA, RNA–DNA duplex 1 and RNA–DNA duplex 2 (Fig. 1*C*) were synthesized by Integrated DNA Technologies and reconstituted in sterile, double-deionized water to 1 mM. The chimeric substrates RNA–DNA duplex 1 and RNA–DNA duplex 2 were annealed by heating to 94 °C for 1 min and slowly cooling to room temperature over 1 h. Oligonucleotide purity was confirmed by size-exclusion chromatography (SEC) (Fig. S5). The equilibrium dissociation constant for the complex between Mss116p/ΔNTE and a fluorescein-labeled U_{10}-RNA (FAM-U_{10}-RNA) was measured under the SAXS buffer conditions (Fig. S4). This K_d was in the nM range and indicates complete binding of ssRNA at the concentrations of protein necessary for SAXS measurements (30–70 μM). Complexes (approximately 50 μM) for analysis by SAXS were assembled from a mixture of 1:1.2–1.5 protein:oligonucleotide in the presence of 0.5–1.0 mM ADP-BeF_x and were formed at room temperature for at least 10 min. ADP-BeF_x was prepared as described (5).

All proteins and complexes were purified to homogeneity by SEC on a BioLogic DuoFlow or an AKTA FPLC using a Superdex 200 10/300 GL analytical gel filtration column equilibrated in the SAXS buffer (20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 5 mM $MgCl_2$, 1 mM DTT for Mss116p or the same buffer with 200 mM KCl for CYT-19). Complex formation was assessed by the A_{260}/A_{280} ratio (Fig. S5). The lack of dissociation observed when purified complexes were reanalyzed by SEC confirmed complex stability (Fig. S5). The relative elution volumes of the Mss116p SAXS samples were compared to those of molecular weight standards to estimate their hydrodynamic radii, R_H. The relative elution volume was calculated as

$$K_{\mathrm{AV}} = \frac{V_e - V_o}{V_g - V_o}, \qquad \text{[S1]}$$

where V_e is the elution volume, V_o is the void volume determined by the elution of Blue Dextran 2,000 (Mr = 2,000 kDa), and V_g is the geometric column volume determined by the elution of free tyrosine (Mr = 0.13 kDa). A standard curve was plotted of K_{AV} versus $\log(R_H)$ (Fig. S5). Molecular weight standards were thyroglobulin (R_H = 85 Å), apoferritin (R_H = 61 Å), β-amylase (R_H = 50.4 Å), alcohol dehydrogenase (R_H = 46 Å), bovine serum albumin (R_H = 35.5 Å), carbonic anhydrase (R_H = 24.3 Å), and cytochrome c (R_H = 17 Å).

Proteins and complexes were concentrated by centrifugation at 4 °C in a 30-kDa MWCO concentrator (Millipore) and concentrations were determined as above. SAXS sample concentrations were 1–3 mg/mL. Care was taken to subject proteins and complexes to only one freeze-thaw cycle before SAXS measurements. Additional complexes with U_{10}-RNA and ADP-BeF_x were assembled at the synchrotron in the presence of excess RNA and nucleotide. In these cases, the flow-through from centrifugation in a 30-kDa MWCO concentrator was used as a buffer blank. Scattering profiles and R_g values for these complexes were in good agreement with those purified by SEC.

Far UltraViolet Circular Dichroism. All measurements were performed in SAXS buffer using a thermostatically controlled 0.01-cm path-length cuvette at 25 °C and a Jasco J-815 spectrometer. Scans were taken between 175 and 260 nm at a scan rate of 0.5 nm s^{-1} with 30 accumulations. Sample concentrations were 1.1–2.5 mg/mL. Spectra were analyzed by the CDSSTR analysis program (6), using the DichroWeb online circular dichroism analysis website (7).

SAXS Data Collection. Synchrotron radiation X-ray scattering data were collected at the Advanced Photon Source beamlines 12-ID-C and 18-ID-D and recorded on two-dimensional CCD detectors. Twenty exposures of 1 s were acquired for each sample at a sample-to-detector distance of 2.0 m and over a range of momentum transfer $0.007 < q < 0.35$ $Å^{-1}$, $q = 4\pi \sin(\theta)/\lambda$. No measurable radiation damage was observed under these conditions. Scattering data were radially averaged to produce one-dimensional profiles of scattering intensity vs. q. Data and a buffer blank were collected for three different concentrations of each sample to check for concentration-dependent scattering effects, such as aggregation or interparticle interference. The scattering intensity at zero-angle ($I(0)$) was calibrated against known concentrations of protein and RNA standards (cytochrome c and U_{10}-RNA, respectively).

SAXS Data Analysis. Analysis of the SAXS data was performed with IGOR-Pro (WaveMetrics) and ATSAS (version 2.4) software. Background-subtracted scattering data were subject to an initial inspection to determine sample monodispersity. Plots of $\log[I(q)]$ versus $\log(q)$ displayed a flat region in the lowest q regime, consistent with the presence of monodisperse particles (8). The zero-angle scattering intensity, $I(0)$, and the radius of gyration, R_g were evaluated using the Guinier approximation for the scattering intensity, $I(q)$, at very low q angles ($qR_g < 1.3$) (Table S1) (9):

$$I(q) = I(0)\exp\left(\frac{-q^2 {R_g}^2}{3}\right). \quad \text{[S2]}$$

Guinier plots of $\ln[I(q)]$ versus q^2 were inspected for linearity, and samples at different concentrations were checked for consistent values of $I(0)/c$, where c is the sample concentration, and R_g to confirm sample monodispersity. $I(0)$ is related to the number of scattering particles per unit volume; particle molecular weights of samples with homogeneous components (MM_p) were estimated using the equation

$$MM_p = (I(0)_p/c_p) \times \frac{MM_{st}}{I(0)_{st}/c_{st}}, \quad \text{[S3]}$$

where $I(0)_p$ and $I(0)_{st}$ are the scattering intensities at zero-angle, MM_p and MM_{st} are the molecular masses, and c_p and c_{st} are the concentrations measured (in g/L), for the sample under study and the protein or RNA standard, respectively (Table S1) (10). The good agreement of these values with those calculated from the primary sequence indicates that the samples are monomeric, free from aggregates, and lack long-range particle interactions under the experimental conditions. In most cases, the scattering data measured for the highest-concentration sample were of high-quality and were used in subsequent structural analysis. Where a slight upturn at very low q in plots of $\log[I(q)]$ versus $\log(q)$ was observed for high-concentration samples, the high q values for these data were merged with low q values measured at a lower concentration to obtain a high-quality dataset for the whole range of q. Final scattering profiles were analyzed with the indirect transform program AUTOGNOM to obtain the maximum particle dimension, D_{max}, the distance distribution function, $P(r)$, and a value for the R_g calculated from the entire scattering profile (11). The latter were in good agreement with those calculated using the Guinier approximation, which demonstrates the internal consistency of the data (Table S1) (8).

Ab Initio Shape Reconstructions. Three-dimensional reconstructions of the SAXS data using ab initio modeling were performed for the constructs of Mss116p and CYT-19 in the free and U_{10}-RNA-bound states. Nucleic-acid components give rise to a significantly greater scattering intensity compared to an equivalent amount of protein because of their higher electron density (12, 13). In this analysis, U_{10}–RNA-protein-ADP-BeF_x complexes (which consist of <5% nucleic-acid) were assumed to be single-component systems of uniform electron densities. Low-resolution particle envelopes were determined by using the simulated annealing procedures implemented by the programs DAMMIN (for $q_{max} < 8/R_g$) and GASBOR (12, 14, 15). In DAMMIN, a single-phase dummy atom protein model is generated from an initial spherical search volume of densely packed beads that can be either protein or solvent. GASBOR uses an assembly of dummy residues that form a polypeptide chain-compatible model to encourage shape reconstructions with protein-like properties. Multiple independent runs were performed for each analysis, and these resulted in reproducible results with good fits to the experimental data (Table S3). Reconstructions were averaged using the DAMAVER suite to confirm the uniqueness and reliability of the solution (16). This method generates the normalized spatial discrepancy (NSD), a quantitative measure of the similarity between sets of three-dimensional points, which was used to compare models of the same resolution. A low value for NSD (approximately 1) indicates a good agreement between models (Table S3) (17). The final DAMMIN model represents a refinement of the average of 10 individual reconstructions against the experimental data (Figs. 2, 3, and 5), whereas the most typical GASBOR model with the lowest average NSD compared to the others is shown in Fig. S1. Low-resolution envelopes were also reconstructed using DAMMIN from SAXS data measured for RNA–DNA duplex 1 and RNA–DNA duplex 2, as described above (Fig. S6).

Homology Modeling of CYT-19. A structural model of CYT-19 was generated from the crystal-structure of Mss116p using the program MODELLER (18) and the published sequence alignment (5, 19). The sequence alignment with Mss116p suggests that the CTE of CYT-19 extends to residue 590, leaving a C-tail of 36 amino acid residues. However, a somewhat larger segment of 49 residues is sensitive to proteolysis (4), indicating either that the final helix of the CTE is also protease sensitive or that the C-tail is somewhat longer than predicted from the sequence alignment. The CYT-19/ΔC-tail mutant was constructed based upon the proteolysis data and therefore has a 49-amino-acid residue C-terminal truncation, which includes an approximately 13 residue segment that could belong either to the CTE or C-tail. The disposition of this approximately 13 residue segment—helical or unstructured—does not significantly affect the modeling comparisons of the SAXS data or the conclusions.

Molecular Modeling. The high-resolution crystal-structure of the helicase core of Mss116p (Fig. 1*B*) and the homology model of CYT-19 were used as additional restraints to generate molecular models of the open and closed states. Multidomain atomic models of Mss116p and CYT-19 were reconstructed from the SAXS data by the program BUNCH, which employs a combination of rigid-body and ab initio modeling (20). The protein domains with known structure are used to generate a theoretical scattering pattern, whereas the unknown regions in each construct are represented by an appropriate number of dummy residues. For open-state models, an additional flexible linker was introduced between the helicase domains 1 and 2 (Fig. 1*A*). BUNCH calculates the relative orientation of the components that is most consistent with the SAXS data. Results from multiple runs gave reproducible results with low average NSD values (Table S3). The low discrepancies between experimental data and computed scattering curves of the final BUNCH models also suggest that a single, unique model is sufficient to describe the SAXS data (Table S3). The molecular models produced using BUNCH were overlaid with the shape reconstructions from DAMMIN using SUPCOMB (17).

Multiphase Modeling. For accurate SAXS reconstructions of DEAD-box proteins bound to large nucleic-acid substrates, it is necessary to take into account the difference in scattering intensities of protein and nucleic-acid. Therefore, the program MONSA, a multiphase version of DAMMIN, was used to obtain ab initio models of protein complexes that include RNA–DNA duplex 1 or RNA–DNA duplex 2 (12, 13). MONSA uses simulated annealing to reconstruct a model of two phases, protein and nucleic-acid, from an assembly of beads inside a defined search volume. Within this volume, each bead is assigned to solvent, protein or nucleic-acid. The program was applied to simultaneously fit three scattering curves of protein, substrate, and complex of protein and substrate, to find the best distribution of beads that corresponds to interconnected phases representing the protein and nucleic-acid. In this analysis, SAXS data from the equivalent protein-U_{10}–RNA complex were used to represent the closed-state protein structure. Several independent runs gave reproducible models with an average NSD of less than 0.65 that described the scattering data well (Table S3). The most typical model is shown with the lowest NSD compared to all other models as calculated using the DAMAVER suite.

Rigid-Body Modeling with Conformational Sampling. SAXS data can be used to investigate the relative compactness of a macromolecule by analysis methods that involve rigid-body modeling with

conformational sampling (21). The ensemble optimization method (EOM) was used to characterize the degree to which Mss116p is extended in solution (21). This method uses the available crystal-structure coordinates and a polypeptide chain of dummy residues to represent the structured and unstructured regions of the protein, respectively. A large pool of conformations is generated in which the unstructured regions are modeled by consecutively adding amino acids with phi and psi angles that are randomly drawn from a library of unstructured loops in the protein data bank. From this pool of 10,000 conformations, an optimized ensemble of 100 conformations is selected that best describes the SAXS data. We used the EOM to model the scattering profiles of the constructs of Mss116p in open and U_{10}-RNA-bound states. The R_g and D_{max} distributions for this optimized ensemble were then compared to those of the random pool (Fig. S3).

1. Halls C, et al. (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol* 365:835–855.
2. Del Campo M, Lambowitz AM (2009) Crystallization and preliminary X-ray diffraction of the DEAD-box protein Mss116p complexed with an RNA oligonucleotide and AMP-PNP. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 65:832–835.
3. Wilkins MR, et al. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552.
4. Grohman JK, et al. (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: The C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry* 46:3013–3022.
5. Del Campo M, Lambowitz AM (2009) Structure of the yeast DEAD-box protein Mss116p reveals two wedges that crimp RNA. *Mol Cell* 35:598–609.
6. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:252–260.
7. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* 89:392–400.
8. Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci* 19:642–657.
9. Guinier A (1939) Diffraction of X-rays of very small angles–application to the study of ultramicroscopic phenomena. *Ann Phys* 12:161–237.
10. Mylonas E, Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J Appl Crystallogr* 40:S245–S249.
11. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI (2007) ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* 40:S223–S228.
12. Svergun DI (1999) Restoring low-resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76:2879–2886.
13. Svergun DI, Nierhaus KH (2000) A map of protein-rRNA distribution in the 70 S *Escherichia coli* ribosome. *J Biol Chem* 275:14432–14439.
14. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80:2946–2953.
15. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346.
16. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860–864.
17. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34:33–41.
18. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
19. Mohr G, et al. (2008) Function of the C-terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. *J Mol Biol* 375:1344–1364.
20. Petoukhov MV, Svergun DI (2005) Global rigid-body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89:1237–1250.
21. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664.

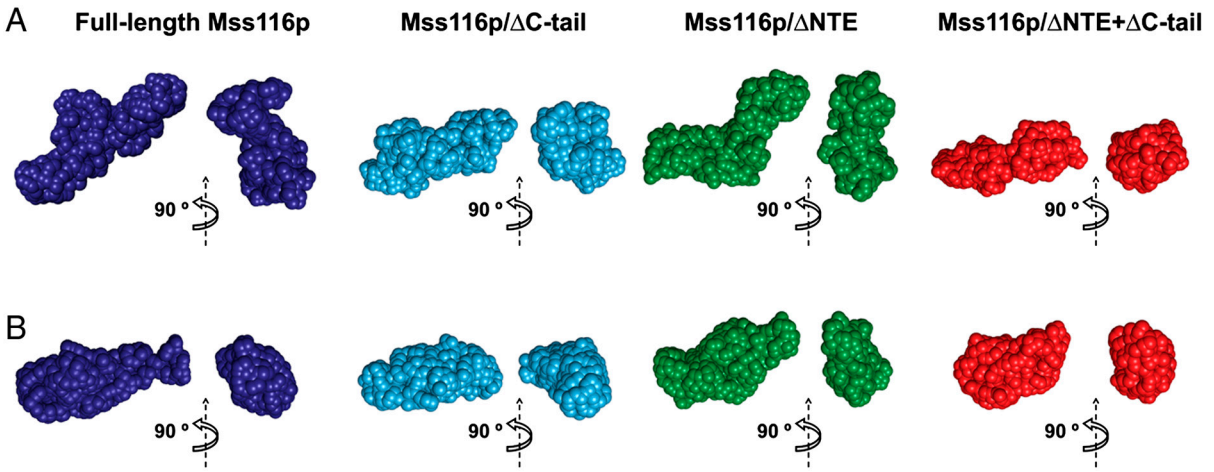

Fig. S1. Reconstructions of SAXS data for Mss116p using modeling with chain-like dummy residues. Low-resolution envelopes calculated from the SAXS data by GASBOR are shown for full-length Mss116p (dark blue), Mss116p/ΔC-tail (light blue), Mss116p/ΔNTE (green), and Mss116p/ΔNTE + ΔC-tail (red) for (*A*) proteins in the absence of ligands and (*B*) proteins bound to ADP-BeF_x and U_{10}-RNA. Views rotated by 90° about the vertical axis are shown for each model.

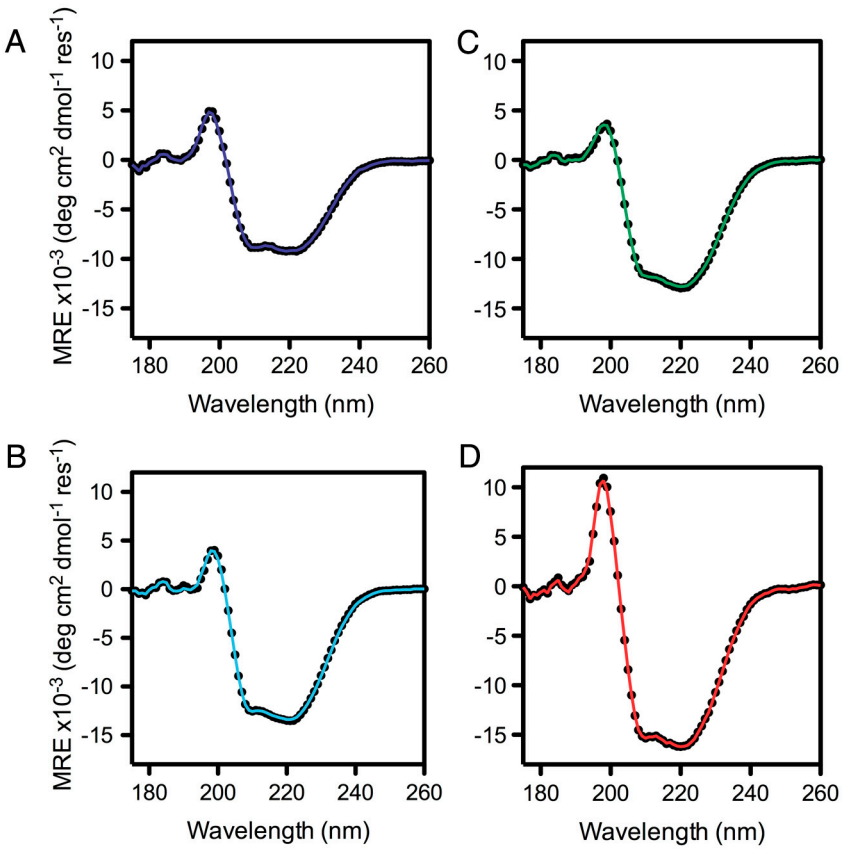

Fig. S2. Far ultraviolet circular dichroism spectra of full-length Mss116p and deletion proteins. Spectra are shown in units of mean residue ellipticity (MRE). Secondary-structure content was calculated from the experimental far-UV CD data by the DichroWeb online server (7) using the CDSSTR analysis program with the SP175 reference dataset (6). The results of this analysis are shown in Table S4. The fit of the theoretical far-UV CD spectra from the analysis to the experimental data (black circles) is shown for (*A*) full-length Mss116p (blue), (*B*) Mss116p/ΔC-tail (light blue), (*C*) Mss116p/ΔNTE (green), and (*D*) Mss116p/ΔNTE + ΔC-tail (red). Normalized-route-mean-square-deviation (NRMSD) values were less than 0.025 for all fits.

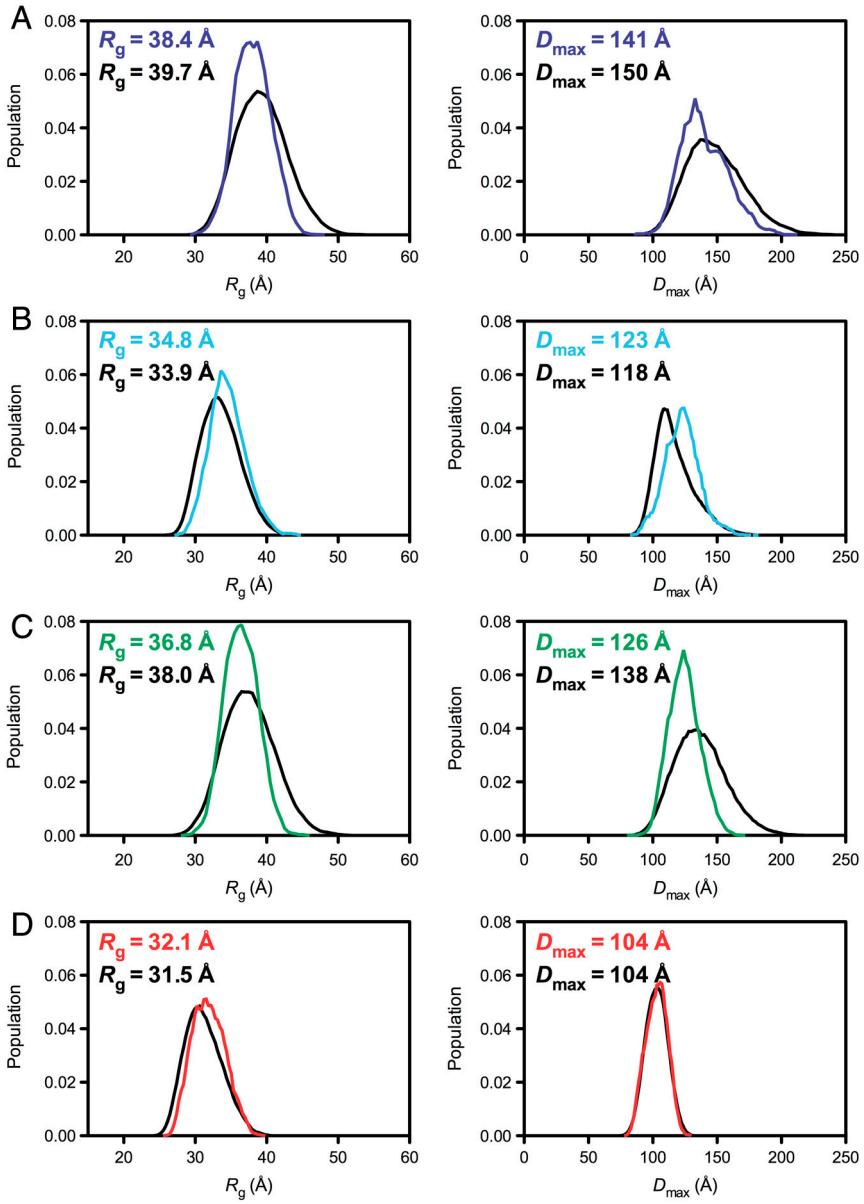

Fig. S3. Size and shape distributions of optimized ensembles for full-length Mss116p and deletion proteins. An optimized ensemble of the 100 models that best describe each scattering profile was selected from a pool of 10,000 random conformers using the EOM. The distributions for the R_g (*Left*) and D_{max} (*Right*) of the models in the optimized ensembles for (*A*) full-length Mss116p (blue), (*B*) Mss116p/ΔC-tail (light blue), (*C*) Mss116p/ΔNTE (green), and (*D*) Mss116p/ΔNTE + ΔC-tail (red) are shown compared to those of the corresponding random pool (black). The average values for the R_g and D_{max} of the optimized ensembles and the random pool are shown in equivalent colors.

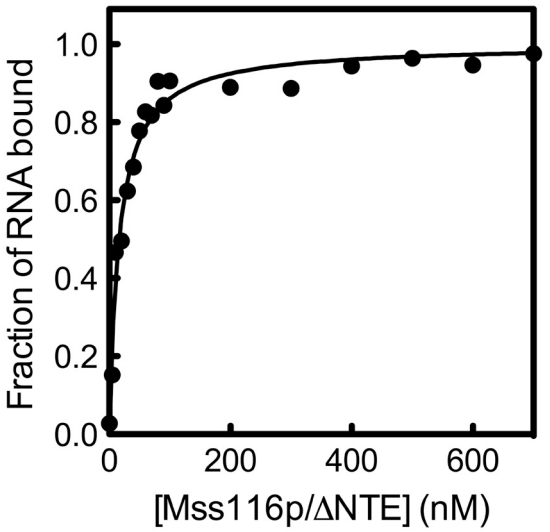

Fig. S4. Equilibrium binding of Mss116p to U_{10}-RNA. Fluorescein-labeled FAM-U_{10}-RNA (20 nM) was incubated with increasing concentrations of Mss116p/ΔNTE in the presence of ADP-BeF_x (0.5 mM) and bovine serum albumin (0.1 mg/mL) in the SAXS buffer of 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$ for 30 min at 22 °C. The fraction of bound RNA was measured by monitoring the change in fluorescence at 510 nm after excitation at 480 nm. Incubations of longer than 30 min gave no additional change in fluorescence, indicating that this time was sufficient to reach equilibrium. The solid line represents the fit of the data to a one-site binding model to give an upper limit for the value of the dissociation constant, K_d, of approximately 20 nM.

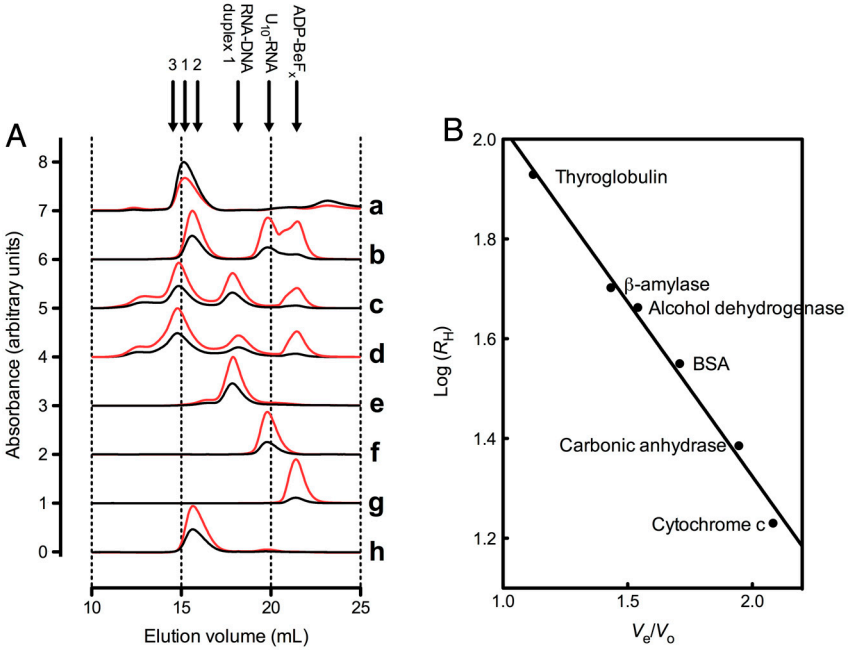

Fig. S5. Purification and characterization of Mss116p nucleic acid complexes using size-exclusion chromatography. SEC was performed using a Superdex 200 10/300 GL column (GE Healthcare) and a BioLogic DuoFlow chromatography system (BioRad) in the SAXS buffer of 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% glycerol, 1 mM DTT, 5 mM $MgCl_2$. (*A*) SEC data are shown for Mss116p/ΔNTE + ΔC-tail and were measured using absorbance at 280 nm (black) and 260 nm (red). Example elution profiles for (a) free protein in the absence of ligands (elution volume, $V_e = 15.1$ mL; peak position 1), (b) U_{10}–RNA protein-ADP-BeF_x complex ($V_e = 15.6$ mL; peak position 2), (c) RNA–DNA-duplex 1-protein-ADP-BeF_x complex ($V_e = 14.8$ mL; peak position 3), (d) RNA–DNA–duplex 2-protein-ADP-BeF_x complex ($V_e = 14.8$ mL; peak position 3), (e) RNA–DNA duplex 1 ($V_e = 17.9$ mL), (f) U_{10}–RNA ($V_e = 19.8$ mL) and (g) ADP-BeF_x ($V_e = 21.4$ mL). Fractions corresponding to the center of the elution peak of the complexes were collected for analysis by SAXS. The stability of the complexes was confirmed by a reanalysis of the isolated complex by SEC to check for dissociation, an example of which is shown in (h) for U_{10}-RNA-Mss116p/ΔNTE + ΔC-tail-ADP-BeF_x ($V_e = 15.6$ mL; peak position 2). The absence of peaks corresponding to either U_{10}–RNA or ADP-BeF_x indicated that the complex remains intact. The ratio of A_{260}/A_{280}, which is approximately 0.5 for free protein and 2 for protein-nucleic-acid complexes, was also used as an indicator of complex formation. (*B*) A calibration curve for the SEC column used in these experiments showing the logarithm of the hydrodynamic radius, R_H, as a function of the relative elution volume, V_e/V_o, where V_o is the void volume of the column. The hydrodynamic radii for the free proteins and complexes of full-length Mss116p and deletion proteins, which are shown in Table S2, were determined by a comparison of their relative elution volumes to those of standard proteins with known hydrodynamic radii (Eq. **S1**).

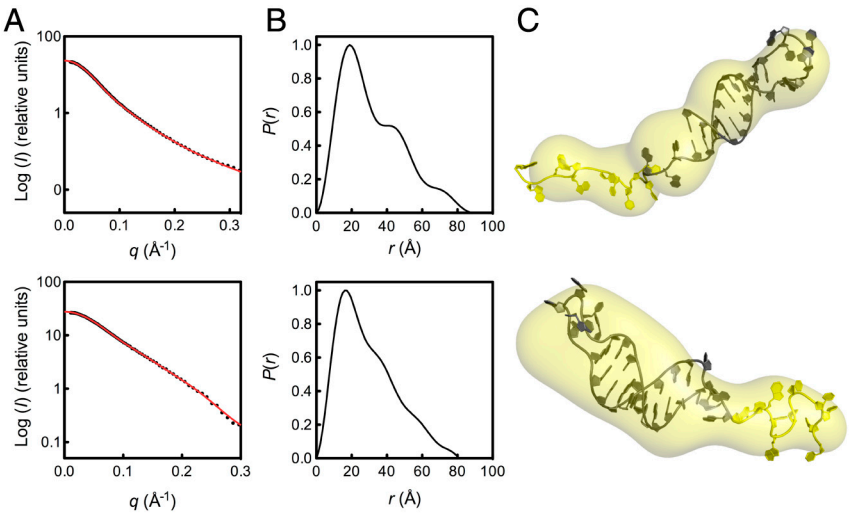

Fig. S6. SAXS data for chimeric nucleic-acid substrates. (*A*) Scattering profiles, (*B*) normalized distance distribution functions, and (*C*) low-resolution DAMMIN reconstructions for RNA–DNA duplex 1 (*Upper*) and RNA–DNA duplex 2 (*Lower*). The red solid line in (*A*) represents the fit of the simulated scattering curve for the DAMMIN envelopes shown in (*C*) to the experimental data ($\chi = 1.02$ and 0.43 for RNA–DNA duplex 1 and RNA–DNA duplex 2, respectively). Atomic models for each substrate, colored according to the sequences in Fig. 1*C*, were generated using the nucleic-acid structure prediction program RNABuilder (1, 2), and were manually placed inside the low-resolution SAXS envelopes.

1 Flores SC, Altman RB (2010) Turning limited experimental information into 3D models of RNA. *RNA* 16:1769–1778.
2 Flores SC, Wan Y, Russell R, Altman RB (2010) Predicting RNA structure by multiple template homology modeling. *Pac Symp Biocomput* 2010:216–227.

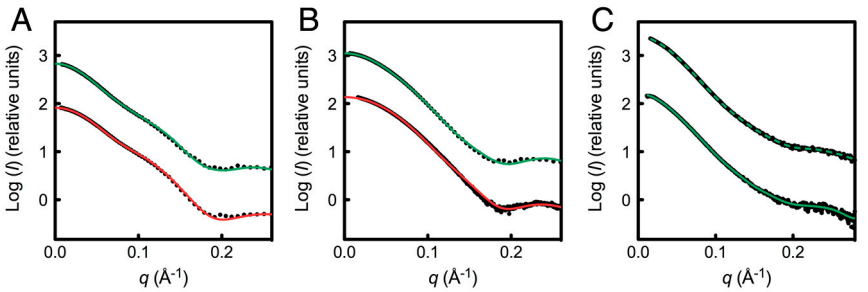

Fig. S7. Scattering data for CYT-19. Scattering profiles were measured for proteins (*A*) in the absence of ligands, (*B*) bound to ADP-BeF_x and U_{10}-RNA, and (*C*) bound to ADP-BeF_x and larger nucleic-acid substrates. SAXS data are shown as the logarithm of the scattering intensity, *I* (black dots), as a function of the momentum transfer, $q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. In (*A*) and (*B*), the solid curves represent the expected scattering profiles of the corresponding BUNCH atomic models or, in the case of U_{10}-RNA-CYT-19/ΔC-tail-ADP-BeF_x, the homology model of the closed-state CYT-19. In (*C*), the fits of the multiphase ab initio model obtained by MONSA for complexes that contain RNA–DNA duplex 1 (solid line) and RNA–DNA duplex 2 (dashed line) are displayed. In panels *A*–*C*, fits shown in green are for full-length CYT-19 and those shown in red are for CYT-19/ΔC-tail.

Table S1. Structural parameters for DEAD-box proteins calculated from SAXS data

Sample	Molecular weight determined from sequence/kDa	Abs $I(0)/c$/ $cm^2\ mg^{-1}$*	Molecular weight determined from SAXS data/kDa†	R_g (Guinier analysis)/Å	R_g (GNOM analysis)/Å	D_{max}/Å
Free protein and nucleic acid						
Full-length Mss116p	72.5	69.1 ± 2.4	73.4 ± 3.0	38.0 ± 0.2	38.9 ± 0.1	135 ± 2
Mss116p/ΔC-tail	64.4	55.9 ± 3.0	59.4 ± 3.5	34.2 ± 0.3	34.7 ± 0.2	120 ± 2
Mss116p/ΔNTE	66.1	56.2 ± 1.5	59.7 ± 2.1	36.9 ± 0.2	38.1 ± 0.3	132 ± 2
Mss116p/ΔNTE+ΔC-tail	58.1	50.3 ± 2.9	53.4 ± 3.4	32.8 ± 0.1	33.7 ± 0.1	115 ± 2
Full-length CYT-19	63.9	—	68‡	36.6 ± 0.1	36.9 ± 0.1	123 ± 2
CYT-19/ΔC-tail	58.2	—	59‡	32.6 ± 0.1	33.2 ± 0.2	115 ± 2
RNA–DNA duplex 1	12.4	62.7 ± 0.3	13.4 ± 1.0	24.6 ± 0.2	24.7 ± 0.1	85 ± 2
RNA–DNA duplex 2	12.4	60.8 ± 0.5	13.0 ± 1.0	23.1 ± 0.3	23.6 ± 0.2	80 ± 2
Cytochrome *c* (horse heart)	12.3	11.6 ± 0.3	—	15.2 ± 0.2	—	—
U_{10}-RNA	3.0	14.0 ± 1.1	—	12.6 ± 0.8	—	—
Protein–nucleic acid complexes						
U_{10}-RNA-full-length Mss116p-ADP-BeF_x	75.9	77.6 ± 1.5	—	33.8 ± 0.1	34.0 ± 0.2	117 ± 2
U_{10}-RNA-Mss116p/ΔC-tail-ADP-BeF_x	67.9	66.7 ± 2.4	—	29.9 ± 0.1	29.8 ± 0.3	102 ± 2
U_{10}-RNA-Mss116p/ΔNTE-ADP-BeF_x	69.6	67.2 ± 3.3	—	30.3 ± 0.1	31.8 ± 0.1	112 ± 2
U_{10}-RNA-Mss116p/ΔNTE + ΔC-tail-ADP-BeF_x	61.5	60.5 ± 1.8	—	26.3 ± 0.2	25.9 ± 0.1	82 ± 2
U_{10}-RNA-full-length CYT-19-ADP-BeF_x	67.4	—	—	30.4 ± 0.1	31.0 ± 0.1	109 ± 2
U_{10}-RNA-CYT-19/ΔC-tail-ADP-BeF_x	61.7	—	—	27.4 ± 0.1	27.5 ± 0.1	95 ± 2
RNA–DNA duplex 1-Mss116p/ ΔNTE-ADP-BeF_x	79.0	92.9 ± 4.9	—	36.0 ± 0.4	36.1 ± 0.1	120 ± 2
RNA–DNA duplex 1-Mss116p/ ΔNTE+ΔC-tail-ADP-BeF_x	70.9	91.5 ± 5.3	—	33.2 ± 0.2	34.0 ± 0.1	118 ± 2
RNA–DNA duplex 2-Mss116p/ ΔNTE-ADP-BeF_x	79.0	83.7 ± 1.7	—	36.1 ± 0.2	37.6 ± 0.1	135 ± 2
RNA–DNA duplex 2-Mss116p/ ΔNTE+ΔC-tail-ADP-BeF_x	70.9	81.3 ± 3.4	—	33.1 ± 0.1	34.3 ± 0.1	119 ± 2
RNA–DNA duplex 1-full-length CYT-19-ADP-BeF_x	76.8	—	—	36.2 ± 0.1	36.6 ± 0.1	126 ± 2
RNA–DNA duplex 2-full-length CYT-19-ADP-BeF_x	76.8	—	—	35.2 ± 0.1	35.2 ± 0.1	120 ± 2

*The forward scattering intensity, $I(0)$, normalized against protein concentration, *c*, measured in mg/mL. Error estimates are the standard errors calculated from linear regression.

†Particle molecular weights of samples with homogeneous components were estimated using Eq. **S3** and standards of cytochrome c and U_{10}–RNA for protein and nucleic-acid samples, respectively. The errors were estimated by propagation of the error calculated for $I(0)/c$.

‡Molecular weights for CYT-19 samples were determined from SAXS using the calculated Porod volume (1).The error estimates for R_g values calculated by Guinier and GNOM analyses are the standard error for linear regression and the standard deviation evaluated by GNOM, respectively. Errors for the maximum particle dimension, D_{max}, are the estimated range for the optimum solution.

1 Porod G (1982) General theory. *Small Angle X-ray Scattering*, eds Glatter O, Kratky O (Academic Press, London), pp 17–51.

Table S2. Hydrodynamic radii and R_g values for Mss116p SAXS samples

Sample	R_H/Å*	R_g/Å†
Free protein		
Full-length Mss116p	35.0	38.0 ± 0.2
Mss116p/ΔC-tail	32.0	34.2 ± 0.3
Mss116p/ΔNTE	34.3	36.9 ± 0.2
Mss116p/ΔNTE+ΔC-tail	31.1	32.8 ± 0.1
Protein–nucleic acid complexes		
U_{10}–RNA-full-length Mss116p-ADP-BeF_x	33.2	33.8 ± 0.1
U_{10}–RNA-Mss116p/ΔC-tail-ADP-BeF_x	30.7	29.9 ± 0.1
U_{10}–RNA-Mss116p/ΔNTE-ADP-BeF_x	31.8	30.3 ± 0.1
U_{10}–RNA-Mss116p/ΔNTE+ΔC-tail-A ADP-BeF_x	28.5	26.3 ± 0.2
RNA–DNA duplex 1-Mss116p/ΔNTE-ADP-BeF_x	35.4	36.0 ± 0.4
RNA–DNA duplex 1-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	32.8	33.2 ± 0.2
RNA–DNA duplex 2-Mss116p/ΔNTE-ADP-BeF_x	36.0	36.1 ± 0.2
RNA–DNA duplex 2-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	33.2	33.1 ± 0.1

*The hydrodynamic radius, R_H, was calculated using SEC (Eq. **S1** and Fig. S5). The error in the value for R_H is estimated to be approximately ±5%, which is based upon the standard error in the calibration curve shown in Fig. S5*B*.

†Radii of gyration are shown here for comparison and are taken from Table S1.

Table S3. Parameters for the ab initio and rigid-body models constructed from SAXS data

	Modeling program					
	DAMMIN or MONSA		GASBOR		BUNCH	
	Single or multiphase ab initio bead modeling, respectively		Ab initio modeling with chain-like dummy residues		Rigid-body modeling	
Sample	χ	NSD	χ	NSD	χ	NSD
Full-length Mss116p	0.98	0.82 ± 0.04	1.09	1.44 ± 0.06	0.83	1.53 ± 0.06
Mss116p/ΔC-tail	1.67	0.92 ± 0.05	0.72	1.26 ± 0.02	0.67	0.96 ± 0.16
Mss116p/ΔNTE	0.45	0.81 ± 0.07	0.52	1.09 ± 0.03	1.03	0.96 ± 0.09
Mss116p/ΔNTE+ΔC-tail	0.94	0.67 ± 0.05	1.00	1.15 ± 0.02	2.13	0.62 ± 0.03
Full-length CYT-19	1.06	0.72 ± 0.04	—	—	1.12	0.69 ± 0.18
CYT-19/ΔC-tail	0.47	0.71 ± 0.19	—	—	1.42	0.77 ± 0.13
RNA–DNA duplex 1	1.01	0.71 ± 0.04	—	—	—	—
RNA–DNA duplex 2	0.43	0.65 ± 0.01	—	—	—	—
U_{10} RNA-full-length Mss116p-ADP-BeF_x	0.54	0.82 ± 0.15	0.59	1.25 ± 0.05	1.19	1.04 ± 0.20
U_{10} RNA-Mss116p/ΔC-tail-ADP-BeF_x	0.38	0.72 ± 0.13	0.52	1.09 ± 0.03	1.43	0.47 ± 0.05
U_{10} RNA-Mss116p/ΔNTE-ADP-BeF_x	1.32	0.70 ± 0.03	2.23	1.20 ± 0.06	1.86	0.46 ± 0.18
U_{10} RNA-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x	0.55	0.92 ± 0.03	0.66	0.98 ± 0.01	1.66	—
U_{10}–RNA-full-length CYT-19-ADP-BeF_x	1.15	0.76 ± 0.12	—	—	0.96	0.27–0.01
U_{10}–RNA-CYT-19/ΔC-tail-ADP-BeF_x	2.03	0.69 ± 0.10	—	—	—	—
RNA–DNA duplex 1-Mss116p/ΔNTE-ADP-BeF_x*	0.67	0.64 ± 0.02	—	—	—	—
RNA–DNA duplex 1-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x*	1.24	0.46 ± 0.02	—	—	—	—
RNA–DNA duplex 2-Mss116p/ΔNTE-ADP-BeF_x*	1.48	0.63 ± 0.09	—	—	—	—
RNA–DNA duplex 2-Mss116p/ΔNTE+ΔC-tail-ADP-BeF_x*	1.44	0.47 ± 0.03	—	—	—	—
RNA–DNA duplex 1-full-length CYT-19-ADP-BeF_x*	2.24	0.63 ± 0.04	—	—	—	—
RNA–DNA duplex 2-full-length CYT-19-ADP-BeF_x*	2.12	0.72 ± 0.05	—	—	—	—

For rigid-body modeling, the final BUNCH models shown in Figs. 2, 3, and 5 represent the most typical structures from multiple runs that have the lowest average NSD value when compared to all the others. The value for χ represents the difference between the calculated scattering profile of the final model and the experimental data. In the case of the U_{10}-RNA-Mss116p/ΔNTE + ΔC-tail-ADP-BeF_x complex where the crystal structure is already known, the χ-value is the difference between the predicted scattering profile of the crystal-structure calculated using CRYSOL (1) and the experimental scattering data of the complex. This low χ value is a good indication that the complex formed in solution resembles that seen in the crystal-structure of Mss116p. In addition, the R_g value calculated by CRYSOL for the crystal-structure is 25.8 Å, which is in good agreement with the value of 26.3 Å calculated from the SAXS data for the equivalent complex by Guinier analysis (Table S1).

1 Svergun D, Barberato C, Koch MHJ (1995) CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.

Table S4. Secondary-structure analysis of far ultraviolet circular dichroism spectra for Mss116p

Construct	Residues	% Helix	% Sheet	% Turns and unordered
Full-length Mss116p	37–664	33 (36)	14 (14)	53 (50)
Mss116p/ΔC-tail	37–597	43 (41)	13 (15)	44 (44)
Mss116p/ΔNTE	88–664	41 (40)	14 (15)	45 (45)
Mss116p/ΔNTE+ΔC-tail	88–597	51 (45)	12 (17)	37 (38)
Crystal structure	88–596	45	17	38

The far-UV CD spectra shown in Fig. S2 were analyzed using the CDSSTR algorithm with the SP175 dataset implemented by the DichroWeb online server for deconvolution of circular dichroism spectra (1, 2). NRMSD values between the calculated and experimental CD data were less than 0.025 in all cases. Protein concentrations ranged from 1.1–2.5 mg/mL and far-UV CD data from 175–260 nm were used in the analysis. The secondary-structure content of the crystal-structure of Mss116p [Protein Data Bank (PDB) code 3I5X] was computed using the PROMOTIF program (3). The numbers in brackets indicate the expected secondary-structure content assuming that residues 37–87 (NTE) and 597–664 (C-terminal basic tail) are unstructured, and residues 88–596 adopt the same secondary-structure as seen in the crystal-structure. The good agreement between the % of turns and unordered regions calculated by analysis of the far-UV CD spectra and that expected if the NTE and C-tail extensions are assumed to be disordered is consistent with these extensions existing as unstructured polypeptide chains. Conditions: 20 mM Tris-HCl (pH 7.5), 500 mM KCl, 10% (v/v) glycerol, 1 mM DTT, 5 mM $MgCl_2$ at 25 °C.

1 Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:252–260.
2 Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* 89:392–400.
3 Hutchinson EG, Thornton JM (1996) PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci* 5:212–220.