Folding of the RNA Recognition Motif (RRM) Domains of the Amyotrophic Lateral Sclerosis (ALS)-linked Protein TDP-43 Reveals an Intermediate State

Received for publication, December 13, 2013, and in revised form, January 29, 2014. Published, JBC Papers in Press, February 4, 2014, DOI 10.1074/jbc.M113.542779

Brian C. Mackness, Meme T. Tran, Shannan P. McClain, C. Robert Matthews, and Jill A. Zitzewitz

From the Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605

Background: TDP-43 aggregates and mutations are observed in patients with ALS and FTLD.
Results: The equilibrium unfolding of the RRM domains reveals a highly populated intermediate in RRM2.
Conclusion: The stability of RRM2 may result from a large hydrophobic cluster, and the intermediate state may be essential for accessing the nuclear export sequence.
Significance: Accessing the RRM2 intermediate state can potentially propagate disease pathogenesis.

Pathological alteration of TDP-43 (TAR DNA-binding protein-43), a protein involved in various RNA-mediated processes, is a hallmark feature of the neurodegenerative diseases amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Fragments of TDP-43, composed of the second RNA recognition motif (RRM2) and the disordered C terminus, have been observed in cytoplasmic inclusions in sporadic amyotrophic lateral sclerosis cases, suggesting that conformational changes involving RRM2 together with the disordered C terminus play a role in aggregation and toxicity. The biophysical data collected by CD and fluorescence spectroscopies reveal a three-state equilibrium unfolding model for RRM2, with a partially folded intermediate state that is not observed in RRM1. Strikingly, a portion of RRM2 beginning at position 208, which mimics a cleavage site observed in patient tissues, increases the population of this intermediate state. Mutually stabilizing interactions between the domains in the tethered RRM1 and RRM2 construct reduce the population of the intermediate state and enhance DNA/RNA binding. Despite the high sequence homology of the two domains, a network of large hydrophobic residues in RRM2 provides a possible explanation for the increased stability of RRM2 compared with RRM1. The cluster analysis suggests that the intermediate state may play a functional role by enhancing access to the nuclear export signal contained within its sequence. The intermediate state may also serve as a molecular hazard linking productive folding and function with pathologi- misfolded aggregation that may contribute to disease.

Amyotrophic lateral sclerosis (ALS) is a highly debilitating and progressive motor neuron disease affecting approximately 1–2 of 100,000 new people each year, with death occurring 2–5 years after onset (1). Only 10% of ALS cases have been linked to genetic mutations in numerous genes (familial ALS), whereas the remaining 90% of cases result from an unknown cause (sporadic ALS) (2). The pathological hallmark of ALS is the presence of ubiquitinated inclusions in the cytoplasm of surviving spinal motor neurons. In 2006, biochemical and immunological approaches identified TAR DNA-binding protein 43 (TDP-43) as a major protein found in post-mortem brain inclusions of patients with both ALS and frontotemporal lobar degeneration with ubiquitinated inclusions (FTLD-U), providing a molecular connection between these diseases (3, 4). In the years since this initial discovery, 50 different TDP-43 mutations in familial and sporadic ALS patients have been identified (see the ALS Online Genetics Database), thereby underscoring a direct role for TDP-43 in ALS pathogenesis. Related research has been rapid (5), with numerous reports of mouse models, biomarkers, and assays for testing disease progression and cellular function. However, a molecular level understanding of how TDP-43 may lead to disease is still lacking (4), in part because of the poor solubility of the full-length protein and its tendency to fragment and aggregate.

TDP-43 is a 414-amino acid protein that contains two RNA recognition motifs (RRMs), a nuclear localization sequence in the N terminus, a nuclear export signal (NES) within RRM2, and a C-terminal glycine-rich domain (accession no. Q13148, UniProtKB/Swiss-Prot) (Fig. 1A). TDP-43 is ubiquitously expressed and has been implicated to play a functional role in many RNA processes, including gene transcription, splicing, mRNA processing, and mRNA stability (6–8). TDP-43 is localized in the nucleus in normal cells but redistributes to form cytoplasmic aggregates composed of hyperphosphorylated and ubiquitinated C-terminal fragments in diseased cells (9). As inferred from studies of other protein aggregates involved in
neurodegenerative diseases (10), TDP-43 aggregates may arise from the population of non-native conformations that probably drive the neurodegeneration directly through a gain-of-function or loss-of-function mechanism. In these scenarios, the formation and accumulation of TDP-43 aggregates generate toxicity or impair normal TDP-43 cellular function, resulting in cell death. Mounting evidence supports a loss-of-function phenotype (11), where sequestration of functional protein into cytoplasmic aggregates would limit the pool of available functional nuclear TDP-43.

Several studies have demonstrated that both RRM2 and the disordered C terminus are required for aggregation and toxicity (12–15). Examination of the domain architecture of TDP-43 suggests a possible interplay between structured and disordered sequences that may play a key role in toggling between appropriate biological function and dysfunction leading to disease (Fig. 1A). Computer algorithms developed for predicting aggregation prone regions in unfolded polypeptide chains (16), WALTZ and TANGO, both show a high propensity for aggregation in RRM2 and the C terminus (Fig. 1A). The sequence-based disorder predictor algorithm, PONDR (17, 18), predicts a high degree of disorder in the C-terminal region of TDP-43. The two RRM domains share significant sequence (Fig. 1B) and structural (Fig. 1, C and D) homology. The NMR solution structures of the isolated RRM domains (Fig. 1, C and D) and the tethered RRM domains (Fig. 1E) show an α + β structure for each domain, composed of two repeating βαβ motifs with an extra β-strand (β4) inserted within the second βαβ motif. Together, these strands form an antiparallel β-sheet across one face of the RRM with the α-helices docked on the opposite face. Although RRM domains can bind a diverse set of targets, including RNA, DNA, and peptides and other proteins (19, 20), most studies on TDP-43 have focused on the role of the highly conserved phe- nylalanine side chains in β3 for RNA recognition (21–23).

To understand how the conformations populated by the RRM domains of TDP-43 may play a role in disease propagation, the equilibrium unfolding and RNA-binding properties of the isolated and tethered RRM domains were probed by a pair of complementary spectroscopic techniques. The results identified a novel intermediate state in the folding of the RRM2 domain. The population of this intermediate is increased in a cleavage fragment but is reduced upon tethering to RRM1. The intermediate state in RRM2 may serve as a molecular hazard that may partition between productive folding and function on the one hand and misfolding and aberrant protein–protein interactions that could lead to disease progression on the other.

**EXPERIMENTAL PROCEDURES**

**Expression and Purification of TDP-43 Fragments**—Gene fragments of TDP-43 encoding RRM1 (amino acids 102–181), RRM2 (amino acids 190–261), an RRM2 fragment mimicking the disease-relevant proteolytic cleavage site within RRM2 (9) (RRM2c: amino acids 208–261), and the tethered RRMs (tRRMs; amino acids 97–261) were purchased from Genescript with BamHI and EcoRI or NcoI and BamHI restriction enzyme digestion sites. The genes were inserted into a modified pGEX-6p1 (GE Healthcare) or pET-3d (Novagen) vector with a His6 tag and either a PreScission or TEV protease cleavage site, respectively, for His6 tag removal. The proteins were overexpressed in BL21 DE3 *Escherichia coli* cells (Stratagene) grown in LB medium until *A*~600~ = 0.8, followed by induction with 1 mM isopropyl 1-thio-β-d-galactopyranoside for 24 h at 20 °C (RRM1 and tRRMs) or for 16 h at 30 °C (RRM2 and RRM2c).

Cells were resuspended in lysis buffer (20 mM NaPi, pH 7.4, 300 mM NaCl, 30 mM imidazole) and lysed by sonication. RRM2 and RRM2c were only present in the insoluble fraction, and isolation from cell pellets was performed in the presence of 6 M urea. Each construct was bound to His60 resin (Clontech) overnight and washed with 10 column volumes of lysis buffer before elution with 300 mM imidazole. The eluted protein was dialyzed against protease cleavage buffer (50 mM Tris, pH 8.0, 1 mM EDTA, and 1 mM DTT), followed by subsequent cleavage with PreScission or TEV protease to remove the His6 tag. Minor contaminants were removed through ion exchange chromatography using Sepharose (RRM1) or Q Sepharose (RRM2, tRRMs, and RRM2c) before dialysis into 10 mM NaPi, pH 7.2, 150 mM KCl, 1 mM β-mercaptoethanol for subsequent studies. After cleavage, protein purity was >98%, as determined by both SDS-PAGE and reverse-phase MALDI-TOF mass spectrometry carried out at the Proteomics and Mass Spectrometry Facility, University of Massachusetts Medical School.

RRM2 contains a second TEV protease-like cleavage site (246EDLIIKG252), as determined by mass spectrometry, preventing isolation of the intact domain. Thus, all RRM2-containing constructs were expressed instead with a PreScission protease site after the N-terminal His tag. Cleavage with TEV protease (24) results in an N-terminal Gly residue before the RRM amino acid sequence, and cleavage with PreScission protease leaves an N-terminal GPLGS sequence, with the LGS sequence being required for cloning.

**Size Exclusion Chromatography to Determine Oligomerization State**—Size exclusion chromatography was performed on all constructs using a 24-ml Superdex 200 10/300 GL column run at a flow rate of 0.2 ml min⁻¹. Oligomerization status of the constructs was monitored as a function of loaded protein concentration by comparison with protein molecular weight standards (GE Healthcare). All size exclusion chromatography was performed at 4 °C in 10 mM KPi, pH 7.2, 150 mM KCl, and 1 mM β-mercaptoethanol.

**Equilibrium Unfolding Experiments**—The native state circular dichroism (CD) spectrum of each construct was collected from 190 to 280 nm on a Jasco-810 spectropolarimeter with a thermostatted temperature control system in a 0.1-cm cuvette (Hellma). Guanidine hydrochloride (GdnHCl)-induced denaturation spectra were collected from 260 to 215 nm at a scan rate of 50 nm min⁻¹ and a response time of 8 s. Samples were prepared from native and unfolded stock solutions mixed in precise amounts by in-house software and a Hamilton series 500 titrator. The resulting solutions were incubated overnight at room temperature or at 37 °C before recording CD spectra. All GdnHCl concentrations were measured using an ABBE refractometer, and all CD measurements were baseline-corrected for buffer contributions. Protein concentration was measured by *A*~280~ absorbance (25), using an extinction coefficient of 15,470 M⁻¹ cm⁻¹ for tRRMs, 13,980 M⁻¹ cm⁻¹ for RRM1, and 1490 M⁻¹ cm⁻¹ for RRM2 and RRM2c. The protein
concentration was varied from 5 to 60 μM for CD experiments. Each CD spectrum was normalized for protein concentration and number of amino acids and reported as mean residue ellipticity (26). Reversibility was confirmed to be $>95\%$ by the coincidence of equilibrium profiles for samples prepared from initial protein stocks in either buffer or denaturant.

Steady-state fluorescence (FL) measurements were performed on a Spex Fluorolog-3 equipped with a wavelength electronics temperature controller. For RRM2 and RRM2c, each GdnHCl titration sample was excited at 274 nm, and tyrosine emission spectra were collected from 280 to 400 nm at 20 or 37 °C with 5-nm slit widths. For the tryptophan-containing RRM1 and tRRMs proteins, excitation was at 295 nm, and tryptophan emission spectra were collected from 300 to 500 nm at 20 °C with 5-nm slit widths.

Denaturation experiments by both CD and FL were performed in replicates of three for each construct to ensure reproducibility. The equilibrium folding data for each construct were analyzed using an appropriate equilibrium folding model with the in-house data analysis software Savuka (27). Each data set was subjected to a global analysis, where the baselines were local parameters, and the free energy of folding in the absence of denaturant ($\Delta G^0_{\text{H2O}}$) and the $m$ value were globally linked between data sets. All of the Trp and Tyr fluorescence equilibrium profiles as well as the CD equilibrium profile for RRM1 were best fit to a two-state model, $N \Leftrightarrow U$. For these experiments, the change in free energy between the native and unfolded states is assumed to depend linearly on the denaturant concentration as shown in Equation 1 (28).

$$\Delta G = \Delta G^0_{\text{H2O}} - m[D] \quad (\text{Eq. 1})$$

The change in free energy in the absence of denaturant, $\Delta G^0_{\text{H2O}}$, can be used to determine the equilibrium constant, $K_{eq}$, from Equation 2.

$$K_{eq} = \exp\left(-\frac{\Delta G^0_{\text{H2O}} - m[D]}{RT}\right) \quad (\text{Eq. 2})$$

For the CD equilibrium unfolding profiles of RRM2 and RRM2c, a three-state model best described the transition between the native and unfolded forms of the protein with the population of a stable intermediate state (I), $N \Leftrightarrow I \Leftrightarrow U$. Tethering the two RRM domains by the natural 15-amino acid residue linker sequence (tRRMs) results in the population of multiple stable intermediates at equilibrium by CD. In this case, the data were best fit with a four-state equilibrium model, $N \Leftrightarrow I1 \Leftrightarrow I2 \Leftrightarrow U$, with the population of two stable intermediate states, I1 and I2.

The fractional population of the intermediate state for RRM2 and RRM2c at a given denaturant concentration was determined from the partition function according to Equation 3 (29),

$$F_i = \frac{\exp\left(-\frac{\Delta G_{Ni}}{RT}\right)}{1 + \exp\left(-\frac{\Delta G_{N1}}{RT}\right) + \exp\left(-\frac{\Delta G_{N2}}{RT}\right) + \exp\left(-\frac{\Delta G_{NU}}{RT}\right)} \quad (\text{Eq. 3})$$

where $\Delta G_{Ni} = \Delta G^0_{\text{NH2O}} - m_{N[I]}[D]$, and $\Delta G_{NU} = \Delta G^0_{\text{NUH2O}} - m_{NU}[D]$. The I2 intermediate in the four-state model for tRRMs corresponds to the single intermediate for RRM2, and its fractional population was calculated as defined in Equation 4,

$$F_i = \frac{\exp\left(-\frac{\Delta G_{N2}}{RT}\right)}{1 + \exp\left(-\frac{\Delta G_{N1}}{RT}\right) + \exp\left(-\frac{\Delta G_{N2}}{RT}\right) + \exp\left(-\frac{\Delta G_{NU}}{RT}\right)} \quad (\text{Eq. 4})$$

where $\Delta G_{N1} = \Delta G^0_{N1} - m_{N1}[D]$, $\Delta G_{N2} = \Delta G^0_{N2} - m_{N2}[D]$, and $\Delta G_{NU} = \Delta G^0_{NU} - m_{NU}[D]$.

**Nucleic Acid Binding Assays—EMSAs and tryptophan (Trp) lifetimes were used to determine the apparent binding affinity of the TDP-43 RRM domains to UUGUGUGUGUG ([U(U)₆]), TGTGTGTGTGTG ([TG]₆), and TTTTTTTTTTTTTT (T₁₂) 12-mer oligonucleotides (IDT Technologies). For EMSA, the oligonucleotides were fluorescently labeled at the 5'-end by 5-carboxyfluorescein (IDT Technologies).**

A typical EMSA consists of 50-μl reactions of 3 nm nucleotides incubated with increasing protein concentrations up to 8 μM. Binding reactions were performed in binding buffer, 10 mM KPO₄, pH 7.2, 150 mM KCl, 2 mM DTT, 10 μg ml⁻¹ tRNA, and 0.01% IGEPAL CA-630 (Sigma-Aldrich) (30) and incubated for 2 h at room temperature. Prior to loading on an 8% polyacrylamide gel, 5 μl of bromocresol blue in 30% glycerol was added to each reaction, and 45 μl of the reaction mixture was added to the acrylamide gel. The samples were run for 1 h at 140 V in 1 Tris-boric acid buffer followed by subsequent imaging with a Fujifilm FLA-5000 system using a 473-nm excitation wavelength. The fraction of bound DNA or RNA, $\theta$, was measured using Multi Gauge software (Fujifilm) to quantify the bound fractions (upper bands) and free DNA fractions (lower bands) from the polyacrylamide gel.

$$\theta = \frac{\text{Bound DNA}}{\text{Bound DNA} + \text{Free DNA}} \quad (\text{Eq. 5})$$

Trp lifetime assays were performed in 500-μl reactions consisting of 2 μM protein incubated with increasing amounts of nucleic acid in the EMSA binding buffer described above. The samples were excited at 295 nm in a 50-μl cuvette (Hellma) using an autosampler configuration to prevent Trp photobleaching. The laser intensity was adjusted using 4 μM N-acetyl-tryptophanamide as a standard to ensure a count rate between $8 \times 10^4$ and $1 \times 10^5$ per second prior to sample acquisition. Trp lifetime decays were measured for 2 min in 30-s intervals for each sample with ~60,000 counts total in the peak channel. The Trp lifetime decays were corrected for buffer contributions and subsequently fit to two exponential decays, which differed slightly, depending on protein construct and DNA/RNA. In comparison with samples containing only protein, the amplitudes of the ~3.8 and ~5.8 ns components decreased and increased, respectively, with increased DNA/RNA concentration. Thus, the amplitude of the faster phase represented the unbound protein, and the slower phase resulted from the DNA/RNA-bound protein.

$$\theta = \frac{\text{Amplitude} \sim 3.8 \text{ ns}}{\text{Amplitude} \sim 3.8 \text{ ns} + \text{Amplitude} \sim 5.8 \text{ ns}} \quad (\text{Eq. 6})$$
FIGURE 1. TDP-43 contains both structured and disordered regions. A, domain architecture of TDP-43 and aggregation (16) and disorder (17) propensity predictions based on the amino acid sequence. The aggregation algorithms TANGO (green) and WALTZ (blue) predict an aggregation-prone stretch in RRM2 (residues 228–233), with WALTZ predicting a second region in the glycine-rich domain (residues 314–319) and TANGO predicting a region near the nuclear localization sequence (NLS; purple). The disorder algorithm PONDR (orange) predicts high amounts of disorder in the C-terminal glycine-rich region (orange), where the ALS-causing mutations are located. Regions of disorder are also predicted in the linker, which lies between the two RRM domains as well as within the nuclear localization sequence. By contrast, very little disorder is predicted within the putative NES (yellow). B, the two RRM domains show high sequence similarity by ClustalW2 sequence alignment (*, identical; :, similar). Residues in β-strands and α-helices are colored blue and red, respectively, with the residues removed in RRM2c boxed in cyan. The predicted aggregation prone sequence in β3 is highlighted in boldface type, and the putative NES sequence is underlined. C and D, RRM domains are structurally similar based on solution NMR (C, RRM1 (Protein Data Bank entry 2cqg); D, RRM2 (Protein Data Bank entry 1wf0)). β-Strands are colored blue, α-helices are red, and loops are gray. The key phenylalanine residues identified for RNA binding in β3 are shown in boldface type (RRM1, Phe113 and Phe172; RRM2, Phe229 and Phe231), and intrinsic fluorophores for following tertiary structure folding are also shown (RRM1, Trp113 and Trp172; RRM2, Tyr232). The RRM2c construct, a model for the cleavage observed in disease, removes the first Phe147 and Phe149; RRM1, Phe229 and Phe231), and intrinsic fluorophores for following tertiary structure folding are also shown (RRM1, Trp113 and Trp172; RRM2, Tyr232). The RRM2c construct, a model for the cleavage observed in disease, removes the first Phe147 and Phe149; RRM1, Phe229 and Phe231), and intrinsic fluorophores for following tertiary structure folding are also shown (RRM1, Trp113 and Trp172; RRM2, Tyr232). The RRM2c construct, a model for the cleavage observed in disease, removes the first Phe147 and Phe149; RRM1, Phe229 and Phe231), and intrinsic fluorophores for following tertiary structure folding are also shown (RRM1, Trp113 and Trp172; RRM2, Tyr232). 

The percentage bound was determined as a function of DNA/RNA concentration and modeled to the quadratic binding equation (Equation 7) using Igor Pro (Wavemetrics, Inc.) to determine the apparent dissociation constant, $K_{d,app}$, (31),

$$\theta = b + (m - b) \times \left( \frac{(L + P + k_{d,app}) - \sqrt{(L + P + k_{d,app})^2 - 4LP}}{2L} \right)$$

where $L$ is the fixed ligand concentration (for EMSA, DNA/RNA; for Trp lifetime assays, RRM1 or tRRMs), $P$ is the independent variable (for EMSA, protein concentration; for Trp lifetime assays, DNA or RNA concentration), $K_{d,app}$ is the apparent dissociation constant, and $b$ and $m$ are the baseline and maximum percentage bound used to normalize the data sets.

RESULTS

RRM Domains Are Monomeric and Well Folded—Each isolated RRM domain (RRM1 and RRM2) could be expressed at high levels, was soluble to concentrations exceeding 5 mg ml$^{-1}$ (~0.5 mM), and was monomeric by size exclusion chromatography (Fig. 2A). Despite their structural similarity by NMR (Fig. 1, C and D), the CD spectra of the two isolated RRM domains were strikingly different from one another. Indeed, CD spectra can vary widely between RRM domains, including RNA-binding proteins that contain a single RRM domain, such as FUS/TLS (32), or multiple RRM domain-containing proteins, such as Musashi-1 (33) and U1A (34). RRM1 contains prominent minima at 208 and 218 nm, suggesting significant α-helical propensity for this domain. RRM1 contains a dramatically increased CD signal compared with RRM2 (Fig. 2B), consistent in part with the enhanced α-helical structure content of RRM1.
A Folding Intermediate in RRM2 of ALS-linked TDP-43

The isolated RRM domains are monomeric and contain significant secondary structure under physiological salt conditions. A, analytical size exclusion chromatography of the isolated (green, RRM1; blue, RRM2; orange, RRM2c) RRM domains reveals that each isolated domain is predominantly monomeric. The tethered domains (tRRMs, black) are also predominantly monomeric; however, some higher order species are also present, probably due to domain swapping. The arrows indicate retention times of molecular weight standards. B, RRM1 contains significant secondary structure compared with RRM2. RRM1 (green) has a minimum at 218 nm with a shoulder at 208 nm, whereas RRM2 (blue) has reduced ellipticity with a minimum at 210 nm and a shoulder at 220 nm. The cleavage fragment, RRM2c (orange) has further reduced ellipticity. All CD spectra were taken at 20 °C unless otherwise noted; physiological temperature (37 °C) does not alter the secondary structures of RRM2 (red) or RRM2c (purple). Tethering the RRRMs (RRRMs; black) results in a CD spectrum resembling the sum of the isolated domains (black dashed line). The CD spectrum for each protein, reported as mean residue ellipticity as a function of wavelength, did not vary in the range of 5–60 μM, and all proteins showed similar unfolded spectra when denatured in 7 M GdnHCl (blue dashed line).

(25%) compared with RRM2 (20%), as predicted by DSSP (35, 36), based on the NMR structures (Fig. 1B). The CD spectrum of RRM2 was approximately half as intense as that of RRM1, with a minimum at 210 nm and a shoulder at 230 nm. Although the two RRM domains have identical β-strand content (30%), the NMR structures indicate that RRM1 has more twist to its β-sheet compared with RRM2, consistent also with the increased intensity of the CD signal for RRM1 (37).

RRM2 Populates an Intermediate State—To probe the equilibrium folding mechanism of the RRM domains and its potential role in ALS, denaturant-induced unfolding was used to sample other protein conformations. The loss in secondary and tertiary structure was monitored by CD and FL, respectively (Fig. 3, A and B). Initial equilibrium folding studies performed in urea (data not shown) showed that RRM2 remains partially folded at high urea concentrations (>8 M). By contrast, both RRM1 and RRM2 were fully unfolded in the presence of 7.5 M GdnHCl.

RRM1 has a CD spectrum with two prominent minima (Fig. 2B) characteristic of an α + β protein. The two tryptophan residues, unique to RRM1 (Fig. 1C), monitor the tertiary structure within this domain upon unfolding. For RRM1 at 20 °C, the CD and Trp FL reveal a single cooperative transition between the native folded state and the unfolded state (Fig. 3A) that is well described by a two-state model for the free energy of folding in the absence of denaturant; ΔG°(20°C) = 3.7 kcal mol⁻¹ (Table 1). The midpoints (Cₘ) of the transition between these two states are coincident between the two spectroscopic techniques (Fig. 4, A and B), consistent with a two-state mechanism of folding for the RRM1 domain.

In comparison with RRM1, RRM2 contains reduced ellipticity and a single tyrosine as a fluorescence probe. The unfolding profile of RRM2 at 20 °C (Fig. 3B) monitored by CD is significantly different from that of RRM1. A change in the cooperativity observed at 4 M GdnHCl indicates a three-state unfolding process with contributions from a stably populated intermediate state. The transitions in the N ⇔ I ⇔ U three-state mechanism contribute 3.6 kcal mol⁻¹ (N ⇔ I) and 3.8 kcal mol⁻¹ (I ⇔ U) to the total stability of RRM2 (Table 1; 7.4 kcal mol⁻¹). The CD and FL transitions are not coincident, further supporting the presence of an intermediate state on the RRM2 unfolding pathway. The Tyr fluorescence becomes insensitive to denaturant in the second transition (>4 M GdnHCl), suggesting that the region surrounding this residue is unfolded in the intermediate state. Using the thermodynamic parameters for the individual transitions (N ⇔ I and I ⇔ U), the population of the intermediate state was calculated as a function of denaturant by Equation 3. As shown in Fig. 4C, the intermediate state is maximally populated (80%) at 4 M GdnHCl, and a small proportion (<1%) of RRM2 populates this partially unfolded intermediate state under native conditions (0 M GdnHCl).

Intermediate Population Is Enhanced by Increased Temperature or Cleavage of RRM2—in ALS patient tissues, fragments of TDP-43 comprising the C terminus and regions of RRM2 are present in cytoplasmic inclusions (9). One of these fragments of RRM2, starting at position Arg208, results in the removal of a region of the native fold in RRM2c. The fragment was predominantly monomeric by size exclusion chromatography (Fig. 2A); however, to enhance the solubility of RRM2c in the absence of denaturation, the 150 mM KCl was eliminated from the folding buffer.

The CD spectrum of RRM2c reveals that cleavage within the domain greatly reduces but does not obliterate its secondary structure (Fig. 2B). The fragment retains the three-state behavior of the RRM2 domain upon denaturation at 20 °C (Fig. 3D) and significant stability (Table 1; 4.1 kcal mol⁻¹). The transitions (N ⇔ I and I ⇔ U) are destabilized compared with the
intact domain (Table 1) and display non-coincident CD and tyrosine FL transitions that support a three-state unfolding model for RRM2c. The FL results suggest that the fragment can fold to a conformation that decreases the solvent accessibility of the Tyr, although the Tyr is close to the new N terminus. Comparing the fraction species plot of the intermediate state as a function of denaturant reveals that 2% of the RRM2c conformational ensemble samples the intermediate state under native conditions at 20 °C (Fig. 4B, compare orange and blue traces). These results indicate that fragments of RRM2 sample partially folded states to a greater extent than the intact domain and may provide a platform for subsequent TDP-43 aggregation (38).

Incubation of RRM1 at physiological temperatures (37 °C) resulted in aggregation (39, 40), whereas RRM2 and RRM2c

| Model | N ≅ U | N ≅ I | I ≅ U | N ≅ II | I ≅ I2 | I2 ≅ U |
|-------|-------|-------|-------|-------|-------|-------|
| CD/FL | 3.73±0.17 | 2.47±0.12 | 1.51±0.10 | 3.73±0.17 | 2.47±0.12 | 1.51±0.10 |
| CD | 3.60±0.15 | 1.30±0.06 | 2.77±0.17 | 3.60±0.15 | 1.30±0.06 | 2.77±0.17 |
| FL | 3.25±0.25 | 1.47±0.09 | 2.21±0.22 | 3.25±0.25 | 1.47±0.09 | 2.21±0.22 |
| CD (37°C) | 2.49±0.08 | 1.06±0.05 | 2.35±0.16 | 2.49±0.08 | 1.06±0.05 | 2.35±0.16 |
| CD | 2.37±0.24 | 1.43±0.14 | 1.66±0.14 | 2.37±0.24 | 1.43±0.14 | 1.66±0.14 |
| FL | 1.71±0.03 | 0.88±0.02 | 1.94±0.06 | 1.71±0.03 | 0.88±0.02 | 1.94±0.06 |
| CD (37°C) | 2.10±0.04 | 0.93±0.03 | 2.26±0.08 | 2.10±0.04 | 0.93±0.03 | 2.26±0.08 |
| CD/FL | 4.67±0.43 | 2.06±0.17 | 2.27±0.28 | 4.67±0.43 | 2.06±0.17 | 2.27±0.28 |

TABLE 1

Thermodynamic parameters of the isolated and tethered RRM domains

Figure 3. The isolated and tethered RRM domains have complex equilibrium unfolding profiles at 20 °C. A, equilibrium unfolding of RRM1 unfolds by a two-state mechanism (N ≅ U) by both CD (filled circles, solid line) and Trp FL (open circles, dashed line). B, RRM2 unfolds through the population of an equilibrium intermediate (N ≅ I ≅ U) by CD (open triangles, solid line). By contrast, the Tyr FL (open triangles, dashed line) unfolds in a two-state process that coincides with the N ≅ I transition observed by CD. C, tethering the two RMs stabilizes the RRM1 transition, resulting in a complex equilibrium unfolding profile with two stable intermediate states (N ≅ I1 ≅ I2 ≅ U) by CD (closed diamonds, solid line). Trp FL (open diamonds, dashed line) follows the denaturation of RRM1 only and coincides with the first transition (N ≅ I) by CD. D, a cleavage fragment of RRM2 (RRM2c) unfolds through a three-state mechanism by CD (filled squares, solid line) with destabilization of both transitions compared with RRM2. Similar to RRM2, the two-state transition observed by Tyr FL (open squares, dashed line) coincides with the N ≅ I transition by CD. Each experiment was consistently reproduced three times.
A Folding Intermediate in RRM2 of ALS-linked TDP-43

FIGURE 4. Cleavage and physiological temperatures destabilize RRM2 and increase the population of the intermediate state. A, fraction apparent plot of the Trp FL of RRM1 (open circles, green dashed line) and tRRMs (open triangles, black dashed line) at 20 °C reveals a shift toward higher denaturant for RRM1 unfolding upon tethering to RRM2. B, fraction apparent plots of all TDP-43 RRM constructs by CD. RRM1 displays a two-state equilibrium profile (green), whereas RRM2 (20 °C, blue; 37 °C, red) and the cleavage fragment, RRM2c (20 °C, orange; 37 °C, purple) display a three-state profile with the population of an intermediate state. tRRMs (black) display a complex unfolding profile with the population of two stable intermediates. C, RRM2 intermediate population as a function of denaturant using Equation 3 for a three-state profile (RRM2 and RRM2c) or Equation 4 for a four-state profile (tRRMs). Increasing temperature (red) or cleavage (orange and purple) within the RRM2 domain (blue) results in an increased intermediate population under native conditions (0 M GdnHCl) with a shift to lower denaturant for the maximum intermediate population. Tethering RRM2 to RRM1 (black) reduces the RRM2 intermediate state under native conditions and shifts the maximum population to higher denaturant. D, physiological temperatures destabilize the N → I transition by CD (filled inverted triangles, solid line) and Tyr FL transition (open inverted triangles, dashed line) with a shift in the C_m to lower denaturant compared with RRM2. E, physiological temperature results in a similar unfolding profile for RRM2c to RRM2 at 37 °C (compare filled inverted triangles with filled hexagons). Tyr FL revealed little change between the folded and unfolded states (open hexagons), suggesting that Tyr^214 resembles the unfolded state at this temperature. Each experiment was consistently reproduced three times.

remain in native conformations (Fig. 2B) with little change in secondary structure. Denaturation of the intact RRM2 at 37 °C (Fig. 4C) also follows a three-state unfolding process, albeit with a reduction in the overall stability (5.7 versus 7.4 kcal mol⁻¹). As with denaturation of RRM2 at 20 °C, the Tyr FL data were not coincident with the CD data (Table 1), supporting the presence of an intermediate in RRM2 even at elevated temperatures. Surprisingly, denaturation of RRM2c at 37 °C revealed a three-state unfolding profile coincident with the intact domain (Fig. 4B) but with a slight increase in total stability upon increasing temperature (4.1 versus 4.8 kcal mol⁻¹). The increase in overall free energy suggests either a self-association of RRM2c or that the folding of this fragment is driven by the hydrophobic effect, which is stronger at higher temperatures (41, 42). Self-association is concentration-dependent, and thus higher concentration would be expected to drive the association reaction and stabilize the protein against denaturation. However, our experiments revealed no concentration dependence for the unfolding transition by CD in the range from 15 to 60 μM RRM2c. Further, the lack of denaturant dependence of the tyrosine emission spectrum for RRM2c shows that the Tyr is exposed to solvent in this RRM2 fragment at physiological temperatures (Fig. 4E). This conformation differs from the Tyr FL at 20 °C, where the Tyr is at least partially buried in the native RRM2c structure. The destabilized native conformations of both RRM2 and RRM2c at physiological temperature suggest that the intermediate state becomes more populated under native conditions at increased temperatures. Indeed, the intermediate state is 5–10-fold more populated at 37 °C compared with 20 °C for both the intact RRM2 and the fragment of RRM2. Together, these results show that either a cleavage event within the RRM2 domain or physiological temperature increases the population of a potentially pathological intermediate state that may contribute to possible misfolding and aggregation.

Tethering the RRM Domains Results in Stabilization of RRM1 and Decreased Access to the RRM2 Intermediate—Under normal cellular conditions, the RNA binding domain of TDP-43 consists of RRM2 tethered to RRM1 by a short 15-amino acid linker. As such, the presence of RRM1 may influence the folding of RRM2 through mutual stabilizing interactions and decrease the population of the RRM2 intermediate state. The tRRM construct, which comprises RRM1, RRM2, and its natural linker, exhibits a CD spectrum that is very similar to the additive sum of the individual RRM domain spectra (Fig. 2B). This observa-
tethering RRM2c to RNA suggests that these domains may serve to stabilize RRM1 (Fig. 4A) rather than provide a major contribution to RNA binding. In fact, RRM1 binds with a weak affinity for UG repeats enhanced by 5-fold compared with the intact domain. This enhancement indicates that RRM2c may access conformations that favor RNA binding compared with the intact RRM2 domain. Interactions with DNA were much weaker than for RNA (>100 μM) and unmeasurable by EMSA. This weak binding of RRM2 and RRM2c to RNA suggests that these domains may serve to stabilize RRM1 (Fig. 4A) rather than provide a major contribution to RNA binding. In fact, RRM1 binds with a $K_{d,app}$ in the nanomolar range compared with the micromolar range of RRM2.

RMR2 binds weakly to RNA and enhances RRM1 affinity—TDP-43 is involved in multiple RNA processes, and thus, RNA probably plays a role in the folding and stability of the RRM domains. As shown in Fig. 4A, RRM1 becomes stabilized against denaturation when the two RRM domains are tethered, suggesting that this interaction could also enhance the affinity of RRM1 and possibly RRM2 for RNA. Two complementary nucleic acid binding assays, EMSA (Fig. 5A) and the tryptophan lifetime-based assay (Fig. 5B), were used to determine the dissociation constant, $K_{d,app}$, for each individual and tethered RRM domain. Based on the results of previous binding affinity measurements on various TDP-43 constructs (21), the oligonucleotides (UG)$_6$ and (TG)$_6$ were selected for comparing the affinities of each domain, and $T_{12}$ was selected as a control oligonucleotide.

In the EMSA, RRM2 bound to the UG repeat sequence very weakly with a dissociation constant of 17.0 μM (Table 2). Strikingly, RRM2c had a binding affinity for UG repeats enhanced by ~5-fold compared with the intact domain. This enhancement indicates that RRM2c may access conformations that favor RNA binding compared with the intact RRM2 domain. Interactions with DNA were much weaker than for RNA (>100 μM) and unmeasurable by EMSA. This weak binding of RRM2 and RRM2c to RNA suggests that these domains may serve to stabilize RRM1 (Fig. 4A) rather than provide a major contribution to RNA binding. In fact, RRM1 binds with a $K_{d,app}$ in the nanomolar range compared with the micromolar range of RRM2.
A Folding Intermediate in RRM2 of ALS-linked TDP-43

TABLE 2

Dissociation constants of the isolated and tethered RRM domains

| RRM domain | Assay     | Dissociation constant (nM) | (UG)_6 | (TG)_6 | T1/2 |
|------------|-----------|---------------------------|---------|---------|------|
| RRM1       | EMSA      | 49.3 ± 12.0               | 232 ± 48 | >8000   |      |
|            | FL        | 34.8 ± 14.1               | 630 ± 75 | >100,000 |      |
| RRM2       | EMSA      | 17,000 ± 3700             | >100,000 | >100,000 |      |
| RRM2c      | EMSA      | 2750 ± 480                | >100,000 | >100,000 |      |
| tRRMs      | FL        | 3.1 ± 1.3                 | 20.2 ± 1.8 | >8000   |      |

(10, 11, 15, 38, 57). The C terminus, which also contains a particularly aggregation-prone stretch (Met111–Met123), was shown to have increased β-strand propensity and aggregation tendency (38), whereas a Gln/Asn-rich region that is critical for proper TDP-43 protein–protein interactions was also sufficient for aggregation of GFP fused with multiple Gln/Asn repeats (58). Taken together, these results suggest that the population of the RRM2 intermediate state could expose aggregation-prone residues, such as β3 and β5 (56), either through the intact domain, cleavage by caspases (13, 59–61), or increased temperatures. This partially folded state could aberrantly interact with RRM1 (39, 62), the aggregation-prone C-terminal peptides (56, 63), or potential sequences within other proteins, such as heterogeneous nuclear ribonucleoprotein A1 and A2/B1 (64, 65), to promote misfolding and propagate the aggregation observed in ALS and FTLD patients.

Biophysical studies on TDP-43 have been hindered by the poor solubility and aggregation tendencies of the full-length protein. Thermal denaturation studies have shown that RRM2 remains well folded and resistant to denaturation beyond 85 °C, whereas RRM1 undergoes a conformational change at 50 °C (38, 66). These results suggest an inherent stability difference between the two domains, which are 30% identical in sequence by ClustalW2 (Fig. 1B). Similarly, in our study, the GdnHCl-induced denaturation of these domains revealed markedly different equilibrium unfolding profiles with significantly different stabilities (Table 1). GdnHCl denaturation fortuitously revealed a partially folded state that was not evident by thermal denaturation. Thus, chemical denaturation provides access to a partially folded state that may be relevant for both function and aggregation.

Hydrogen exchange experiments on the TIM barrel protein family have shown that clusters of isoleucine, leucine, and valine residues can accurately predict cores of stability (67) and early folding events (68, 69). Using an in-house algorithm (see the Clusters of Branched Aliphatic Side Chains in Globular Proteins (BASIC) Web site) (70, 71), clusters of the branched aliphatic side chain residues in RRM1 (Fig. 6D) and RRM2 (Fig. 6A) were identified using the NMR structures. As shown in Fig. 6, RRM2 contains one large cluster with 12 ILV residues, eight of which provide multiple contacts (Fig. 6C). RRM1 (Fig. 6E), on the other hand, contains three clusters of ILV residues with two single-contact clusters and a remaining cluster that consists of eight loosely connected residues. Upon tethering to RRM2 and the addition of RNA, the Leu156–Leu177 contact (Fig. 6, D and E, purple) and Val161 in a2 are incorporated into the RRM1 cluster and possibly enhance the stability of RRM1. In RRM2c (Fig. 6B), three ILV residues (Val183, Val186, and Leu207) (Fig. 5D, cyan) are removed, and the remaining contacts may provide a significantly different hydrophobic core for folding (Fig. 2A). This networked cluster of ILV residues in RRM2 and RRM2c (Fig. 6C) probably contributes to the presence of a stable intermediate and increased stability of RRM2 compared with RRM1 (Fig. 6E).

Alternatively, the RRM2 intermediate may serve a functional role in RNA processing. Normal TDP-43 shuttling to the cytoplasm is probably governed by a leucine-rich NES (72) located within RRM2 (residues 239–250). Alanine mutation of any of

(Aggregation of TDP-43. TANGO and WALTZ (16) predict an aggregation-prone region, β3 (Fig. 1A), localized within RRM2 that may serve as a template for aggregation propagation. In combination with the C terminus, intact or fragmented RRM2 was shown to severely enhance cellular aggregation and toxicity in cell models (12, 13, 15, 38, 57). The C terminus, which also contains a particularly aggregation-prone stretch (Met111–Met123), was shown to have increased β-strand propensity and aggregation tendency (38), whereas a Gln/Asn-rich region that is critical for proper TDP-43 protein–protein interactions was also sufficient for aggregation of GFP fused with multiple Gln/Asn repeats (58). Taken together, these results suggest that the population of the RRM2 intermediate state could expose aggregation-prone residues, such as β3 and β5 (56), either through the intact domain, cleavage by caspases (13, 59–61), or increased temperatures. This partially folded state could aberrantly interact with RRM1 (39, 62), the aggregation-prone C-terminal peptides (56, 63), or potential sequences within other proteins, such as heterogeneous nuclear ribonucleoprotein A1 and A2/B1 (64, 65), to promote misfolding and propagate the aggregation observed in ALS and FTLD patients.

Biophysical studies on TDP-43 have been hindered by the poor solubility and aggregation tendencies of the full-length protein. Thermal denaturation studies have shown that RRM2 remains well folded and resistant to denaturation beyond 85 °C, whereas RRM1 undergoes a conformational change at 50 °C (38, 66). These results suggest an inherent stability difference between the two domains, which are 30% identical in sequence by ClustalW2 (Fig. 1B). Similarly, in our study, the GdnHCl-induced denaturation of these domains revealed markedly different equilibrium unfolding profiles with significantly different stabilities (Table 1). GdnHCl denaturation fortuitously revealed a partially folded state that was not evident by thermal denaturation. Thus, chemical denaturation provides access to a partially folded state that may be relevant for both function and aggregation.

Hydrogen exchange experiments on the TIM barrel protein family have shown that clusters of isoleucine, leucine, and valine residues can accurately predict cores of stability (67) and early folding events (68, 69). Using an in-house algorithm (see the Clusters of Branched Aliphatic Side Chains in Globular Proteins (BASIC) Web site) (70, 71), clusters of the branched aliphatic side chain residues in RRM1 (Fig. 6D) and RRM2 (Fig. 6A) were identified using the NMR structures. As shown in Fig. 6, RRM2 contains one large cluster with 12 ILV residues, eight of which provide multiple contacts (Fig. 6C). RRM1 (Fig. 6E), on the other hand, contains three clusters of ILV residues with two single-contact clusters and a remaining cluster that consists of eight loosely connected residues. Upon tethering to RRM2 and the addition of RNA, the Leu156–Leu177 contact (Fig. 6, D and E, purple) and Val161 in a2 are incorporated into the RRM1 cluster and possibly enhance the stability of RRM1. In RRM2c (Fig. 6B), three ILV residues (Val183, Val186, and Leu207) (Fig. 5D, cyan) are removed, and the remaining contacts may provide a significantly different hydrophobic core for folding (Fig. 2A). This networked cluster of ILV residues in RRM2 and RRM2c (Fig. 6C) probably contributes to the presence of a stable intermediate and increased stability of RRM2 compared with RRM1 (Fig. 6E).

Alternatively, the RRM2 intermediate may serve a functional role in RNA processing. Normal TDP-43 shuttling to the cytoplasm is probably governed by a leucine-rich NES (72) located within RRM2 (residues 239–250). Alanine mutation of any of
the hydrophobic residues within this sequence resulted in nuclear mutants and endogenous aggregates (73). The NES forms /H92512 and /H92514 within RRM2 to possibly sequester the hydrophobic residues (Leu243, Leu248, and Ile250) and contribute to the ILV cluster (Fig. 6B). Another ALS-linked protein, FUS/TLS, also contains a NES within an RRM domain (74) with low affinity for RNA (75). Strikingly, we have observed that the RRM domain of FUS/TLS also unfolds through an equilibrium intermediate state.3 The presence of an intermediate state in the RRM unfolding pathways of TDP-43 and FUS/TLS suggests that these RRMs may serve to stabilize the other RNA-binding domain and sequester the NES within the hydrophobic core to control cytoplasmic shuttling or to prevent aggregation. RRM2 must partially unfold to expose the NES for recognition by exportin (76), which mediates the transport to the cytoplasm. This hypothesis suggests that cleavage within RRM2, loss of RRM1, or external stressors would increase the population of the intermediate state and disrupt the normal cellular distribution. Up-regulated export to the cytoplasm could increase the formation of dysfunctional complexes with other proteins and cytoplasmic aggregates that would ultimately decrease the population of functional nuclear TDP-43. Further experiments to investigate the role of the intermediate state on TDP-43 nuclear export would delineate the consequences of populating this partially folded state.

RRM2 may also enhance the function of TDP-43 through interactions with RRM1 and RNA. Tethering RRM2 to RRM1 increases the stability of RRM1 (Fig. 4C) and the affinity to UG-rich sequences (Fig. 5C). Studies on RNA binding in disease progression have indicated that removal of RRM1 or mutational analysis of key phenylalanine residues in RRM1 results in non-functional TDP-43 with neurotoxicity and aggregation phenotypes (21–23, 77, 78). A cell-free system revealed that incubation with (UG)₆ or (TG)₆ resulted in an increase in TDP-43 solubility and reduced the tendency to aggregate compared with mutant RRM1 TDP-43 (22). These results suggest

3 B. C. Mackness, M. T. Tran, S. P. McClain, C. R. Matthews, and J. A. Zitzewitz, unpublished data.

FIGURE 6. Extensive and networked ILV cluster in RRM2. A, solution NMR structure of RRM2 (Protein Data Bank code 1wf0) with isoleucine (I), leucine (L), and valine (V) residues highlighted that contact other ILV residues (blue). B, residues from the large cluster in RRM2 that are removed in RRM2c are highlighted in cyan, and the NES (residues 239–250), which is still part of the large cluster in RRM2c, is highlighted in yellow. C, ILV cluster contact map for RRM2 displayed on the secondary structure elements of RRM2. The RRM2 ILV cluster contains a single highly connected cluster (blue) spanning all α-helices and β-strands, and the cluster is unaffected by tethering to RRM1. The residues removed in RRM2c (cyan) are marked with an asterisk. The NES (yellow) contributes three residues to the large ILV cluster (Leu243, Leu248, and Ile250), which may contribute to the increased stability of RRM2 compared with RRM1. D, RRM1 (Protein Data Bank code 2cqg) contains three small clusters (purple, red, and green), two of which were single surface residue contacts (red and purple). The remaining cluster (green) is small and less connected compared with RRM2. E, the RRM1 cluster contains a small ILV cluster with single contacts centered in β1–β3. Upon tethering to RRM2 and RNA binding, the cluster exhibits a few contact changes that include the addition of Val161 and the purple cluster, resulting in a more networked RRM1 compared with the isolated RRM1.
that RNA binding may be playing a critical role in limiting access to the partially folded intermediate in RRM2 and reducing aggregation. RNA binding assays by our group and others (21, 66) suggest that RRM2 binds weakly to RNA compared with RRM1 (Table 2). RRM2 could contribute to RNA binding through two potential mechanisms: 1) RRM2 may contribute to binding through allosteric interactions with RRM1, and 2) RRM2 may indirectly contribute to binding by stabilizing RRM1. In either scenario, mutual interactions between RRM1 and RRM2 would reduce access to the RRM2 intermediate state. Structural information on the DNA/RNA-bound tethered domains would provide critical insights into the RNA-binding modes for each of these domains and could further delineate the role of RNA in the RRM folding pathway.

The thermodynamic and DNA/RNA binding experiments suggest a mechanism for potential TDP-43 dysfunction and aggregation through the RRM2 intermediate (Fig. 7). Normal cellular TDP-43 is involved in multiple RNA processes with protein partners in the nucleus. TDP-43 probably exists in equilibrium between two states: the folded TDP-43 and functional TDP-43 that populates the RRM2 intermediate. This intermediate could serve two potential roles: 1) a productive functional intermediate allowing for NES recognition by the cellular export machinery or 2) a non-productive misfolding intermediate with exposed aggregation-prone peptides, such as β3 and β5, that may aberrantly interact with other intact and fragmented TDP-43 as well as other known protein partners. Normally, only a small fraction of TDP-43 populates the intermediate state, and the equilibrium heavily favors the folded TDP-43 with a properly folded RRM2. The NES remains sequestered in the hydrophobic core of the molecule for the nuclear localization of TDP-43. However, cellular stresses, such as oxidative damage, could result in TDP-43 cleavage and loss in RNA binding or protein partners. Such events could potentially shift the equilibrium toward the misfolding of the intermediate state. These misfolded, non-functional proteins may increase transport to the cytoplasm, where exposure of the hydrophobic residues in the NES or aggregation-prone peptides in RRM2 results in the formation of dysfunctional complexes and aggregates of TDP-43. In any case, the reduced functional pool of TDP-43 in the nucleus would influence neuronal death. Thus, the RRM2 intermediate state may link the productive folding and function of TDP-43 with non-productive misfolding and aggregation that leads to disease progression. Therapeutic intervention strategies can be developed that take into account approaches that limit pathological access to this intermediate state, potentially offering a viable drug target for treating ALS and FTLD.

Acknowledgements—We thank Muska Hassan, Daniel Lindstrom, and Amanda Winn for help with molecular biology and/or protein purification and Dr. Sagar Kathuria for assistance with the hydrophobic cluster analysis program. We are grateful to Drs. Francesca Massi and Sean Ryder and the members of their respective laboratories for many helpful suggestions. We also thank Drs. Selma Aven-dano-Vazquez, Osman Bilsel, Daryl Bosco, Larry Hayward, Bill Royer, Zuoqiang Xu, and Chunxiao Yang for many insightful discussions.

REFERENCES

1. Da Cruz, S., and Cleveland, D. W. (2011) Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Curr. Opin. Neurol. 24, 904–919
2. Turner, M. R., Hardiman, O., Benatar, M., Brooks, B. R., Chio, A., de Carvalho, M., Ince, P. G., Lin, C., Miller, R. G., Mitsumoto, H., Nicholson, G., Ravits, J., Shaw, P. J., Swash, M., Talbot, K., Traynor, B. J., van den Berg, L. H., Veldink, J. H., Vucic, S., and Kiernan, M. C. (2013) Controversies and priorities in amyotrophic lateral sclerosis. Lancet Neurol. 12, 310–322
3. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, C., Clark, C. M., McCuskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314, 130–133
4. Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., and Oda, T. (2006) TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem. Biophys. Res. Commun. 351, 602–611
5. Chen-Plotkin, A. S., Lee, V. M., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease. Nat. Rev. Neurol. 6, 211–220
6. Lee, E. B., Lee, V. M., and Trojanowski, J. Q. (2012) Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. Nat. Rev. Neurosci. 13, 38–50
7. Buratti, E., and Baralle, F. E. (2008) Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. Front. Biosci. 13, 867–878
8. Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D. W. (2010) TDP-43 and FUS/TLS: emerging roles in RNA processing and neurode-
A Folding Intermediate in RRM2 of ALS-linked TDP-43

9. Igaz, L. M., Kwong, L. K., Xu, Y., Truxa, A. C., Uryu, K., Neumann, M., Clark, C. M., Elman, L. B., Miller, B. L., Grossman, M., McCluskey, L. F., Trojanowski, J. Q., and Lee, V. M. (2008) Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Am. J. Pathol.* 173, 182–194

10. Ross, C. A., and Poirier, M. A. (2004) Protein aggregation and neurodegenerative disease. *Nat. Med.* 10, S10–S17

11. Xu, Z.-S. (2012) Does a loss of TDP-43 function cause neurodegeneration? *Mol. Neurodegener.* 7, 27

12. Johnson, B. S., McCaffery, J. M., Lindquist, S., and Gitler, A. D. (2008) A genus of ordered proteins in human diseases: introducing the D2 concept. *Biopolymers* 89, 1018–1029

13. Yang, C., Tan, W., Whittle, C., Qiu, L., Cao, L., Akbarian, S., and Xu, Z. (2012) Does loss of TDP-43 function cause neurodegeneration? *PLoS One* 7, e15787

14. Daubner, G. M., Cléry, A., and Allain, F. H. (2013) RNA-RNA recognition motifs. Boring? Not quite. *Curr. Opin. Struct. Biol.* 24, 290–298

15. Cléry, A., Blatter, M., and Allain, F. H. (2008) RNA recognition motifs. Boring? Not quite. *Curr. Opin. Struct. Biol.* 23, 100–108

16. Buratti, E., and Baralle, F. E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* 276, 36337–36343

17. Huang, Y.-C., Lin, K.-F., He, R.-Y., Tu, P.-H., Koubek, J., Hsu, Y.-C., and Lin, C.-D. (2009) Expression of TDP-43 C-terminal fragments in vitro recapitulates pathological features of TDP-43 proteinopathies. *J. Biol. Chem.* 284, 8516–8524

18. Zhang, Y.-J., Kuo, P.-H., Chuang, C.-H., Liang, J.-R., Chen, Y.-R., Wang, S., Shen, S. C., and Yuan, H. S. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* 288, 9049–9057

19. Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takashashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takashashi, R., Kitahara, R., Ito, K., Fujitaka, N., and Ushiratani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* 288, 14886–14905

20. Chang, C.-K., Chuang, M.-H., Toh, E. K.-W., Chang, C.-F., and Huang, T.-H. (2013) Molecular mechanism of oxidation-induced TDP-43 RRM1 aggregation and loss of function. *FEBS Lett.* 587, 557–562

21. Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 83, 8069–8072

22. Schellman, J. A. (1997) Temperature, stability, and the hydrophobic interaction. *Biophys. J.* 73, 2960–2964

23. Luksavsky, P. J., Daujotyte, D., Tollervey, D., J. R., Ule, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* 20, 1443–1449

24. Rosen, D. R., Siddique, M., Patterson, B. M., Figlewicz, D. A., Sapp, P., Pentti, A., Donaldson, D., Goto, J., O’Regan, J. P., and Deng, H. X. (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362, 59–62

25. Kabashi, E., Valdmanis, P. N., Dion, P., Spiegelman, D., McConkey, B. J., Vande Velde, C., Bouchard, J.-P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P.-F., Camus, W., Meinger, V., Dupre, N., and Rouleau, G. A. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* 40, 572–574

26. Sreedharan, J., Blair, I. P., Tripathi, V. B., Xu, Y., Vance, C., Rogelj, B., Ackerley, S., Durnall, J. C., Williams, K. L., Buratti, E., Baralle, F., de Blerberoche, J., Mitchell, J. D., Leigh, P. N., Al-Chalabi, A., Miller, C. C., Nicholson, G., and Shaw, C. E. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 319, 1668–1672

27. Kawiakowski, T. J., Jr., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T., Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., de Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., Ticozzi, N., etc...
A Folding Intermediate in RRM2 of ALS-linked TDP-43

Siddique, T., McKenna-Yasek, D., Sapp, P. C., Horvitz, H. R., Landers, J. E., and Brown, B. R., Jr. (2009) Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208

48. Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P. J., Sapp, P. C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D. M., Kost, J. E., Gonzalez-Perez, P., Fox, A. D., Adams, J., Taroni, F., Tiloca, C., Leclerc, A. L., Chafe, S. C., Mangroo, D., Moore, M. J., Zitzewitz, J. A., Xu, Z.-S., van der Berg, L. H., Glass, J. D., Siciliano, G., Cirillii, E. T., Goldstein, D. B., Salachas, F., Meiningter, V., Rossoll, W., Ratti, A., Gellera, C., Bosco, D. A., Bassell, G. J., Silani, V., Dory, V. E., Brown, R. H., Jr., and Landers, J. E. (2012) Mutations in the prion 1 gene cause familial amyotrophic lateral sclerosis. *Nature* **489**, 499–503

49. DeJesus-Hernandez, M., Mackenzie, I. R., Boone, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. W., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Nature Neurol* **72**, 245–256

50. Ronton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollison, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetua, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Morck, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orr ell, R. W., Neal, J. M., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richard son, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peurlinna, T., Janson, I., Isoviita, V.-M., Kaivironne, A.-L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Renström, M., Aalto, S., Vilkka, A., Stolt, N., Rossoll, W., Wang, C.-H., and Huang, J. J. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268

51. Chen, S., Sayana, P., Zhang, X., and Le, W. (2013) Genetics of amyotrophic lateral sclerosis: an update. *Mol. Neurodegener*. **8**, 28

52. Youmans, K. L., and Wolozin, B. (2012) TDP-43. A new player on the AD field? *Exp. Neurol.* **237**, 90–95

53. Rayaprolu, S., Fujioka, S., Traynor, S., Soto-Ortolaza, A. I., Petrucelli, L., Liu, X., Niu, C., Ren, J., Zhang, X., Zhu, H., Feng, W., and Griffin, W. E. (2012) Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Science* **337**, 1108–1109

54. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. W., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256

55. Geser, F., Lee, V. M., and Trojanowski, J. Q. (2010) Amyotrophic lateral sclerosis. *Ann. Neurol.* **67**, 1577–1583

56. Baralle, F. E. (2005) TDP-43 binds heterogeneous nuclear ribonucleoprotein A2 through its C-terminal tail: an important region for the inhibition of cytosolic fibril transmembrane conductor regulate exon 9 splicing. *J. Biol. Chem.* **280**, 37572–37584

57. A. Ambrogio, A., Buratti, E., Stuani, C., Guarracina, C., Romano, A., Ayala, Y. M., and Baralle, F. E. (2009) Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res.* **37**, 4116–4126

58. Kuo, P.-H., Doudova, L. G., Wang, Y.-T., Chen, C.-K., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37**, 1799–1808

59. Gu, Z., Zitzewitz, J. A., and Matthews, C. R. (2007) Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: the roles of motif and sequence for the indole-3-glycerophosphate synthase from *Sulfobolus solfatarius*. *J. Mol. Biol.* **368**, 582–594

60. Gu, Z., Rao, M. K., Forsyth, W. R., Finke, J. M., and Matthews, C. R. (2007) Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerolphosphate synthase, by hydrogen exchange mass spectrometry and Gō model simulation. *J. Mol. Biol.* **374**, 528–546

61. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X., and Matthews, C. R. (2007) A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the α subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* **366**, 1624–1638

62. Kathuria, S. V., Day, I. J., Wallace, L. A., and Matthews, C. R. (2008) Kinetic traps in the folding of beta alpha-repeats proteins: CheY initially misfolds before accessing the native conformation. *J. Mol. Biol.* **382**, 467–484

63. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* **15**, 327–332

64. la Cour, T., Kiemer, L., Melgaard, A., Gupta, R., Skriver, K., and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.* **17**, 527–536

65. Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., and Lee, Y. M. (2008) Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *J. Biol. Chem.* **283**, 13302–13309

66. Ito, D., Seki, M., Tsunoda, Y., Uchiyama, H., and Suzuki, N. (2011) Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. *Ann. Neurol.* **69**, 152–162

67. Liu, X., Niu, C., Ren, J., Zhang, J., Xie, X., Zhu, H., Feng, W., and Gong, W. (2013) The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site. *Biochim. Biophys. Acta* **1832**, 375–385

68. Kawahara, Y., and Mieda-Sato, A. (2012) TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3347–3352

69. Ayala, Y. M., Zago, P., A. Ambrogio, A., Xu, Y.-F., Petrucelli, L., Buratti, E., and Baralle, F. E. (2008) Structural determinants of the cellular localization and shuttling of TDP-43. *J. Cell Sci.* **121**, 3778–3785

70. Ibara, R., Matsukawa, K., Nagata, Y., Kunugi, H., Tsuji, S., Chihara, T., Kuranaga, E., Miura, M., Wakabayashi, T., Hashimoto, T., and Iwatsubo, T. (2013) RNA binding mediates neurotoxicity in the transgenic *Drosophila* model of TDP-43 proteinopathy. *Hum. Mol. Genet.* **22**, 4474–4484
Folding of the RNA Recognition Motif (RRM) Domains of the Amyotrophic Lateral Sclerosis (ALS)-linked Protein TDP-43 Reveals an Intermediate State
Brian C. Mackness, Meme T. Tran, Shannan P. McClain, C. Robert Matthews and Jill A. Zitzewitz

J. Biol. Chem. 2014, 289:8264-8276. doi: 10.1074/jbc.M113.542779 originally published online February 4, 2014

Access the most updated version of this article at doi: 10.1074/jbc.M113.542779

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 76 references, 19 of which can be accessed free at http://www.jbc.org/content/289/12/8264.full.html#ref-list-1