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  
*University of Massachusetts Medical School*

Meme T. Tran  
*University of Massachusetts Medical School*

Shannan P. McClain  
*University of Massachusetts Medical School*

See next page for additional authors

Follow this and additional works at: [https://escholarship.umassmed.edu/gsbs_sp](https://escholarship.umassmed.edu/gsbs_sp)

Part of the [Biochemistry Commons](https://escholarship.umassmed.edu/biochem), [Nervous System Diseases Commons](https://escholarship.umassmed.edu/neurodis), [Neuroscience and Neurobiology Commons](https://escholarship.umassmed.edu/neurobiol), and the [Structural Biology Commons](https://escholarship.umassmed.edu/structbio)

Repository Citation

[https://escholarship.umassmed.edu/gsbs_sp/2046](https://escholarship.umassmed.edu/gsbs_sp/2046)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Student Publications by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state

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

Keywords
Amyotrophic Lateral Sclerosis (Lou Gehrig’s Disease), Circular Dichroism (CD), Fluorescence, Frontotemporal Lobar Degeneration (FTLD), Neurodegenerative Diseases, Protein Folding, Protein Misfolding, Protein-Nucleic Acid Interaction, RNA-binding Protein, Thermodynamics

Rights and Permissions
© 2014 by The American Society for Biochemistry and Molecular Biology, Inc. Publisher PDF posted after 12 months as allowed by the publisher’s author rights policy at http://www.jbc.org/site/misc/Copyright_Permission.xhtml.
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.

Amyotrophic lateral sclerosis (ALS)² 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...

The abbreviations used are: ALS, amyotrophic lateral sclerosis; FL, fluorescence; FTLD, frontotemporal lobar degeneration; GdnHCl, guanidine hydrochloride; RRM, RNA recognition motif; RRM1, first RRM domain, comprising residues 102–181; RRM2, second RRM domain, comprising residues 190–261; RRM2c, fragment of RRM2, comprising residues 208–261; RRM, tethered RRM domain construct, comprising residues 97–261; NES, nuclear export signal; TEV, tobacco etch virus.

*This work was supported, in whole or in part, by National Institutes of Health Grant GM54836 (to C. R. M.). This work was also supported by the ALS Therapy Alliance, Inc. (to J. A. Z.), the ALS Association (to J. A. Z.), and a University of Massachusetts Medical School Faculty Scholar Award (to J. A. Z.).

1To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 344 Plantation St., Worcester, MA 01605. Tel: 508-856-6739; Fax: 508-856-6464; E-mail: jill.zitzewitz@umassmed.edu.

2The abbreviations used are: ALS, amyotrophic lateral sclerosis; FL, fluorescence; FTLD, frontotemporal lobar degeneration; GdnHCl, guanidine hydrochloride; RRM, RNA recognition motif; RRM1, first RRM domain, comprising residues 102–181; RRM2, second RRM domain, comprising residues 190–261; RRM2c, fragment of RRM2, comprising residues 208–261; RRM, tethered RRM domain construct, comprising residues 97–261; NES, nuclear export signal; TEV, tobacco etch virus.
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 repeated 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 RRM domains (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 His$_6$ tag and either a PreScission or TEV protease cleavage site, respectively, for His$_6$ tag removal. The proteins were overexpressed in BL21 DE3 _Escherichia coli_ cells (Strategene) 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 Na$_p$, 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 His$_6$ tag. Minor contaminants were removed through ion exchange chromatography using S Sepharose (RRM1) or Q Sepharose (RRM2, tRRMs, and RRM2c) before dialysis into 10 mM K$_p$, 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 (246EDLLIKG252), 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 K$_p$, 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 thermostatic 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
A Folding Intermediate in RRM2 of ALS-linked TDP-43
corresponds to the single intermediate for RRM2, and its fractional population was calculated as defined in Equation 4,

\[ F_i = \frac{\exp\left(-\frac{\Delta G_{NI}}{RT}\right)}{1 + \exp\left(-\frac{\Delta G_{NI}}{RT}\right) + \exp\left(-\frac{\Delta G_{NU}}{RT}\right) + \exp\left(-\frac{\Delta G_{NT}}{RT}\right)} \]

(Eq. 4)

where \( \Delta G_{NI} = \Delta G_{NI}^0 - m_{NI[D]} \), \( \Delta G_{NU} = \Delta G_{NU}^0 - m_{NU[D]} \), and \( \Delta G_{NT} = \Delta G_{NT}^0 - m_{NT[D]} \).

Nucleic Acid Binding Assays—EMSA and tryptophan (Trp) lifetimes were used to determine the apparent binding affinity of the TDP-43 RRM domains to UGUGUGUGUGUGU ([U(U)]_6), TGTTGTGTGTGTG ([TG]_6) (Trp), and TTTTTTTTTTTTTT (T 12) 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 ng nucleotides incubated with increasing protein concentrations up to 8 μM. Binding reactions were performed in binding buffer, 10 mM KPi, 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 bromoresol blue in 30% glycerol was added to each reaction, and 45 μl of the mixture was added to the acrylamide gel. The samples were run for 1 h at 140 V in 1× Tris-boric acid 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.

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 × 10^4 and 1 × 10^6 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.

\[ \theta = \frac{\text{Amplitude } \sim 3.8 \text{ ns}}{\text{Amplitude } \sim 3.8 \text{ ns} + \text{Amplitude } \sim 5.8 \text{ ns}} \]

(Eq. 6)
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, $K_{d,app}$, for bound RNA.

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

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$^{-1}$ ($\sim$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 $\alpha$-helical propensity for this domain. RRM1 contains a dramatically increased CD signal compared with RRM2 (Fig. 2B), consistent in part with the enhanced $\alpha$-helical structure content of RRM1.
A Folding Intermediate in RRM2 of ALS-linked TDP-43

FIGURE 2. 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 (tRRMs; 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°H2O20°C = 3.7 kcal mol⁻¹ (Table 1). The midpoints (C_m) 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 tryptophan 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 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).

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 (tRRMs; 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).
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
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 plots 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 Cm 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 Tyr116 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. 4D) 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-
A Folding Intermediate in RRM2 of ALS-linked TDP-43

The slight differences may arise through a small decrease in α-helical content of the RRM1 domain when tethered, as suggested by the decrease in the 190 nm band and the shift in the ratio of the CD signal at 222 nm to 208 nm for tRRMs compared with the additive spectra (Fig. 2B).

However, upon denaturation with GdnHCl, tRRMs displayed a complex equilibrium unfolding profile when monitored by CD that was best described by N ↔ I1 ↔ I2 ↔ U, suggesting the population of two discrete intermediate states (Fig. 3C). To elucidate the extent that each RRM domain contributes to the intermediates, the Trp FL data were used as a constraint to define the first transition. Because Trp residues are only present within RRM1, the Trp FL provides structural insights into the unfolding of only RRM1 when tethered to RRM2. Interestingly, comparison of the Trp FL of the isolated RRM1 with the tethered RRRMs (Fig. 4A) reveals a shift in the transition midpoint (Cm) to higher denaturant (from 1.5 to 2.3 M) upon tethering. These results suggest that RRM1 and RRM2 interact in the absence of RNA, with interdomain interactions stabilizing the native state of each RRM domain by 0.9 kcal mol⁻¹ (Table 1). Indeed, a recent NMR study on the tethered RRM domain revealed that in the absence of RNA, the two RRMs do not tumble independently of one another (43), further supporting mutual stabilizing interactions between the domains. The Trp FL also suggests that RRM1 is completely unfolded at the first intermediate, I1, of the CD equilibrium unfolding profile. The extent of the RRM2 folding at the I1 intermediate is unclear because this domain lacks Trp residues, and the contributions from the single tyrosine residue are masked by RRM1. However, the thermodynamic parameters derived from the CD profile suggest that RRM2 is at least partially unfolded because the I1 ↔ I2 transition of tRRMs is not coincident with N ↔ I of RRM2 (Fig. 4B).

The third transition in the equilibrium profile of tRRMs (I2 ↔ U) corresponds to the I ↔ U transition of RRM2 (Table 1) because both transitions provide similar stability and midpoints. Therefore, the I2 species in the tethered domains probably correspond to the same RRM2 intermediate state. Using Equation 4 to identify the population of the RRM2 intermediate as a function of denaturant reveals that tethering shifts the maximal population of RRM2 to higher denaturant (5.0 M GdnHCl) compared with the individual RRM2 domain (4.0 M Gdn; Fig. 4C), suggesting that RRM1 inhibits the formation of the RRM2 intermediate. Indeed, based on the stability measurements from the four-state model, RRM1 contributes ~0.7 kcal mol⁻¹ of stability to the RRM2 native state, which would shift the N ↔ I transition to higher denaturant concentration and reduce the intermediate state population under native conditions. These results suggest that in the intact TDP-43 with RRM2 tethered to RRM1, mutual interactions between these domains can serve two purposes: 1) stabilizing RRM1 against populating unfolded conformations and 2) decreasing the RRM2 intermediate to almost negligible populations under native conditions (<0.1%). Isolating or fragmenting RRM2 removes stabilizing contributions from RRM1 and allows this region of TDP-43 to sample a potentially pathogenic partially folded state.

**RRM2 Binding 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 RRMs 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 trypanothione lifetime-based assay (Fig. 5B), were used to determine the dissociation constant, Kd, 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 T₁₂ 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 Kd in the nanomolar range compared with the micromolar range of RRM2.
Aggregation studies on RRM2 revealed increased formation of fibrillar aggregates with truncated RRM2 and peptide fragments (Table 2), supporting RRM1 as the major contributor to RNA binding. Notably, tethering the two RRM domains enhances binding affinity by an order of magnitude to result in dissociation constants for UG and TG repeat sequences in the low nanomolar range (Fig. 5C and Table 2). Trp lifetimes on the RRM domain reveal a reduction of Trp intensity and lifetime upon binding RNA (Fig. 5B) and provide structural insights into the binding interaction because the Trp residues are located on the β-sheet of RRM1. The RNA binding results suggest that the presence of RNA will play a role in defining the populations of all species on the folding free energy surface of TDP-43 by stabilizing the RRM domains and modulating access to the partially folded state in RRM2.

**DISCUSSION**

Although the ALS-linked proteins, such as SOD1 (44), TDP-43 (45, 46), FUS/TLS (47), profilin 1 (48), and C9ORF72 (49, 50) among others (51), are functionally different, common mechanisms for cellular toxicity and aggregation may govern the folding pathways of these proteins to result in disease. Indeed, the RNA-binding proteins, TDP-43 and FUS/TLS, have been linked not only to ALS but also to other neurodegenerative diseases, including FTLD-U (1), Alzheimer disease (52), and Parkinson disease (53), suggesting a potential common mechanism of disease pathogenesis between these proteins. Here, we performed denaturation and RNA binding studies to probe the equilibrium unfolding pathways of the isolated and tethered RRM domains of TDP-43. The data revealed a highly populated stable folding intermediate within RRM2 (Figs. 3B and 4C) that may link the pathways governing TDP-43 folding and function with those of misfolding and aggregation. Specifically, the intermediate state may have a normal cellular function in nuclear export, but at the same time, populating this intermediate state may sow the seeds for misfolding and aggregation in disease.

A major focus of the TDP-43 field has been to investigate the potential impacts of TDP-43 loss-of-function and aggregation propensity on the propagation of ALS and FTLD (1, 8, 11, 54, 55). Aggregation studies on RRM2 revealed increased formation of fibrillar aggregates with truncated RRM2 and peptide fragments consisting of β3 and β5 (56). These results suggest that the core sequences within RRM2 may directly participate in driving the 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 (Met311–Met323), 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 more stable 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 Leu106-Leu177 contact (Fig. 6, D and E, purple) and Val161 in α2 are incorporated into the RRM1 cluster and possibly enhance the stability of RRM1. In RRM2c (Fig. 6B), three ILV residues (Val193, Val195, 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 /H9251 and /H9252 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

---

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.

**REFERENCES**

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.,
m ents in TAR DNA-binding protein-43 cytoplasmic inclusions in brain
but not in spinal cord of frontotemporal lobar degeneration and amy-
generative disease. Nat. Med. 10, S10–S17
Mol. Neurodegener. 7, 27
stable form of recombinant tobacco etch virus protease.
J. Biol. Chem. 283, 4051–4055
EMBO J. 20, 215–246
A Folding Intermediate in RRM2 of ALS-linked TDP-43


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

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