RNA-binding proteins with basic-acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer's disease

Isaac Bishof, Emory University
Eric Dammer, Emory University
Duc M. Duong, Emory University
Sean R. Kundinger, Emory University
Marla Gearing, Emory University
James Lah, Emory University
Allan Levey, Emory University
Nicholas Seyfried, Emory University

Journal Title: Journal of Biological Chemistry
Volume: Volume 293, Number 28
Publisher: American Society for Biochemistry and Molecular Biology |
2018-07-13, Pages 11047-11066
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1074/jbc.RA118.001747
Permanent URL: https://pid.emory.edu/ark:/25593/v2p99

Final published version: http://dx.doi.org/10.1074/jbc.RA118.001747

Copyright information:
© 2018 Bishof et al.

Accessed February 5, 2020 6:42 AM EST
The U1 small nuclear ribonucleoprotein 70 kDa (U1-70K) and other RNA-binding proteins (RBPs) are mislocalized to cytoplasmic neurofibrillary Tau aggregates in Alzheimer’s disease (AD), yet the co-aggregation mechanisms are incompletely understood. U1-70K harbors two disordered low-complexity domains (LC1 and LC2) that are necessary for aggregation in AD brain extracts. The LC1 domain contains highly repetitive basic (Arg/Lys) and acidic (Asp/Glu) residues, referred to as a basic-acidic dipeptide (BAD) domain. We report here that this domain shares many of the properties of the Gln/Asn-rich LC domains in RBPs that also aggregate in neurodegenerative disease. These properties included self-assembly into oligomers and localization to nuclear granules. Co-immunoprecipitations of recombinant U1-70K and deletions lacking the LC domain(s) followed by quantitative proteomic analyses were used to resolve functional classes of U1-70K-interacting proteins that depend on the BAD domain for their interaction. Within this interaction network, we identified a class of RBPs with BAD domains nearly identical to that found in U1-70K. Two members of this class, LUC7L3 and RBM25, required their respective BAD domains for reciprocal interactions with U1-70K and nuclear granule localization. Strikingly, a significant proportion of RBPs with BAD domains had elevated insolubility in the AD brain proteome. Furthermore, we show that the BAD domain of U1-70K can interact with Tau from AD brains but not from other tauopathies. These findings highlight a mechanistic role for BAD domains in stabilizing RBP interactions and in potentially mediating co-aggregation with the pathological AD–specific Tau isoforms.

The molecular processes that contribute to neurodegenerative diseases are not well understood. Recent observations suggest that numerous neurodegenerative diseases are promoted by the accumulation of RNA-binding protein (RBP) aggregates (1–3). This includes Alzheimer’s disease (AD), where pathological RNA–protein aggregates are often, but not exclusively, associated with Tau neurofibrillary tangles in brain (4). For example, U1 small nuclear ribonucleoprotein 70 kDa (U1-70K) and other core components of the spliceosome complex co-aggregate with Tau in both sporadic and familial human cases of AD, but not other tauopathies (5–7). Furthermore, RNA-sequencing analysis from AD and control brains revealed a significant accumulation of unspliced pre-mRNA–related transcripts in AD consistent with a loss of U1-spliceosome function (7, 8). Currently, our knowledge of the specific mechanisms underlying U1-70K and Tau co-aggregation in AD is limited. This has proved to be a barrier to developing cellular models that would further our understanding of U1-70K and related RBP aggregation events in the pathogenesis of AD. Supporting evidence indicates that a select group of RBPs are poised for aggregation because they self-assemble to form structures, including RNA granules (9), which are membrane-free organelles composed of RNA and RBPs (1, 9–12). RNA granules form via liquid–liquid phase separation (LLPS), which is driven by a dynamic network of multivalent interactions between structurally disordered low complexity (LC) domains (13–15) that have limited diversity in their amino acid composition (16). LLPS allows specific RBPs to concentrate and separate, leading to the formation of higher-order structures, including oligomers, granules, and ultimately aggregates (16–18). Notably, several RBPs that aggregate...
BAD RNA-binding proteins aggregate in AD

gate in neurodegenerative disease contain LC domains, including TDP-43 and fused in sarcoma (FUS) (19–21). The LC domains found in TDP-43 and FUS mediate self-association, are necessary for RNA granule formation, and polymerize into amyloid-like aggregates (9, 22, 23). Mutations within the LC domains of TDP-43 and FUS cause amyotrophic lateral sclerosis (ALS) and increased RNA granule stability, highlighting a critical role for LC domains in disease pathogenesis (24–26).

We recently reported that human AD brain homogenates induce the aggregation of soluble U1-70K from control brain as well as recombinant U1-70K, making U1-70K detergent-insoluble (5). The C terminus of U1-70K, which contains two LC domains (LC1 and LC2), is necessary for this aggregation (5). Furthermore, the LC1 domain (residues 231–308) of U1-70K is sufficient for robust aggregation, and through cross-linking studies it was found to directly interact with insoluble U1-70K in AD brain homogenates (5). Collectively, these observations led to a hypothesis that pathological aggregation of RBPs in neurodegenerative diseases, including U1-70K, is driven by LC domains. However, unlike the prion-like Gln/Asn-rich LC domains of TDP-43 and FUS, the LC1 domain of U1-70K contains highly repetitive basic (Arg/Lys) and acidic (Asp/Glu) residues that we refer to as a Basic-Acidic Dipeptide (BAD) domain. These structurally unique motifs were originally described by Perutz (27), who proposed their ability to self-assemble and form higher-order structures termed polar zippers. Currently, the physiological role of BAD domains in U1-70K and other RNA-binding proteins is unclear, and understanding their role in protein–protein interactions may shed light on the mechanisms underlying RBP aggregation and their association with Tau in AD (7, 28).

Here, we report that the BAD domain of U1-70K shares many of the properties of the Gln/Asn-rich LC domains found in TDP-43 and FUS, despite having a vastly different amino acid composition. These properties include the ability to self-assemble with homotypic selectivity into high-molecular weight oligomers and to associate with nuclear granules in cells. Coupling quantitative proteomics and network-based systems biology, we mapped classes of U1-70K–interacting proteins that show a dramatic reduction in their association with Tau under native conditions (7, 28). These complexes were more evident following the transfer to a membrane and Western blot analysis with Myc antibodies (Fig. 2C). Notably, a higher proportion of the LC1/BAD domain formed dimers compared with trimers (n = 2 and n = 4) compared with trimers (n = 3), suggesting that dimer intermediates are favored over trimer intermediates for tetramer formation (Fig. 2D). These in vitro findings demonstrate that the LC1/BAD domain can directly self-associate to form oligomers, including high molecular weight species, which implicates homologous species for physiological conditions is not known. To test this question, we overexpressed full-length recombinant GST-fused, Myc-tagged U1-70K (rU1-70K) in HEK293 cells with serial deletions lacking one or both LC domains followed by co-immunoprecipitation (co-IP) and Western blot analysis (Fig. 1A). Full-length rU1-70K (WT) and the ΔLC2 mutant co-IP endogenous U1-70K (~55 kDa), while rU1-70K variants that lack the LC1/BAD domain (ΔLC1 and ΔLC1 + 2), show a significant decrease in the ability to co-IP native U1-70K. Together, these observations indicate that the LC1 domain is necessary for U1-70K self-association.

To determine whether the LC1 domain is sufficient for self-association, co-IPs were performed following the overexpression of specific rU1-70K domains (Fig. 1B), including the N-terminal domain alone (residues 1–99), the N terminus, and RNA recognition motif (residues 1–181), the LC1/BAD domain alone, and the LC2 domain alone. Only the LC1/BAD domain was sufficient for self-association with native U1-70K. Furthermore, this interaction was likely not influenced by the presence of RNA, as treatment of the lysates with RNase prior to co-IP did not impair the ability of rU1-70K to interact with native U1-70K (Fig. 1C). Thus, our findings show that the LC1/BAD domain is necessary and sufficient for U1-70K self-association and that this interaction is predominantly RNA-independent.

U1-70K BAD domain oligomerizes in vitro

Although our results show that the LC1/BAD domain of rU1-70K is sufficient to interact with native U1-70K in cells, it is unclear whether this interaction is direct or facilitated by indirect interactions with additional RBPs. To determine whether the LC1/BAD domain of U1-70K can directly self-associate, we performed blue native gel-PAGE (BN-PAGE) of the GST-purified LC1/BAD domain (residues 231–310) and N-terminal domain (residues 1–99) of rU1-70K; the latter was unable to interact with native U1-70K (Fig. 1B). In contrast to SDS-PAGE, which resolves proteins under denaturing conditions, BN-PAGE is used to determine native protein complex masses, including high molecular weight oligomeric states, and to identify physiological protein–protein interactions (29). Under the denaturing conditions of SDS-PAGE (Fig. 2A), both the LC1 and N-terminal domain have equivalent molecular masses (~65 kDa) compared with purified GST (~20 kDa). However, under native conditions (Fig. 2B), the BAD domain formed dimers, trimers, tetramers, and high-molecular mass oligomers in the megadalton range (>1,236 kDa). In contrast, the N-terminal domain mainly existed in the monomeric and dimeric state with some evidence of lower abundance high-molecular mass oligomers; GST alone was almost exclusively monomeric (Fig. 2B). These complexes were more evident following the transfer to a membrane and Western blot analysis with Myc antibodies (Fig. 2C). Notably, a higher proportion of the LC1/BAD domain formed dimers (n = 2) and tetramers (n = 4) compared with trimers (n = 3), suggesting that dimer intermediates are favored over trimer intermediates for tetramer formation (Fig. 2D). These in vitro findings demonstrate that the LC1/BAD domain can directly self-associate to form oligomers, including high molecular weight species, which implicates

Results

BAD domain of U1-70K is necessary and sufficient for self-association

The U1-70K LC1 domain (residues 231–308), which we also refer to as the LC1/BAD domain, contains highly repetitive basic (Arg/Lys) and acidic (Asp/Glu) residues. This domain is necessary and sufficient for robust aggregation in AD brain homogenates (5). However, whether this domain is required for endogenous U1-70K self-association under physiological conditions is unclear. To test this question, we overexpressed full-length recombinant GST-fused, Myc-tagged U1-70K (rU1-70K) in HEK293 cells with serial deletions lacking one or both LC domains followed by co-immunoprecipitation (co-IP) and Western blot analysis (Fig. 1A). Full-length rU1-70K (WT) and the ΔLC2 mutant co-IP endogenous U1-70K (~55 kDa), while rU1-70K variants that lack the LC1/BAD domain (ΔLC1 and ΔLC1 + 2), show a significant decrease in the ability to co-IP native U1-70K. Together, these observations indicate that the LC1 domain is necessary for U1-70K self-association.

To determine whether the LC1 domain is sufficient for self-association, co-IPs were performed following the overexpression of specific rU1-70K domains (Fig. 1B), including the N-terminal domain alone (residues 1–99), the N terminus, and RNA recognition motif (residues 1–181), the LC1/BAD domain alone, and the LC2 domain alone. Only the LC1/BAD domain was sufficient for self-association with native U1-70K. Furthermore, this interaction was likely not influenced by the presence of RNA, as treatment of the lysates with RNase prior to co-IP did not impair the ability of rU1-70K to interact with native U1-70K (Fig. 1C). Thus, our findings show that the LC1/BAD domain is necessary and sufficient for U1-70K self-association and that this interaction is predominantly RNA-independent.

U1-70K BAD domain oligomerizes in vitro

Although our results show that the LC1/BAD domain of rU1-70K is sufficient to interact with native U1-70K in cells, it is unclear whether this interaction is direct or facilitated by indirect interactions with additional RBPs. To determine whether the LC1/BAD domain of U1-70K can directly self-associate, we performed blue native gel-PAGE (BN-PAGE) of the GST-purified LC1/BAD domain (residues 231–310) and N-terminal domain (residues 1–99) of rU1-70K; the latter was unable to interact with native U1-70K (Fig. 1B). In contrast to SDS-PAGE, which resolves proteins under denaturing conditions, BN-PAGE is used to determine native protein complex masses, including high molecular weight oligomeric states, and to identify physiological protein–protein interactions (29). Under the denaturing conditions of SDS-PAGE (Fig. 2A), both the LC1 and N-terminal domain have equivalent molecular masses (~65 kDa) compared with purified GST (~20 kDa). However, under native conditions (Fig. 2B), the BAD domain formed dimers, trimers, tetramers, and high-molecular mass oligomers in the megadalton range (>1,236 kDa). In contrast, the N-terminal domain mainly existed in the monomeric and dimeric state with some evidence of lower abundance high-molecular mass oligomers; GST alone was almost exclusively monomeric (Fig. 2B). These complexes were more evident following the transfer to a membrane and Western blot analysis with Myc antibodies (Fig. 2C). Notably, a higher proportion of the LC1/BAD domain formed dimers (n = 2) and tetramers (n = 4) compared with trimers (n = 3), suggesting that dimer intermediates are favored over trimer intermediates for tetramer formation (Fig. 2D). These in vitro findings demonstrate that the LC1/BAD domain can directly self-associate to form oligomers, including high molecular weight species, which implicates
Figure 2. LC1/BAD domain of U1-70K directly self-interacts and oligomerizes in vitro. A, SDS-PAGE of GST alone, purified N-terminal domain (N-term), and purified LC1/BAD domain of rU1-70K. Both the LC1/BAD and N-term domains have equivalent molecular masses (~65 kDa), whereas GST alone is (~20 kDa). B, BN-PAGE of GST alone, the N-terminal domain, and the LC1 domain of rU1-70K, respectively. The LC1/BAD domain formed higher molecular weight species (*) consistent with dimers (~130 kDa), trimers (~195 kDa), and high-molecular weight (HMW) oligomers (~400 kDa). C, Western blotting detection of blue native complexes using Myc antibodies. D, densitometry of monomeric, dimeric, trimeric, and high-molecular weight species of the N-term domain (blue) and LC1/BAD domain (red) of rU1-70K. Each form is represented as the fraction of total signal intensity in each sample analyzed in technical replicate (n = 2). Error bars represent standard deviation.

Figure 1. LC1/BAD domain of U1-70K is necessary and sufficient for self-association. A, full-length (WT) recombinant GST-fused, Myc-tagged U1-70K (rU1-70K) and variants lacking one or both LC domains (ΔLC1, ΔLC2, and ΔLC1 + ΔLC2) were overexpressed in HEK293 cells and immunoprecipitated (IP) with anti-Myc antibodies. IP with a nonspecific IgG was also performed from mock-transfected cells as a negative control. Western blotting for recombinant Myc-tagged proteins (green) and native U1-70K (red) are shown for both the inputs and co-IPs (A, bottom panels). B, full-length WT and rU1-70K truncations, including the N terminus (1–99 residues) alone, the N terminus and RRM (1–181 residues), LC1/BAD alone (231–308 residues), and the LC2 domain alone (residues 317–407). IP with a nonspecific IgG was also performed from mock-transfected cells as a negative control. Western blotting for recombinant Myc-tagged proteins (green) and native U1-70K (red) are shown for both the inputs and co-IPs. C, WT rU1-70K was immunoprecipitated from untreated and RNase (50 ng/μl)-treated lysates followed by Western blotting for the Myc tag recombinant protein (green) and native U1-70K (red). IP with a nonspecific IgG was also performed from mock-transfected cells as a negative control.
direct BAD–BAD domain interactions as a mechanism of U1-70K self-association.

**U1-70K BAD domain is necessary and sufficient for robust nuclear granule localization**

To explore whether the LC1/BAD domain influences nuclear localization and granule formation in cells, full-length rU1-70K or variants lacking one or both LC domains (Fig. 1A) were overexpressed in HEK293 cells. The cells were then fractionated into nuclear and cytoplasmic pools followed by Western blot (WB) analysis for both recombinant Myc-tagged proteins (green) and native U1-70K (red). Western blotting for histone H3 (bottom panel) was used as a positive control in the nuclear fraction. A, densitometry analysis was performed to calculate the levels of cytoplasmic and nuclear WT rU1-70K and variants, and the percent nuclear intensity for each rU1-70K protein is reported. Each experiment was performed in biological triplicate (n = 3) with error bars representing the standard deviation (S.D.). Both the ΔLC1 and ΔLC1 + 2 rU1-70K fragments were significantly less nuclear than full-length rU1-70K (**, p value < 0.01 by ANOVA compared with WT).

**Figure 3. LC1/BAD domain of U1-70K is necessary and sufficient for nuclear granule localization.** A, WT full-length rU1-70K or variants lacking one or both LC domains were overexpressed in HEK293 cells. The cells were then fractionated into nuclear and cytoplasmic pools followed by Western blot (WB) analysis for both recombinant Myc-tagged proteins (green) and native U1-70K (red). Western blotting for histone H3 (bottom panel) was used as a positive control in the nuclear fraction. A, densitometry analysis was performed to calculate the levels of cytoplasmic and nuclear WT rU1-70K and variants, and the percent nuclear intensity for each rU1-70K protein is reported. Each experiment was performed in biological triplicate (n = 3) with error bars representing the standard deviation (S.D.). Both the ΔLC1 and ΔLC1 + 2 rU1-70K fragments were significantly less nuclear than full-length rU1-70K (**, p value < 0.01 by ANOVA compared with WT). B, densitometry analysis was performed to calculate the levels of cytoplasmic and nuclear WT rU1-70K and variants, and the percent nuclear intensity for each rU1-70K protein is reported. Each experiment was performed in biological triplicate (n = 3) with error bars representing the standard deviation (S.D.). Both the ΔLC1 and ΔLC1 + 2 rU1-70K fragments were significantly less nuclear than full-length rU1-70K (**, p value < 0.01 by ANOVA compared with WT). C, immunocytochemistry for WT rU1-70K and deletion variants that lacked the LC1/BAD, LC2, or both LC domains was performed and visualized by confocal microscopy. Scale bar equates to 10 μm. D, overexpression of the rU1-70K LC1/BAD domain alone (red) resulted in nuclear granule association and sequestration of native U1-70K (green). The EM439 antibody detects an extreme C-terminal epitope not present in the LC1/BAD rU1-70K protein, which allows discrimination between the recombination protein and native U1-70K. DAPI-stained nuclei are shown in blue. Scale bar equates to 10 μm.

**BAD RNA–binding proteins aggregate in AD**
ment for the LC1/BAD domain in subnuclear granule localization. Furthermore, the LC1/BAD domain alone co-localizes with native U1-70K in nuclear granules (Fig. 3D), consistent with the ability of the LC1/BAD domain to interact with native U1-70K from cell lysates (Fig. 1B). Collectively, these results demonstrate that the LC1/BAD domain is important for U1-70K subcellular nuclear localization and suggest a role for LC1/BAD-mediated intermolecular interactions as a mechanism of nuclear granule formation.

**Protein–protein interaction network analysis resolves functionally distinct classes of U1-70K–interacting proteins**

To further assess the physiological function of the LC domains of U1-70K, we performed co-IPs of full-length rU1-70K and various rU1-70K variants lacking either or both LC domains from HEK293 cells. Co-interacting proteins were identified by LC coupled to tandem MS (LC-MS/MS). Each co-IP was performed in biological quadruplicate (n = 4), and an equal number of mock IPs were performed using a nonspecific immunoglobulin (IgG) as a negative control. Protein abundance was determined by peptide ion–intensity measurements across LC-MS/MS runs using the label-free quantification (LFQ) algorithm in MaxQuant (33). In total, 45,223 peptides mapping to 3,458 protein groups were identified. To limit the number of nonspecific interactors, proteins with less than a 1.5-fold enrichment over IgG were not considered. This resulted in the final quantification of high-confidence interactors falling into 716 protein groups mapping to 713 unique gene symbols (Table S1).

The Weighted co-expression network analysis (WGCNA) is typically used for large-scale transcriptome and proteome datasets to categorize gene products into biologically meaningful complexes, molecular functions, and cellular pathways (34). Here, we sought to leverage co-enrichment patterns to better classify protein–protein interactions (PPIs) across WT and rU1-70K deletions to determine whether specific classes of proteins selectively favor interactions with the LC domains. In WGCNA, correlation coefficients between each protein pair in the dataset are calculated, and groups of highly correlated proteins are segregated into modules (35). In our dataset, a total of seven modules were defined (Fig. 4A). These modules range from 292 proteins in turquoise to 19 in black (Table S1). The premise of co-expression, or in this case protein co-enrichment analysis, is that the strong correlation between two or more proteins is indicative of a physical interaction, functional relationship, and/or co-regulation. We therefore hypothesized that following a co-IP for rU1-70K, specific modules would reflect biologically relevant PPIs and thus highlight distinct complexes. As expected, modules were significantly enriched for biologically meaningful gene ontologies (GO) as well as established cellular functions and/or organelles as determined by GO-Elite (Table 1).

Each module has an abundance profile for all member proteins across the rU1-70K co-IP conditions, termed the eigenprotein (Fig. 4, A and B). Notably, six (M1–6) of the seven modules showed a significantly higher level of co-enrichment in the WT and rU1-70K deletions compared with the IgG negative controls, indicative of specific interactions for members of these modules (Fig. 4, A and B). In contrast, M7 (Fig. 4A, black) had essentially equivalent levels across all conditions and was the only module with positive correlation to the nonspecific IgG. Therefore, it was considered nonspecific and not considered for further analysis (Fig. 4A). The protein interactors with reduced affinity for rU1-70K following LC1/BAD domain deletion (e.g. ΔLC1 and ΔLC 1/2) include members of both the mRNA processing (blue) and snRNP assembly (green) modules (Table 1). Therefore, the LC1/BAD domain-dependent interactors are defined by membership in these modules. In contrast, modules enriched with large ribosomal subunit components (turquoise) and mitochondrial ribosome subunits (brown) display increased levels following deletion of the LC1/BAD domain or both LC domains, suggesting that the LC1/BAD domain negatively regulates their interactions with U1-70K (Table 1 and Fig. 4, A and B). Finally, the modules enriched with proteins involved in protein folding (Fig. 4, A and B, red) and the small ribosomal subunit (yellow) show little difference across the co-IP conditions, suggesting that these protein interactions are mainly with the N terminus and/or RRM domain of U1-70K. The hub proteins, with the highest correlation to the module abundance profile (i.e. eigenprotein), are highlighted in Fig. 4B. These findings demonstrate that a weighted PPI network analysis of the U1-70K interactome successfully resolves biologically and structurally distinct complexes.

**Confirmation of U1-70K BAD domain–dependent interacting proteins**

To validate and extend the module assignments, we performed both in silico and biochemical analysis. First, to visualize the relationships among modules with an independent clustering method, the T-distributed stochastic neighbor embedding (tSNE) algorithm was used to map the relatedness of proteins of top module members. The tSNE analysis largely agreed with and confirmed module assignments, whereby the majority of proteins clustered with their own module members as assigned by WGCNA (Fig. 5A). The tSNE analysis also allows for visualization of module relatedness, with similar modules in close proximity to each other and dissimilar modules further apart. For example, modules involved in translation cluster together (Fig. 5A, brown and turquoise), whereas those involved in mRNA processing (blue) and snRNP assembly (green) formed a separate cluster.

To experimentally validate the module assignments, co-IP of WT rU1-70K and deletion variants was performed followed by Western blotting for interactors of the snRNP assembly module (SNRPD1 and U1A), mRNA-processing module (SRSF1, RRM25, and LUC7L3), and the protein-folding module (TDP-43) (Fig. 5B). Members of both the mRNA processing and snRNP assembly modules have higher abundance in the WT and ΔLC2 co-IPs compared with that of the IgG, ΔLC1, and ΔLC1 + 2 IP samples. This mirrored the pattern observed for the blue and green eigenprotein values, confirming the proteomic findings (Fig. 4B). In contrast, TDP-43 shows a similar level of interaction across WT rU1-70K and the various deletions, consistent with TDP-43 being an N-terminal interactor.
of U1-70K and not influenced by the absence of the disordered LC domains.

**mRNA processing module is enriched with RNA-binding proteins containing BAD domains**

U1-70K–interacting proteins that map to the mRNA processing and snRNP assembly modules are related by their affinity for the LC1/BAD domain, yet they contain RBPs with distinct biological functions (Table 1). For example, all U1 snRNP components and assembly factors, including U1A, U1C, Sm proteins, and the SMN complex, are enriched in the green module. The SMN complex is responsible for loading the Sm proteins onto the snRNA scaffold, a critical step in U1 snRNP assembly (36). This module also contains the components of the 7SK snRNP, which regulates snRNA transcription and is present in Cajal bodies, a site of U1 snRNP maturation (37, 38). In contrast, the mRNA processing module (blue) is enriched with proteins associated with RNA splicing, polyadenylation, mRNA export, and “nuclear specks,” the latter being an analogue term for splicing speckles (39). However, it is their respective association with rU1-70K LC deletion variants that discriminates the mRNA processing and snRNP module members. For example, proteins involved in snRNP assembly are less influenced by the loss of the LC2 domain, yet associations of members of the mRNA-processing module are affected, suggesting that proteins involved in granule/speckle assembly interact in part via both LC domains of U1-70K, whereas core spliceosome assembly factors do not favor interactions with the LC2 domain (40, 41).

Our observations also reveal that several members of the mRNA-processing module (blue) contain stretches of highly repetitive basic (Arg/Lys) and acidic (Asp/Glu) dipeptides, analogous to the LC1/BAD domain of U1-70K (Fig. 6A). To examine the relationship between this sequence similarity and U1-70K–interacting proteins, a list of LC1-like (i.e. BAD domain) proteins was created using the Uniprot protein Basic Local Alignment Search Tool (BLAST) feature. Many proteins (n = 255) in the proteome were determined to have significant sequence overlap to the BAD domain of U1-70K. These include other members of the mRNA-processing module such as RBM25, ZC3H18, DDX46, and LUC7L3 among others. Although not identical in length, sequence alignment highlights the similar stretches of highly repetitive basic (Arg/Lys) and acidic (Asp/Glu) dipeptides across these distinct proteins (Fig. 6A). Indeed, a one-tailed Fisher’s exact test revealed that the mRNA-processing module is significantly enriched with proteins harboring BAD domains (Fig. 6B). In contrast, a similar analysis comparing disordered RNA-binding proteins containing prion-like (Gln/Asn-rich) domains (42), including TDP-43 and FUS, showed no enrichment in any of the modules of U1-70K–interacting proteins. Notably, the mRNA-processing module also has a significant over-representation of nuclear proteins that selectively precipitate after treatment with biotinylated isoxazole (43). Many of these proteins participate in RNA granule assembly and form hydrogels in vitro (9, 44). Collectively, these results suggest that structurally similar BAD domains, analogous to the U1-70K LC1/BAD domain, engage in protein–protein interactions, which are essential for nuclear granule assembly and mRNA processing.

**BAD domains in LUC7L3 and RBM25 are necessary for reciprocal interactions with U1-70K and nuclear granule localization**

Based on their related structural and functional roles in nuclear speckle assembly and affinity for the BAD domain of U1-70K, we tested whether members of the mRNA-processing module co-localize with U1-70K in cells. Both RBM25 and LUC7L3 contain BAD domains analogous to the LC1/BAD domain of U1-70K with similar E-values of 5.5E-27 and 1.8E-26, respectively, and amino acid overlap of 55.1 and 44.2%, respectively (Fig. 6A). As expected, all three proteins localize to nuclear granules (30, 45, 46), where U1-70K shows strong co-localization with LUC7L3 and RBM25 (Fig. 7A). Our current findings support a model where the LC1/BAD domain is necessary and sufficient for U1-70K self-association and nuclear granule association. By extension, we hypothesized that the BAD domains in LUC7L3 and RBM25 would similarly be important in mediating interactions with U1-70K and other structurally similar proteins. To test this possibility, we overexpressed full-length recombinant GST-fused, Myc-tagged rRBM25 or rLUC7L3 in HEK293 cells, and their respective deletion variants lacking the BAD domains (ΔBAD) as well as the BAD domains alone followed by IP and Western blot analysis (Fig. 7B). The full-length rLUC7L3 and the BAD domain were each able to co-IP endogenous LUC7L3, mirroring the self-association observed for U1-70K (Fig. 1A). Unfortunately, the rLUC7L3-ΔBAD domain protein migrated at a similar molecular weight to endogenous LUC7L3, and thus, we were unable to determine whether the BAD domain is necessary for self-association in cells. Full-length rLUC7L3 and the BAD domain interact with endogenous U1-70K and RBM25 (Fig. 7B), whereas the rLUC7L3-ΔBAD variant does not (Fig. 7B).

Similarly, full-length rRBM25 interacts with both endogenous U1-70K and LUC7L3, whereas the rRBM25-ΔBAD variant does not (Fig. 7C). In contrast, the BAD domain of rRBM25 is not sufficient to interact with U1-70K or LUC7L3, perhaps due to misprocessing, post-translation modifications, or size.

Given the role of the U1-70K LC1/BAD domain in nuclear granule localization, we sought to determine whether the BAD domains of LUC7L3 and RBM25 influence nuclear localization
of these proteins to granules (Fig. 7, D and E). Both full-length rLUC7L3 and rRBM25 localize to nuclear granules by immunocytochemistry, consistent with the endogenous LUC7L3 and RBM25 localization pattern in cells (Fig. 7A). However, rLUC7L3/BAD diffusely localizes to the cytoplasm, whereas rRBM25/BAD is primarily nuclear but not localized to nuclear granules. Consistent with U1-70K, the BAD domain of rLUC7L3 was sufficient to localize to nuclear granules, likely due to interactions with U1-70K and other BAD RBPs. However, the BAD domain of rRBM25, which does not interact by co-IP with BAD RBPs, does not localize to nuclear granules in cells. Taken together, our findings suggest a shared functional role for BAD domains in stabilizing protein–protein interactions that likely play a role in nuclear granule assembly.

**RNA-binding proteins with BAD domains have elevated insolubility in AD brain**

Based on the ability of U1-70K to aggregate in AD brain homogenate, and the key role of the BAD domain in U1-70K oligomerization *in vitro*, we hypothesized that proteins harboring similar BAD domains would preferentially aggregate in AD brain. To test this hypothesis, we assessed the distribution of insoluble proteins with BAD domains in a recently published comprehensive analysis of the Sarkosyl-insoluble proteome (n = 4,643 proteins quantified) from individual human control and AD cases (47). Protein ratios for all pairwise comparisons (i.e. control versus AD) were converted into log2 values and the resulting histogram fit to a normal Gaussian distribution (Fig. 8A). Compared with the normal distribution of all proteins in the AD insoluble proteome (blue histogram), quantified BAD proteins (yellow histogram) show a global shift toward insolubility in AD (Fig. 8A). This increase is significant using a one-tailed Fisher exact test (p value = 2.028868e-09). Consistently, BAD proteins within the top 10th percentile (n = 28) are significantly elevated in AD cases compared with controls, similar to Aβ and Tau levels (Fig. 8, B and C). Strikingly, 68% of the AD-enriched BAD proteins within the top 10th percentile, including LUC7L3, are members of the blue module identified

---

**Table 1**

U1-70K protein–protein interaction network generates modules enriched with specific gene ontology (GO) terms

<table>
<thead>
<tr>
<th>Module color</th>
<th>Ontology type</th>
<th>Top GO terms</th>
<th>Fisher exact p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turquoise (n = 292)</td>
<td>Biological process</td>
<td>ncRNA processing, rRNA metabolic process, Endocrine pancreas development</td>
<td>2.07E-11</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>Nucleic acid binding, Structural constituent of ribosome, Protein–DNA loading ATPase activity</td>
<td>9.99E-09</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Cytosolic large ribosomal subunit, Nucleolus, Intracellular</td>
<td>2.16E-10</td>
</tr>
<tr>
<td>Blue (n = 190)</td>
<td>Biological process</td>
<td>mRNA processing, RNA splicing, mRNA export from nucleus</td>
<td>1.01E-44</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>RNA binding, RS domain binding, Nucleotide binding</td>
<td>6.57E-06</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Nuclear speck, Spliceosomal complex, Nucleus</td>
<td>3.55E-14</td>
</tr>
<tr>
<td>Brown (n = 61)</td>
<td>Biological process</td>
<td>Translation, Structural constituent of ribosome, Translation regulator activity</td>
<td>1.82E-15</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>Mitochondrion, Ribosome</td>
<td>9.67E-34</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Mitochondrial large ribosomal subunit</td>
<td>1.51E-11</td>
</tr>
<tr>
<td>Yellow (n = 46)</td>
<td>Biological process</td>
<td>Endocrine pancreas development, Viral transcription, Viral infectious cycle</td>
<td>9.99E-09</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>Structural constituent of ribosome, mRNA binding</td>
<td>1.28E-20</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Cytosolic small ribosomal subunit, Ribosome</td>
<td>1.78E-44</td>
</tr>
<tr>
<td>Green (n = 40)</td>
<td>Biological process</td>
<td>Spliceosomal snRNP assembly, Regulation of cyclin-dependent protein kinase activity</td>
<td>2.67E-20</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>snRNA binding, Small nuclear ribonucleoprotein complex, U12-type spliceosomal complex</td>
<td>4.85E-08</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Ciliary body, Intraflagellar body</td>
<td>1.05E-07</td>
</tr>
<tr>
<td>Red (n = 20)</td>
<td>Biological process</td>
<td>Protein folding, Response to abiotic stimulus, Oxidation–reduction process</td>
<td>3.65E-06</td>
</tr>
<tr>
<td></td>
<td>Molecular function</td>
<td>Heat-shock protein binding, Purine ribonucleoside triphosphate binding</td>
<td>7.28E-07</td>
</tr>
<tr>
<td></td>
<td>Cellular component</td>
<td>Microtubule, Membrane, Intrinsic to membrane</td>
<td>4.79E-05</td>
</tr>
</tbody>
</table>

BAD RNA-binding proteins aggregate in AD

in the rU1-70K interactome studies (Fig. 4), with shared functions in RNA binding, splicing, and processing (Fig. 8, D and E). However, RBM25 is not significantly elevated in AD brain, despite containing a BAD domain. Thus, although RBPs with BAD domains clearly have a higher likelihood of insolubility and aggregation in the AD brain, the presence of a BAD domain alone is not sufficient for aggregation.

BAD domain of U1-70K interacts with pathological Tau specifically from AD brain

We have previously reported an association of aggregated U1-70K with Tau neurofibrillary tangles in both sporadic and familial cases of AD but not in other tauopathies (5–7, 48). However, mechanisms underlying the specificity of Tau-U1-70K co-aggregation in AD are poorly understood. Similar to the biophysical properties of RBPs, recent evidence indicates that Tau undergoes LLPS in vitro (49). This process is enhanced by polyamines, such as spermidine (49) and RNA (50), as well as phosphorylation on Tau (49). Based on these observations, we sought to assess whether the LC1/BAD domain of U1-70K interacts with pathological Tau from the human AD brain. Equivalent amounts of GST-purified, Myc-tagged LC1/BAD domain or, as a control, the N-terminal domain of rU1-70K were added to AD brain homogenates and immunoprecipitated with anti-Myc antibodies followed by a Western blotting for Tau (Fig. 9A). Compared with the N-terminal domain, the LC1/BAD domain of rU1-70K co-immunoprecipitated significantly more Tau, including modified Tau species of altered molecular weights (Fig. 9, A and B).

To test whether the interaction between the BAD domain of rU1-70K and Tau was specific to the AD brain, purified LC1/BAD or the N-terminal rU1-70K proteins were added to homogenates generated from control (n = 6), AD (n = 6), or non-AD tauopathy (n = 6) brain tissue. The latter group included progressive supranuclear palsy (n = 1) and corticobasal degeneration (n = 5) cases (51, 52). Following IP with anti-Myc antibodies, samples were analyzed by MS to identify and quantify Tau. Consistent with Western blotting results, significantly more Tau in AD brain homogenates interacted with the LC1/BAD domain compared with the N-terminal domain (Fig. 9, A–C). Furthermore, MS analysis revealed that the LC1/BAD domain interacts with pathological Tau from AD brain homogenates but not other tauopathies (Fig. 9C). This suggests that BAD domains in U1-70K and other RBPs mediate interactions with pathological Tau isoforms specific to AD.

Discussion

In this study, we show that the BAD domain of rU1-70K can directly self-interact in vitro to form high molecular weight oligomers and that this domain is also necessary and sufficient for U1-70K self-association in cells. Using quantitative proteomics, functional classes of U1-70K-interacting proteins were identified that favored interactions with the BAD domain. This analysis revealed a class of structurally similar RBPs that also contained analogous BAD LC domains. We show that for at least two other RBPs, LUC7L3 and RBM25, their respective BAD domains are required for reciprocal interactions with U1-70K and for proper localization to nuclear granules. Comprehensive

**Figure 5.** Confirmation of U1-70K-interacting partners that favor interactions via the LC1/BAD domain. A, to visualize the relationship between modules and validate the WGCNA results, the tSNE algorithm was used to map the relatedness of proteins with a kME score of 0.5 or greater, which is a measure of intramodular connectivity, defined as the Pearson correlation between the expression pattern of a protein and the module eigenprotein. The tSNE analysis overlaid with module assignments determined by WGCNA allows for visualization of module relatedness, with the distance between proteins representing similarity of co-enrichment, with the more similar clusters of proteins (related modules) in closer proximity to each other compared with dissimilar modules. Proteins with similar co-enrichment across the rU1-70K co-IPs are highly correlated to one another and are related to modules with distinct biological functions (Table 1). B, Western blot analysis for select interactors of the snRNP assembly (SNRPD1 and U1A), mRNA processing (SRSF1, RBM25, and LUC7L3), and protein folding (TDP-43) modules following co-IP across the five experimental conditions (IgG, WT rU1-70K, and deletions ΔLC1, ΔLC2, and ΔLC1 + ΔLC2).
analysis of the detergent-insoluble proteome in human brain revealed elevated levels of BAD RBPs in AD. Finally, we show that the LC1/BAD domain of U1-70K (residues 231–308) contains highly repetitive dipeptide repeats of basic (Arg/Lys) and acidic (Asp/Glu) residues. A list of 255 proteins that shared greater than 20% similarity to the LC1/BAD domain of U1-70K (E-values less than 0.005) was created using the Uniprot protein BLAST feature. Using Clustal Omega, an alignment was performed on the U1-70K LC1/BAD domain and the four most structurally similar proteins to highlight the repetitive basic and acidic residues in the sequence. Residues that can be phosphorylated (Ser/Thr/Tyr) are also highlighted as they can be negatively charged after modification. A one-tailed Fisher’s exact test was used to assess structural overlap of LC1-like BAD proteins from BLAST analysis with module membership for U1-70K-interacting partners (upper panel). The same analysis was repeated using Gln/Asn-rich prion-like RNA-binding proteins (middle panel) or RNA-binding proteins that were precipitated from nuclear extracts using biotin–isoxazole compound (bottom panel). Benjamini-Hochberg multiple comparison corrected p values for the module enrichment are highlighted. Significance is demonstrated by the color scales, which go from 0 (white) to ≥3 (red), representing −log(p).

We propose that BAD proteins be considered a class of proteins due to their related biological function and shared primary structure. For example, we provide evidence to support the function of the LC1/BAD domain in U1-70K and more broadly other BAD domains in nuclear granule assembly. Thus, under physiological conditions the reciprocal interactions of BAD domains could form the “glue” that drives granule assembly. BAD domains, such as those found in U1-70K, RBM25, and LUC7L3, have been proposed to self-assemble through the formation of polar zippers (27). Furthermore, the promiscuous nature of BAD domain interactions could be crucial to facilitating the complicated and multifaceted cooperative function of RNA-processing proteins (16, 53). The presence of distinct LC domains like the LC2 (317–407 residues) in U1-70K may further refine the dynamics of this process. For example, U1-70K interacts with both the spliceosome and members of the polyadenylation complex, including FIP1L1. Analogous to U1-70K, FIP1L1 harbors a BAD domain and is the top hub protein in the blue module (54). Thus, the presence of BAD domains in U1-70K enables physical, if not also functional, cross-talk between the role of U1-70K in 5’-splice site recognition and the polyadenylation complex in mRNA processing.

Nuclear U1-70K is found mislocalized to cytoplasmic Tau-immunoreactive neurofibrillary aggregates in AD neurons (6), which may contribute to a loss of spliceosome function, given the recently identified RNA splicing deficits in the disease (8). Mislocalization of other RBPs also contributes to neurodegenerative disease (55, 56). Here, we show that the LC1/BAD domain is important for nuclear localization of U1-70K, supporting a link between aberrant BAD domain interactions and mislocalization of U1-70K. Our findings also shed light on previous studies in which the C-terminal domain (residues 161–437) of U1-70K was found to be sufficient for nuclear localization (30, 57). Moreover, our ΔLC1 + 2 deletion mutant mimics findings for the 1–199 U1-70K C-terminal truncation (57), wherein this N-terminal fragment localized to the nucleus but not to granules. Perhaps, the nuclear localization sequence within the LC1/BAD domain was missed by earlier studies due to the selected sites of truncation, as the LC1/BAD domain was never expressed in its entirety (57).

We show that U1-70K and other BAD proteins share many of the same properties as the Gln/Asn-rich prion-like RBPs, despite their difference in primary sequences. These properties include the homotypic selectivity to self-assemble into high molecular weight oligomers, localize to nuclear granules in cells, and promote aggregation (58). The formation of RNA granules has been viewed as an intermediary step toward pro-
tein aggregation (1, 9, 59), and our observations place U1-70K and other BAD RBPs among prion-like RBPs, such as TDP-43, FUS, hnRNPA1, and TIA-1 that associate with granules and aggregate in neurodegenerative disease (4, 21, 26, 60). Furthermore, repeat-associated non-ATG (RAN) translation of C9orf72 generates dipeptide repeat proteins (DPRs) that form pathological aggregates in ALS and frontotemporal dementia (61). Similar to our findings for the LC1/BAD domain of U1-70K, MS analysis of C9orf72 DPR-interacting proteins showed that arginine-containing DPRs, poly-Gly–Arg and
poly-Pro–Arg, interact with RBPs with LC domains that mediate the assembly of RNA granules (62). Thus, we hypothesize that the ability of the LC1/BAD domain within U1-70K and other RBPs to self-interact poise them for pathological aggregation in neurodegenerative diseases, which is consistent with their increased insolubility in AD.

Both U1-70K and Tau co-localize to neurofibrillary tangles in late-onset sporadic and familial cases of AD but not in other tauopathies (5–7, 48, 63). U1-70K also aggregates in preclinical or asymptomatic AD cases (63), defined by significant Aβ deposition in the absence of significant cortical Tau deposition and cognitive impairment. Although mechanisms underlying the relationship between Aβ, Tau, and U1-70K aggregation are incompletely understood, our findings suggest that Aβ precedes and influences U1-70K interactions with Tau in brain. For example, many RBPs shuttle between the nucleus and cytoplasm (3), potentially bringing them into contact with Tau and/or under the influence of downstream intracellular signaling events triggered by extracellular Aβ (63). These signaling events could influence Tau post-translational modification and tertiary structures that favor interactions with U1-70K and BAD RBPs in AD. For example, Tau undergoes LLPS via electrostatic interactions in vitro, referred to as coacervation (50, 64, 65), mediated by the Tau microtubule-bundling repeats (residues 244–369) (66). This is notable, as a recent study examining the physical structure of Tau filaments in AD brain revealed an exposed BAD motif within the tertiary structure of Tau, composed of residues 338–358 in the microtubule-binding repeat domain (67). Thus, it is tempting to speculate that pathological Tau may behave like other BAD RBPs and sequester U1-70K to neurofibrillary tangles or vice versa (Fig. 10). Specifically, we propose that this BAD surface in pathological Tau in part mediates physical interactions between Tau and the U1-70K LC1/BAD domain. These structural conformations of Tau are likely specific to AD brain, compared with other tauopathies, which may in part underlie its selective co-aggregation with U1-70K and other BAD RBPs in AD. Furthermore, Tau–U1-70K hetero-oligomers may have a unique aggregation propensity, although additional determinants of aggregation may reside in the cytoplasm, including RNA (48).

In summary, we have identified novel functional roles for BAD domains in protein–protein interactions, nuclear localization, granule assembly, and pathological aggregation. We show similarities between BAD domains and Gln/Asn-rich domains found in RBPs and DPRs that aggregate in neurodegenerative diseases. We also demonstrate how a weighted protein–protein interaction network analysis can be used to resolve biologically and structurally distinct complexes. Notably, RBPs with BAD domains showed elevated insolubility in AD brain, and the BAD domain of U1-70K is sufficient to interact with pathological Tau in AD brain but not other tauopathies. These findings support a hypothesis where BAD domains in U1-70K and related RBPs could mediate cooperative interactions with Tau isoforms specific to AD.

**Materials**

Primary antibodies used in these studies include the following: an in-house rabbit polyclonal antibody raised against a synthetic keyhole limpet hemocyanin-conjugated peptide corresponding to a C-terminal epitope of U1-70K (EM439) (7); an anti-Myc tag (clone 9B11, Cell Signaling); an anti-GST (catalog no. ab6613, Abcam) antibody; an anti-LUC7L3 (catalog no. HPA018484-100UL, Sigma); an anti-RBM25 (ab72237, Abcam); an anti-histone H3 (catalog no. ab1791, Abcam); an anti-U1-70K monoclonal (catalog no. 05-1588, Millipore); an anti-Tau (catalog no. ab54193, Abcam); and IgG mouse control (catalog no. 550339, Pharmingen). Secondary antibodies were conjugated to either Alexa Fluor 680 (Invitrogen) or IRDye800 (Rockland) fluorophores.

**Plasmids and cloning**

The original cDNA of U1-70K containing C-terminal Myc and DDK tags was cloned from pCMV6-Entry vector (Origene) and inserted into the HindIII/BamHI sites in the pCDNA3.1 vector (5). Full-length and U1-70K deletions sequences were subsequently cloned into the EcoRV/Xhol sites in the pLEX-GST vector for the expression of N- and C-terminal GST-tagged proteins. Similar cloning strategies were performed using LUC7L3 (catalog no. RG214406, Origene) and RBM25 (catalog no. RC212256, Origene) plasmids. All cloning was performed by the Emory Custom Cloning Core Facility, and plasmids were confirmed by DNA sequencing.

**Immunoprecipitation**

Human embryonic kidney (HEK) 293T cells (CRL-3216, ATCC) were cultured in Dulbecco’s modified Eagle’s medium (high glucose (Gibco)) supplemented with 10% (v/v) fetal bovine serum (Gibco) and penicillin/streptomycin (Gibco) and maintained at 37 °C under a humidified atmosphere of 5% (v/v) CO2 in air. For transient transfection, the cells were grown to 80–90% confluence in 10-cm2 culture dishes and transfected with 10 μg of expression plasmid and 30 μg of linear polyethamid dom in LUC7L3 and RBM25 are necessary for reciprocal interactions with U1-70K and nuclear granule localization. A, immunocytochemistry (ICC) was performed to assess the co-localization of native U1-70K (green) with RBM25 (red) or LUC7L3 (red). DAPI-stained nuclei are shown in blue. Scale bar equates to 10 μm. B, full-length (WT) recombinant GST-fused, Myc-tagged LUC7L3 (rLUC7L3), and variants lacking the BAD domain or the BAD domain alone (upper panel) were overexpressed in HEK293 cells and immunoprecipitated with anti-Myc antibodies. IP with a nonspecific IgG was also performed from mock-transfected cells as a negative control. Western blotting (WB) for recombinant Myc-tagged proteins (green) and native LUC7L3 (red) are shown for both the inputs and co-IPs (bottom panels). Membranes were also re-probed for native U1-70K (red) or RBM25 (red). C, full-length (WT) recombinant GST-fused and Myc-tagged RBM25 (rRBM25) and deletion variants lacking the BAD domain or the BAD domain alone (upper panel) were overexpressed in HEK293 cells and immunoprecipitated with anti-Myc antibodies. A nonspecific IgG was also used on mock-transfected cells as a negative control. Western blotting for recombinant Myc-tagged proteins (green) and native U1-70K (red) or LUC7L3 (red) are shown for both the inputs and co-IPs (bottom panels). D, immunocytochemistry for full-length rLUC7L3, a variant lacking the BAD domain (LUC7L3-ΔBAD), and the BAD domain alone (LUC7L3-BAD) were expressed in HEK293 cells and visualized by confocal microscopy. E, immunocytochemistry for full-length rRBM25, a variant lacking the BAD domain (RBM25-ΔBAD), and the BAD domain alone (RBM25-BAD) were expressed in HEK293 cells and visualized by confocal microscopy. DAPI was used to visualize nuclei (blue). Scale bars are 10 μm for both D and E.
BAD RNA–binding proteins aggregate in AD
yleneimine. Cells were homogenized in ice-cold immunoprecipitation (IP) buffer containing 50 mM HEPES, pH 7.4, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 0.5% (v/v) Nonidet P-40, 0.5% (v/v) CHAPS, Halt phosphatase inhibitor mixture (1:100, ThermoFisher Scientific). Samples were sonicated for 5 s on 5 s off at 30% amplitude for a total of 1.5 min (13 cycles). The samples were cleared (14,000 g for 10 min), and protein concentrations were determined using a standard bicinchoninic acid (BCA) assay (Pierce). Protein A–Sepharose 4B beads (catalog no. 101042, Invitrogen; 20 µl per IP), were washed twice in IP buffer and then blocked with 0.1 mg/ml BSA (catalog no. 23209, ThermoFisher Scientific) and washed three additional times in IP buffer. Anti-Myc (4 µg) mouse mAb (catalog no. 2276, Cell Signaling) or 4 µg of IgG control (catalog no. 550339, Pharmin-
gen) was allowed to incubate rotating with the bead slurry in IP buffer (500 μl) for a minimum of 90 min to allow antibody conjugation to beads. Beads were washed three times in IP buffer. The pre-cleared protein lysates were added to beads (1.5 mg per IP) and incubated by rotating overnight at 4 °C. The beads were washed three times in IP wash buffer (IP buffer without glycerol or CHAPS) by centrifugation at 500 × g for 5 min at 4 °C and then resuspended in IP wash buffer. Following the last wash, the bead suspension was transferred to a new Eppendorf tube to minimize contamination. The bound protein was eluted with 8 M urea buffered in 10 mM Tris, pH 8.0. For proteomics assays, four independent biological replicates were performed for each condition. For protein digestion, 50% of the eluted protein samples was reduced with 1 mM dithiothreitol (DTT) at 25 °C for 30 min, followed by 5 mM iodoacetamide at 25 °C in the dark. Protein was digested with 1:100 (w/w) lysyl endopeptidase (Wako) at 25 °C for 2 h and diluted with 50 mM NH₄HCO₃ to a final concentration of less than 2 M urea. Samples were further digested overnight with 1:50 (w/w) trypsin (Promega) at 25 °C. Resulting peptides were desalted with in-house stage tips and dried under vacuum.

**RNase A treatment**

Cells were lysed in IP buffer with the addition of 5 mM MgCl₂. Following sonication and centrifugation as described above, the lysates were split and treated with RNase A (Sigma) or buffer alone (control) to a final concentration of 50 μg/ml of RNase A. The RNase A treated and control samples were incubated for 30 min at room temperature followed by centrifugation at 10,000 × g for 10 min at 4 °C. The supernatant was added to beads, and the IP was completed as detailed above.

**Blue native gel electrophoresis**

Recombinant N-terminal (residues 1–99) and LC1/BAD U1-70K fragments (residues 231–308) were purified, and their concentrations were determined as described previously (5). The purified GST used as control was kindly gifted by the Dr. Richard Kahn (Department of Biochemistry, Emory University). Prior to analysis, purified rU1-70K fragments were cleared by centrifugation at 20,000 × g for 15 min at 4 °C to remove any insoluble precipitates. Each protein (0.8 μg) was added to blue native gel loading buffer (5% glycerol, 50 mM tris(2-carboxyethyl)phosphine, 0.02% (w/v) G-250 Coomassie, 1× Native Page Running Buffer (catalog no. BN2001, Invitrogen)) and allowed to incubate at room temperature for 30 min. Samples were loaded onto a 3–12% native-PAGE BisTris gel (catalog no. BN2011BX10, Invitrogen) in addition to a native gel molecular weight marker (catalog no. LC0725, ThermoFisher Scientific). Samples were resolved by electrophoresis at 150 V for 1.5 h in anode native-PAGE Running Buffer (catalog no. BN2001, Invit-
BAD RNA–binding proteins aggregate in AD

Immunocytochemistry

HEK293 cells were plated on Matrigel (catalog no. 356234, Corning)-coated coverslips and prepared for transfection using Lipofectamine (ThermoFisher Scientific) according to the manufacturer’s protocol. Immunocytochemistry was performed 48–72 h after transfection essentially as described (68). After the blocking step, slides were dabbed to remove excess liquid and incubated in primary antibody overnight at 4 °C. Primary antibodies included the following: rabbit anti-U1-70K (EM439); mouse anti-Myc; mouse anti-LUC7L3; mouse anti-RBM25; mouse anti-U1-70K. The slides were washed three times with PBS with 0.05% (v/v) saponin and then incubated with secondary antibody (Dylight 549, Alexa 488) for 1 h of shaking at room temperature. Again, the slides were washed three times with PBS with 0.05% saponin. DAPI diluted in PBS was added to each slide and incubated for at least 30 min while rotating at room temperature. Following additional rinses in PBS, cells were mounted in Vectashield (Vector Laboratories, Burlingame, CA) and sealed with nail polish. Images were captured on an Fluoview FV1000 confocal laser-scanning microscope (Olympus).

Liquid chromatography coupled to tandem MS (LC-MS/MS)

Tryptic peptides were analyzed by LC-MS/MS essentially as described (69). Peptides were resuspended in loading buffer (0.1% formic acid, 0.03% TFA, 1% acetonitrile) and separated on a self-packed C18 (1.9 μm Dr. Maisch, Germany) fused silica column (20 cm × 75 μm internal diameter; New Objective, Woburn, MA) by a NanoAcquity UHPLC (Waters) and monitored on a Q-Exactive Plus mass spectrometer (ThermoFisher Scientific, San Jose, CA). Elution was performed over a 140-min gradient at a rate of 300 nl/min with buffer B ranging from 3 to 80% (buffer A: 0.1% formic acid and 5% DMSO in water; buffer B: 0.1% formic acid and 5% DMSO in acetonitrile). The mass spectrometer cycle was programmed to collect one full MS scan (300–1,800 m/z range, 1,000,000 AGC, 100-ms maximum ion time) were collected at a resolution of 70,000 at m/z 200 in profile mode, and the MS/MS spectra (2 m/z isolation width, 28 normalized collision energy, 50,000 AGC target, 50 ms maximum ion time) were acquired at a resolution of 17,500 at m/z 200. Dynamic exclusion was set to exclude previous sequenced precursor ions for 30 s. Precursor ions with +1 and +6 or higher charge states were excluded from sequencing. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008260.

Database search

Raw data files were analyzed using MaxQuant version 1.5.2.8 with Thermo Foundation 2.0 for RAW file reading capability (70). The search engine Andromeda was used to build and search a concatenated target-decoy UniProt Knowledgebase (UniProtKB) containing both Swiss-Prot and TrEMBL human reference protein sequences (90,411 target sequences downloaded April 21, 2015), plus 245 contaminant proteins included as a parameter for Andromeda search within MaxQuant (71). Methionine oxidation (+15.9949 Da), asparagine and glutamine deamidation (+0.9840 Da), and protein N-terminal acetylation (+42.0106 Da) were variable modifications (up to five allowed per peptide); cysteine was assigned a fixed carbamidomethyl modification (+57.0215 Da). Only fully tryptic peptides were considered with up to two miscleavages in the database search. A precursor mass tolerance of ±20 ppm was applied prior to mass accuracy calibration and ±4.5 ppm after internal MaxQuant calibration. Other search settings included a maximum peptide mass of 6,000 Da, a minimum peptide length of six residues, and 0.05-Da tolerance for high resolution MS/MS scans. The false discovery rate for peptide spectral matches, proteins, and site decoy fraction were all set to 1%. The LFQ algorithm in MaxQuant (33, 72) was used for protein quantitation. One limitation of data-dependent LFQ proteomics methods is the inherent missing data (i.e., missing protein identifications or abundance values), especially for low abundance proteins (73). Thus, proteins with 10 or more missing values across the 20 individual samples were not included in the subsequent bioinformatic analysis.

Protein–protein interaction network analysis

The R package WGCNA was used to sort proteins into functional groups by examining relative levels of co-enrichment (74). In WGCNA, correlation coefficients between each protein pair in the dataset are first calculated and transformed continuously with the power adjacency function to generate an adjacency matrix that defines the connection strength between protein pairs. This adjacency matrix is then used to calculate a topological matrix (TO), which measures the interconnectedness or correlation between two proteins and all other proteins in the matrix. All proteins are then hierarchically clustered (e.g., average linkage) using 1-TO as a distance measure, and module assignments are subsequently determined by dynamic tree cutting (74). Threshold power Beta for reduction of false-positive correlations (i.e., the beneficial effect of enforcing scale-free topology) was sampled in increments of 0.5, and as the target scale free topology $R^2$ was approached, 0.1. The power selected was the lowest power at which scale-free topology $R^2$ was ~0.80 or, in the case of not reaching 0.80, the power at which a horizontal asymptote (plateau) was nearly approached before further increasing the power, had a negative effect on scale-free topology $R^2$. Other parameters were selected as optimized previously for protein abundance networks (69). Thus, for the signed network built on protein LFQ abundances obtained from IP-LC-MS/MS, parameters were input into the WGCNA::blockwiseModules() function as follows: Beta 10.9; mergeCutHeight 0.07; pamStage TRUE; pamRespectsDendro TRUE;
reassignThreshold $p < 0.05$; deepSplit 4; minModuleSize 15; corType bicor; and maxBlockSize greater than the total number of proteins. T-Distributed Stochastic Neighbor Embedding (tSNE) analysis was performed as described (75). Proteins with WGCNA intramodular kME $> 0.50$ were retained, and all duplicated values were removed, as well as proteins with any missing values for the 16 non-IgG measurements. Then Rtsne R package Barnes-Hut-Stochastic Neighbor Embedding (SNE) Rtsne function was run on the LFQ expression matrix to reduce dimensionality from 16 to 2. The remaining points or proteins ($n = 375$) were colored according to WGCNA module membership. GO Elite analysis on each module was performed as described previously (69).

**Bioinformatic analysis of BAD proteins in the detergent-insoluble proteome in AD brain**

Quantitative proteomic analysis using isobaric tagging of Sarkosyl-insoluble fractions (frontal cortex) from eight AD and six control cases was previously performed as described (47). Supplementary proteomic data were downloaded, and R was used to generate histograms, Fisher exact $p$ values, box plots, and the clustered heat map with the Nonnegative Matrix Factorization (NMF) package.

**Nuclear and cytoplasmic fractionation**

The fractionation protocol was performed as essentially described in Ref. 76, with slight modifications. Briefly, after transfections with full-length rU1-70K plasmids or respective variants, HEK293 cells were harvested by scraping and washed with PBS, including $1 \times$ Protease Inhibitor Mixture (Sigma). Cells were then spun down at $1000 \times g$ at 4 °C for 5 min and carefully resuspended in hypotonic buffer (10 mM HEPES, pH 7.9, 20 mM KCl, 0.1 mM EDTA, 1 mM DTT, 5% glycerol, 0.5 mM phenylmethylsulfonyl fluoride, and Halt phosphatase inhibitor mixture (1:100, ThermoFisher Scientific)) and incubated on ice for 15 min. The detergent Nonidet P-40 was then added to a 0.1% (v/v) final concentration, and cells were briefly agitated by vortex and left on ice for 5 additional min followed by centrifugation for 10 min at 4 °C at 15,600 $\times g$, affording the supernatant (S1) as the cytoplasmic fraction and the pellet (P1) as the nuclear fraction. To determine whether differences in nuclear and cytoplasmic distributions were significant across conditions, repeated measures ANOVA with post-hoc Tukey was performed in GraphPad Prism.

**Immunoprecipitation of rU1-70K fragments from human brain homogenates**

Post-mortem frontal cortex tissue from pathologically confirmed AD cases were provided by the Emory Alzheimer’s Disease Research Center (ADRC) brain bank (Table S2). Neuro-pathological evaluation of amloid plaque distribution was performed according to the Consortium to Establish a Registry for AD (CERAD) semi-quantitative scoring criteria (77), whereas neurofibrillary tangle pathology was assessed in accordance with the Braak staging system (78). The cortico-basal degeneration and progressive supranuclear palsy cases included in this study also underwent extensive neuropathological characterization required for diagnosis. Tissues were homogenized in Nonidet P-40 lysis buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P-40, 5% glycerol + protease + phosphatase inhibitors) using a bullet blender (69) followed by centrifugation at $10,000 \times g$ for 10 min at 4 °C to clear tissue debris. Immunoprecipitation was performed from 1 mg of brain homogenate from three independent AD cases. Homogenates were first pre-cleared using 30 $\mu$L of protein A-Sepharose–conjugated beads (Invitrogen catalog no. 101041) rotating at 4 °C for 1 h. GST-purified rU1-70K fragments (4 $\mu$g) were added independently to the pre-cleared homogenates. IP was performed using anti-Myc tag (clone 9B11, Cell Signaling) in samples containing the rU1-70K proteins. An IgG control antibody (catalog no. 550339, Pharmin-gen) was used as negative control. Immunocomplexes were captured using the Dynabeads protein G magnetic beads (catalog no. 1003D, Invitrogen), which were washed three times using wash buffer (50 mM Tris- HCl, pH 8, 150 mM NaCl, and 1% Nonidet P-40) followed by 5 min boiling in Laemmli sample buffer to elute bound proteins prior to Western blot analysis. Samples were prepared for MS as described above and analyzed on the Orbitrap Fusion mass spectrometer (ThermoFisher Scientific) (79). A total of 25 Tau (MAPT) were also added to an inclusion list to increase the likelihood of identification and quantification following database searching and quantification by MaxQuant as described above. The intensities of Tau were compared across conditions. ANOVA was used to determine significance with Tau levels across samples in GraphPad Prism.

**Western blotting**

Western blotting was performed according to standard procedures as reported previously (5, 10, 68). Samples in Laemmli sample buffer (8% glycerol, 2% SDS, 50 mM Tris, pH 6.8, 3.25% $\beta$-mercaptoethanol) were resolved by SDS-PAGE before an overnight wet transfer to 0.2-μm nitrocellulose membranes (Bio-Rad) or a semi-dry transfer using the iBlot2 system. Membranes were blocked with casein blocking buffer (catalog no. B6429, Sigma) and probed with primary antibodies (see under “Materials”) at a 1:1,000 dilution overnight at 4 °C. Membranes were incubated with secondary antibodies conjugated to Alexa Fluor 680 (Invitrogen) or IRDye800 (Rockland) fluorophores for 1 h at room temperature. Images were captured using an Odyssey IR Imaging System (Li-Cor Biosciences), and band intensities were quantified using Odyssey imaging software.

**Assessment of protein similarity to LC1/BAD domain of U1-70K**

The U1-70K LC1/BAD domain protein similarity list was created using the Uniprot pBLAST feature (http://www.uniprot.org/blast/) using the following parameters: Target Database Human containing 160,363 entries (updated March 2015), E-threshold: 10, Matrix: Auto, Filtering: None, Gapped: Yes, Hits: 1000. The input blast entry was the LC1/BAD domain of U1-70K (residues 231–308). The resulting list of proteins was subsequently filtered to remove unreviewed entries, producing 255 proteins with E-values less than 0.005 and similarity to the LC1/BAD domain of higher than 20% (Table S3). The “biotin-
isoxazole" list originates from a previous study (43) using the biotinylated isoxazole compound to precipitate proteins from HEK293 nuclear extracts. The "prion-like" list of RNA-binding proteins originates from a previous study using in silico methods to identify Gln/Asn-rich prion-like proteins (42). Finally, protein alignment was done using Clustal Omega multiple sequence alignment (https://www.ebi.ac.uk/Tools/msa/clustalo/) (80).


**Acknowledgments**—We acknowledge Drs. Anita Corbett (Emory Department of Biology) and Daniel Reines (Emory Department of Biochemistry) for providing helpful comments to the manuscript. We also thank Dr. Measho Abreha for technical advice on brain homogenization and immuno precipitation conditions. Finally, we acknowledge David S. Sanders who on Twitter created the Basic-Acidic Dipetidyl (BAD) domain acronym after reading our paper on bioRxiv (https://www.biorxiv.org/). Research reported in this publication was supported in part by the Emory Neuroscience NINDS Core Facilities (Grant P30NS055077) from the National Institutes of Health, NINDS.

**References**

25. Liu-Yesucevitz, L., Gan, S. Y., Abata, A., Boon, J. Y., Yeid, W., Xu, Y. F., Kobrin, K., Murphy, G. J., Petrucelli, L., and Wolozin, B. (2014) ALS-linked...


