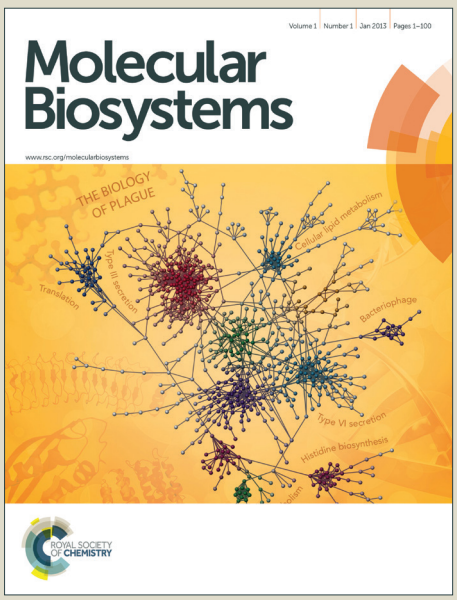
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Repeats are one of the main characteristics of RNA-binding proteins with prion-like domains

Oxana V. Galzitskaya

[1]*Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region;*

* Corresponding author: O.V. Galzitskaya, *e-mail:* ogalzit@vega.protres.ru

Abstract

It is not surprising that at a large number of diseases related to amyloid fibril depositions are formed in various organs. Therefore, it is necessary to understand the transformation of native proteins into amyloid fibrils in order to clarify which key elements of this process determine the pathway of protein misfolding. Significant attention has been directed recently to investigating the mechanism of the formation of cross-β structures that have the properties of liquids but can also exist in gel-like forms, thus facilitating the retention of both RNAs and RNA-binding proteins. Proteins that form stress granules are believed to do this rapidly, and they are expected to contain a prion-like domain that can facilitate this process. By analyzing the known yeast prion proteins and 29 RNA-binding proteins with prion-like domains, we demonstrate here that the existence of repeats is one of the general characteristics of prion-like domains. The presence of repeats should help to determine the border of prion domains as in the case of Rnq1: five found repeats shift the border of the prion domain from the 153-rd to at least the 133-th residue. One can suggest that such repeats assist in the rapid initiation of the process of assembly and formation of cross-β structures and such domains most likely should be disordered. These repeats should contain aromatic amino acid residues for the formation of hydrogel because its amino acid context modulates the strength of interaction. The key factors determined here can be used to control the process of aggregation to prevent the development of pathologies and diseases caused by prion-like domains.

Keywords: proteinopathies, aggregation, repeats, stress granules, cross-β structure

Introduction

Prions are infectious proteins that can assume various conformations, including amyloid fibrils that can serve as a matrix and "infect" other proteins both within and between the cells and between organisms [1,2]. However, it was observed that yeast prions facilitate their adaptation to diverse conditions of the environment [3]. In mammals, they can both cause diseases, sometimes even leading to death, and perform useful functions, such as activating cell-based immunity (protein MAVS), providing for long-term memory (protein CPEB), and forming stress-granule assemblies in conditions unfavorable for the cell [4].

For specific purposes living organisms widely utilize the property of protein molecules to form various amyloid structures. Normally, certain organisms form amyloid fibrils to perform various functions. One of the best studied examples of such functional amyloids is curlin, which is used by *E.coli* to colonize inert surfaces and is a mediator upon binding with proteins of other organisms. Another example is the bacterium *Streptomyces coelicolor*, which, due to the formation of amyloid fibrils with chaplin proteins, forms hyphae that are used for spore spreading [5].

In the examples above, the process of amyloid nucleation that initiates the aggregate growth depends on the ambient conditions and is controlled by a definite cascade of reactions. The controlled formation of functional amyloid aggregates occurs in mammals as well. For example, melanosomes are organelles that differentiate in melanocytes and are responsible for melanin biogenesis in skin cells, and they contain fibril structures on which melanin granules are formed. Such fibrils have much in common with amyloids: like amyloids, they are formed from a proteolytically cleaved domain of a membrane protein, specifically Pmel17 [6].

Long-term memory is also provided by the principle of fibril formation, in which the protein CPEB (an RNA-binding protein capable of controlling the local translation of mRNAs in dendrites) plays an essential role. This protein can stimulate mRNA polyadenylation, and its aggregation activates translation of the "silent" mRNA accumulated in synaptic end-feet [7]. The N-domain of CPEB is rich in asparagine and glutamine residues, a feature which is specific to prion-like domains [8].

The protein MAVS, located on the surface of mitochondria membranes, activates the innate antiviral cell immunity in a similar manner. When aggregated, it can interact with the cytoplasmic receptors that recognize the patterns specific to most pathogens, which triggers a

cascade of reactions leading to the synthesis of β-interferon [9,10]. This protein can also aggregate by the prion mechanism.

The examples above demonstrate that even in highly organized organisms, the formation of amyloids located in a strictly defined place and rigidly controlled can be physiologically beneficial for performing specific, specialized biological functions.

The aggregation of proteins can occur as a result of cell stress. To prevent unfavorable external effects (heat shock, oxidative stress, UV irradiation, viral infections, and many other factors) that can damage the cell, the eukaryotic cell has developed a mechanism for assembling the nontranslatable mRNA and RNA-binding proteins that accompany their transition into special ribonucleoprotein (RNP) complexes: stress granules (SGs) and RNA-processing organelles (P-bodies) [11–15]. Whereas SGs function as transient "repositories" of mRNA and protect it from proteins, P-bodies perform the selective degradation of mRNA during stress and during the recovery period, constantly exchanging material with the SGs [16].

SGs are highly dynamic structures and include RNA helicases, kinases, various signal molecules, and ribosomal subunits, in addition to mRNA and RNA-binding proteins. In post-stress conditions, SGs rapidly dissociate, and the cell resumes its functioning [17,18].

Many RNA-binding proteins required for the formation of SGs contain prion-like domains that, because of the protein-protein interaction, develop dynamic cross-β structures capable of the rapid aggregation and dissociation that are so important for the correct functioning of SGs [4]. In addition, the aggregated forms of prion-like domains in mammals become resistant to the action of proteases [19].

When an RNA-binding protein has a propensity for various conformational transitions, including the formation of amyloid fibrils, these prion-like domains enable folding and cause the protein to transform from an unfolded three-dimensional structure into intermediate states [20–22]. As a rule, such proteins are in a dynamic equilibrium between the two forms: unfolded soluble monomers and molten oligomers. The latter can be involved in multiple conformational states (Figure 1).

In keeping with one scenario, they can assemble into structured amyloidogenic oligomers that are then transformed into pathological aggregates of a non-amyloid type or into amyloid fibrils. The amyloid fibrils can, in turn, play the role of a "matrix" for incorrect folding, as do prions. According to another scenario, protein molecules can form amorphous aggregates consisting of both soluble monomers and molten oligomers, arranged as dynamic cross-β structures that have the properties of liquids but can develop into gel-like forms [23–25]. The two

transition states (liquid – hydrogel) are vital for the formation of various nonmembrane structures, including transport RNA granules, Cajal bodies, gemini of Cajal bodies, SGs, etc. [26,27]. X-ray diffraction and EM studies have revealed that the hydrogel is composed of uniformly polymerized amyloid-like fibers. Unlike pathogenic fibers, these polymers are dynamic and accommodate heterotypic polymerization [22]. X-ray diffraction patterns give strong evidence for the presence of amyloid-like polymers as the structural basis of the hydrogel architecture [22]. Numerous studies have demonstrated that prion-like domains are capable of helping to localize RNA-binding proteins to P-bodies in yeast[28,29] and mammalian SGs [4]. It is likely that under the action of definite factors, these nonmembrane structures can transform into pathological protein aggregates and substitute the amorphous structure by an amyloid one [15]. It is likely that this mechanism is used by some SGs to become precursors of pathological protein inclusions under (neuro) degenerative proteinopathies.

When the stress conditions end, SGs dissociate rapidly, and the "released" mRNA recommences functioning. But, if the residence time of such proteins in SGs increases, or if their concentration in SGs increases, the dissociation of SGs may be impeded by the effects of the stable protein-protein interactions of the prion-like domains. This may lead to the appearance of an "aggregation initiation center" that gives rise to pathological protein aggregation [30]. This process can be described as a type of "stabilization" of SGs with subsequent evolution in a pathological manner. The data available make it possible to propose the existence of a fine balance between the physiological and pathological aggregation of proteins because both types of aggregation are controlled by prion-like domains. At present, an important task is to establish the factors that tip this balance and trigger a cascade of uncontrolled aggregation of RNA-binding proteins.

In this paper, we analyze RNA-binding proteins with prion-like domains to elucidate their characteristics. It is demonstrated that the prion-like domains in these proteins are intrinsically disordered regions and have several repeats (or tandem repeats). Based on the known properties of prion-like domains, such as the formation of stress granules (as an example of the reversible aggregation of proteins), one can hypothesize that a large number of repeats is necessary to accelerate the process of reversible cross-β structure/amyloid formation.

Results and Discussion

Repeats in the disordered domains of known yeast prion proteins

Let us consider the amino acid sequences of the known yeast prion proteins. It should be noted that PrP and Sup35 have seven and six imperfect repeats in their N-terminal domains (PHGGGWGQ and PQGGYQQYN, respectively; see Table 1). One can suggest that the aggregation process should occur faster for a larger number of repeats. In reality, it has been shown for Sup35 that two additional repeats of R2 lead to a decreased lag time for the NM portion, whereas the deletion of the R2-R5 repeats results in an increased lag time [31]. As has been suggested, the repeats should facilitate the correct alignment of the intermolecular contacts between molecules. Moreover, faster aggregation kinetics occurs for a longer polyQ [32]. The efficiency of fibril formation at seeding aggregation can strongly decrease upon an increase in dissimilarity in the primary structure [33–35]. For immunoglobulin domains with a different primary structure, it was demonstrated that coaggregation of various types of domains does not occur when the identity of the protein primary structure is less than 30-40% [35]. It was observed that the protofibril-to-mature-fibril transition of a peptide (109-122) from a Syrian hamster prion protein proceeds through the alignment of originally unaligned β-regions to form a potential fibril [36]. This alignment includes the isolation of β-regions and their subsequent inclusion into the potential fibril; however, internal rearrangement of β-regions is also possible and has been observed under certain conditions [37].

The prion-like domain of the protein URE2 is the N-terminal domain, which includes amino acids 1-94 and is rich in asparagine and glutamine residues [38] (see Figure 2 and Table 1). The C-terminal catalytic domain (95-354 a.a.) is responsible for catabolic repression [39]. The transition to the prion state of the N-terminal domain inactivates protein URE2. The C-terminal domain has regions that affect the capacity of the N-terminal domain to become a prion domain [40].

The prion domain of Rnq1p (153-405) contains many glutamine residues and includes five repeats not reported previously in the literature and in the Uniprot database, such as R1 (ASGLAALASQF), R2 (FTALASLASSF), R3 (FGALASMASSF), R4 (FSSLASMAQSY), and R5 (FSALASMASSY) (see Figure 2 and Table 1). It should be noted that these repeats have been found by eyes, because the program T-REKS does not find them. These hydrophobic repeats are flanked by other repeats enriched by such amino acid as Q/N/G. It should be noted that the borders for prion domains are not precisely defined by functional criteria [41]. Based on the found repeats the border should be corresponding to at least the 133-rd but not to the 153-th residue as considered before. Thereby, the presence of repeats helps to determine the border of prion domains.

Four imperfect repeats in the prion-domain of RNQ1 are noted by Vitrenko et al.[42] including F**LAS*A*S, which agrees with R2-R5 repeats in 6 out of the 9 specified positions F**LASM_LASSF_Y (we allowed one mismatching out of five, in the case of two we used a subscript to denote the amino acid). The authors investigated how repeats can influence the aggregation and prion propagation. It has been found that the three hydrophobic repeats retained in Rnq-Δ4:GFP are sufficient to position and polymerize the protein on wild-type Rnq1 aggregates, but to maintain the prion state in the absence of wild-type Rnq1 all four (in our case five) repeats (Rnq1-Δ6) are required. The authors speculate that "some of the repeats, when exposed on the surface of [PIN+] aggregates could specifically bind to, and stabilize, prionogenic Sup35 oligomers" in addition to the other possible function of repeats that are targets for chaperones shown to interact with Rnq1 [43–45]. Moreover, the authors hypothesized that "the number, high hydrophobicity index and spacing of these motifs might be crucial for aggregation and prion propagation"[42]. An important result of this paper is that there are some other features of the amino acid composition contributing to the higher prionogenicity of longer Rnq1 fragments in addition to the Q/N content.

Search for repeats in RNA-binding proteins with prion-like domains

Based on the examples above, it can be hypothesized that prion-like domains should contain repeats of simple motifs for the faster formation of a dynamic cross-β structure. To verify this assertion, we considered the database published in [46] that consists of 29 RNA-binding proteins with prion-like domains.

First, we predicted disordered regions using the IsUnstruct program, which operates as well as the PONDR-FIT meta-server, for 29 RNA-binding proteins of the human proteome, which included the predicted prion-like domains [47]. The comparison in this work was between the RNA-binding protein FUS (further FUS) and the DAZ1 protein (see Figure 3). All 29 proteins considered here have disordered regions (see Supplementary Materials and Table 2).

Regarding the prediction of prion-like domains, the first, second, and third best among the 29 candidates of RNA-binding proteins with a prion-like domain are FUS, TAF15 (TATA-binding protein-associated factor 2N), and EWS (RNA-binding protein EWS) [48]. For these three proteins, our program predicts the existence of disordered domains, usually at the N- or C-terminus of the polypeptide chain, which correspond to the prion-like domain (see Figure 4, Table 2, and Supplementary Materials).

The prion-like domains of FUS, TAF15, and EWS are critical for the aggregation of proteins associated with human neurodegenerative diseases [49]. Proteins of this family are

involved not only in neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) [50], Huntington's disease, spinocerebral ataxy, and dentatorubral pallidoluysian atrophy [51], but also in the formation of human mixsoid liposarcoma [52–54].

The N-terminus of the FUS protein is enriched with amino acid residues such as Gln/Gly/Ser/Tyr (see Figure 4). Gapped multiple-sequence alignment was performed to represent a profile of the selected motifs for FUS [55] (QPGQGYS-QQSS, see Supplementary Materials and Table 2). The C-terminus of the FUS protein includes four imperfect repeats, such as DDRRGGRGGY.

Although RGG repeated motifs have been considered for the TAF15 protein, more thorough consideration reveals a more attractive repeat at the C-terminus (407-575): 21-23 approximate tandem repeats of DR[S,G]GGYGG. The N-terminus (1-208) is enriched with amino acid residues such as Gln/Gly/Ser/Tyr (see Table 2 and Supplementary Materials) and repeats such as SYGQSGGEQQ.

The prion-like domain of the EWS protein includes 31 approximate tandem repeats (8-285), such as SYSQAPS at the N-terminus and six imperfect DRGRGGPGG repeats at the C-terminus (see Figure 4, Table 2, and Supplementary Materials).

In addition to prion-like domains rich in asparagine, glutamine, and tyrosine residues, many RNA-binding proteins have regions with large amounts of glycine. It has been suggested that these regions facilitate the formation of RNP-complexes and are involved in the splicing of pre-mRNA, as well as having an effect on the posttranslational modifications of proteins [56–58]. Because having a large amount of glycine helps maintain the protein in an unfolded state, the neighboring amino acid residues determine part of its function. For example, the FUS structure has domains rich in glycine and arginine (RGG) that are responsible for the protein-protein and protein-RNA interactions. These processes are controlled by arginine methylation [59]. It has also been observed that the FUS structure contains a region rich in glycine and serine. The function of this region has not been determined yet, but it is proposed that phosphorylation of its serine residues affects the mutual interaction of the prion-like domains [21,58].

Additionally, other RNA-binding proteins with long homorepeats of glycines involved in neurodegenerative diseases have been observed: for example, the proteins HNRNPA0 (heterogeneous nuclear ribonucleoprotein A0) and HNRNPA1 (heterogeneous nuclear ribonucleoprotein A1, see №5 in Table 2). Aggregation of these proteins is highly toxic. Thus, pathogenic aggregation is predicted for the protein HNRNPA3 (heterogeneous nuclear ribonucleoprotein A3), which also has long glycine homorepeats (GNFGGRG) at the C-terminus

but has not been tested yet for toxicity (see Supplementary Materials, №6 in Table 2). It should be mentioned that these glycine-rich peptides belong to prion-like domains and that hydrogel formation was observed for the prion-like domain of HNRNPA2 [22].

RNA-binding domains exist that have a clearly pronounced structural organization, as predicted by the IsUnstruct[47] and PONDR-FIT[60] programs (see Figure 3). Several examples are the proteins that are involved in the spermatogenesis of DAZ-1,-2,-3 and are capable of forming toxic aggregates [50]. These proteins are characterized by the recurrence of a large motif (RRM domain) and a smaller motif (AYPHSPGQVITGCQLLVYNYQ – DAZ-like repeats that are essential and mediate the interaction with DAZAP1 and DAZAP2), but these two types of motifs occur a different number of times in each of the proteins. Although the capacity of the protein DAZ-4 to form toxic aggregates has not yet been demonstrated experimentally, the presence of the same motifs suggests the extremely toxic aggregation of this protein because the large domain recurs two times and the smaller one nine times (see Supplementary Materials). It is known that proteins from the DAZ family can retain mRNA in stress granules in sperm cells and thus protect them from the degradation of transcripts during stress actions [61].

A careful consideration of the repeats in the prion-like domain of RNA-binding proteins shows that they include aromatic amino acid residues (see Table 2). This was stated when analyzing yeast prion proteins. In the case of the three FET proteins, the repeats found by us at the N- and C-termini also contain aromatic residues. Studies were performed in which it was shown that namely phenylalanine is critical for the formation of hydrogel in the case of protein Nsp1p. Like many other nuclear pore proteins, Nsp1p contains multiple phenylalanine-glycine (FG) repeats. In this protein mutation of all phenylalanine residues to serine blocked the hydrogel formation, yet mutation of the same residues to tyrosine did not. The authors suggested that phenylalanine is the most important amino acid for the formation of hydrogel [62]. However later it became clear that all the properties of hydrogel cannot be explained by only phenylalanine [63]. Solid-state NMR spectroscopy allowed finding an additional type of intragel interaction. They are intermolecular hydrogen-bond interactions within β-sheets between polar amino acid residues.

The N-terminal domain of FUS contains 27 different versions of GYG, GYS, SYG and SYS tripeptides (which can be designated as [G/S]Y[G/S] repeats). Four mutants with different numbers of tyrosine residues substituted for serine were prepared to demonstrate that namely tyrosine residues are responsible for the formation of hydrogel. The number of substitutions was 5, 9, 15 and all 27. Neither of the mutants could form hydrogel, but with a different strength

these mutants could bind to hydrogel. Mutants with 5 and 9 substituted residues could yet bind to hydrogel, but the other mutants could not [22].

Thus, one can hypothesize that the presence of repeats/tandem repeats or homorepeats in prion-like IDRs should result in an acceleration of the formation of a dynamic cross-β structure.

Materials and Methods

A database consisting of 29 RNA-binding proteins containing prion-like domains was considered[64] (see Supplementary Materials,

http://bioinfo.protres.ru/papers/Supplementary_Table.pdf). The prediction of prion-like domains was made using the algorithm developed by Alberti et al., [8] based on the choice of protein regions of 60 amino acid residues that are similar in their amino acid composition to the prion domains of yeast proteins, such as Sup35, URE2, and Rnq1p [65]. Usually, these regions are rich in hydrophilic amino acid residues, such as glutamine, asparagine, and tyrosine.

Search for disordered residues

Disordered residues were predicted using the IsUnstruct program, which is based on the Ising model [66]. The parameters of the program were determined and optimized on the basis of protein-structure statistics. The tests demonstrated that the program yields reliable predictions. The program is available at http://bioinfo.protres.ru/IsUnstruct[47]. The PONDR-FIT method was used to check the reliability of the predictions; a meta-server yields a consensus prediction for ten programs [60].

Search for tandem repeats

The program T-reks [55] and the UniProt database were used to search for tandem repeats.

Conclusion

The elucidation of the molecular mechanisms of prion diseases and neurodegenerative diseases, and specifically the identification of the regions in the protein sequences responsible for their development, is one of the most important problems in the area of life sciences. Indeed, the ability to control reversible protein aggregation could enable biomarker discovery and targeted drug development. In this work, RNA-binding proteins with prion-like domains were

analyzed by bioinformatics tools to understand the possible role of the repeats in the aggregation process. After considering RNA-binding proteins with prion-like domains, one can hypothesize that (tandem) repeats assist in the rapid initiation of the process of the assembly and formation of cross-β structures for prion-like domains. The presence of repeats should help to determine the border of prion-like domains and should be taken into account in the programs for searching prion-like domains. Moreover, aromatic amino acids are critical for the formation of hydrogels modulating the strength of interaction.

Acknowledgments

We are grateful to D. Reifsnyder Hickey, T.B. Kuvshinkina, E.I. Leonova, and I.V. Sokolovsky for assistance in the preparation of the paper. This study was supported by the Russian Science Foundation (grant number 14-14-00536).

References

1 A. Aguzzi and L. Rajendran, *Neuron*, 2009, **64**, 783–790.
2 S. B. Prusiner, M. R. Scott, S. J. DeArmond and F. E. Cohen, *Cell*, 1998, **93**, 337–348.
3 R. Halfmann, S. Alberti and S. Lindquist, *Trends Cell Biol.*, 2010, **20**, 125–133.
4 N. Gilks, N. Kedersha, M. Ayodele, L. Shen, G. Stoecklin, L. M. Dember and P. Anderson, *Mol. Biol. Cell*, 2004, **15**, 5383–5398.
5 D. Claessen, R. Rink, W. de Jong, J. Siebring, P. de Vreugd, F. G. H. Boersma, L. Dijkhuizen and H. A. B. Wosten, *Genes Dev.*, 2003, **17**, 1714–1726.
6 J. F. Berson, A. C. Theos, D. C. Harper, D. Tenza, G. Raposo and M. S. Marks, *J. Cell Biol.*, 2003, **161**, 521–533.
7 P. Tompa and P. Friedrich, *Neuroscience*, 1998, **86**, 1037–1043.
8 S. Alberti, R. Halfmann, O. King, A. Kapila and S. Lindquist, *Cell*, 2009, **137**, 146–158.
9 K. Onoguchi, K. Onomoto, S. Takamatsu, M. Jogi, A. Takemura, S. Morimoto, I. Julkunen, H. Namiki, M. Yoneyama and T. Fujita, *PLoS Pathog.*, 2010, **6**, e1001012.
10 J. L. Jacobs and C. B. Coyne, *J. Mol. Biol.*, 2013, **425**, 5009–5019.
11 P. Anderson and N. Kedersha, *J. Cell Biol.*, 2006, **172**, 803–808.
12 P. Anderson and N. Kedersha, *Trends Biochem. Sci.*, 2008, **33**, 141–150.
13 N. Kedersha and P. Anderson, *Methods Enzymol.*, 2007, **431**, 61–81.
14 N. Kedersha, G. Stoecklin, M. Ayodele, P. Yacono, J. Lykke-Andersen, M. J. Fritzler, D. Scheuner, R. J. Kaufman, D. E. Golan and P. Anderson, *J. Cell Biol.*, 2005, **169**, 871–884.
15 Y. R. Li, O. D. King, J. Shorter and A. D. Gitler, *J. Cell Biol.*, 2013, **201**, 361–372.
16 R. Parker and U. Sheth, *Mol. Cell*, 2007, **25**, 635–646.
17 N. Kedersha, M. R. Cho, W. Li, P. W. Yacono, S. Chen, N. Gilks, D. E. Golan and P. Anderson, *J. Cell Biol.*, 2000, **151**, 1257–1268.
18 N. L. Kedersha, M. Gupta, W. Li, I. Miller and P. Anderson, *J. Cell Biol.*, 1999, **147**, 1431–1442.
19 D. Riesner, *Br. Med. Bull.*, 2003, **66**, 21–33.

20 R. Halfmann, S. Alberti, R. Krishnan, N. Lyle, C. W. O'Donnell, O. D. King, B. Berger, R. V. Pappu and S. Lindquist, *Mol. Cell*, 2011, **43**, 72–84.
21 T. W. Han, M. Kato, S. Xie, L. C. Wu, H. Mirzaei, J. Pei, M. Chen, Y. Xie, J. Allen, G. Xiao and S. L. McKnight, *Cell*, 2012, **149**, 768–779.
22 M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko and S. L. McKnight, *Cell*, 2012, **149**, 753–767.
23 Z. Sun, Z. Diaz, X. Fang, M. P. Hart, A. Chesi, J. Shorter and A. D. Gitler, *PLoS Biol.*, 2011, **9**, e1000614.
24 R. B. Wickner, K. L. Taylor, H. K. Edskes and M. L. Maddelein, *Curr. Biol.*, 2000, **10**, 335–337.
25 L. Li and S. Lindquist, *Science*, 2000, **287**, 661–664.
26 T. E. Kaiser, R. V. Intine and M. Dundr, *Science*, 2008, **322**, 1713–1717.
27 S. P. Shevtsov and M. Dundr, *Nat. Cell Biol.*, 2011, **13**, 167–173.
28 C. J. Decker, D. Teixeira and R. Parker, *J. Cell Biol.*, 2007, **179**, 437–449.
29 M. A. M. Reijns, R. D. Alexander, M. P. Spiller and J. D. Beggs, *J. Cell Sci.*, 2008, **121**, 2463–2472.
30 D. A. Bosco, N. Lemay, H. K. Ko, H. Zhou, C. Burke, T. J. Kwiatkowski, P. Sapp, D. McKenna-Yasek, R. H. Brown and L. J. Hayward, *Hum. Mol. Genet.*, 2010, **19**, 4160–4175.
31 J. Shorter and S. Lindquist, *Nat. Rev. Genet.*, 2005, **6**, 435–450.
32 N. C. Kar, A. Cracchiolo 3rd, J. Mirra and C. M. Pearson, *Am. J. Clin. Pathol.*, 1976, **65**, 220–228.
33 M. Tanaka, P. Chien, K. Yonekura and J. S. Weissman, *Cell*, 2005, **121**, 49–62.
34 M. R. H. Krebs, L. A. Morozova-Roche, K. Daniel, C. V. Robinson and C. M. Dobson, *Protein Sci. Publ. Protein Soc.*, 2004, **13**, 1933–1938.
35 C. F. Wright, S. A. Teichmann, J. Clarke and C. M. Dobson, *Nature*, 2005, **438**, 878–881.
36 S. A. Petty, T. Adalsteinsson and S. M. Decatur, *Biochemistry (Mosc.)*, 2005, **44**, 4720–4726.
37 S. A. Petty and S. M. Decatur, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 14272–14277.
38 A. A. Komar, T. Lesnik, C. Cullin, W. C. Merrick, H. Trachsel and M. Altmann, *EMBO J.*, 2003, **22**, 1199–1209.
39 P. W. Coschigano and B. Magasanik, *Mol. Cell. Biol.*, 1991, **11**, 822–832.
40 M. L. Maddelein and R. B. Wickner, *Mol. Cell. Biol.*, 1999, **19**, 4516–4524.
41 H. K. Edskes and R. B. Wickner, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99 Suppl 4**, 16384–16391.
42 Y. A. Vitrenko, M. E. Pavon, S. I. Stone and S. W. Liebman, *Curr. Genet.*, 2007, **51**, 309–319.
43 M. E. Bradley, H. K. Edskes, J. Y. Hong, R. B. Wickner and S. W. Liebman, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99 Suppl 4**, 16392–16399.
44 N. Lopez, R. Aron and E. A. Craig, *Mol. Biol. Cell*, 2003, **14**, 1172–1181.
45 R. Aron, N. Lopez, W. Walter, E. A. Craig and J. Johnson, *Genetics*, 2005, **169**, 1873–1882.
46 O. D. King, A. D. Gitler and J. Shorter, *Brain Res.*, 2012, **1462**, 61–80.
47 M. Y. Lobanov, I. V. Sokolovskiy and O. V. Galzitskaya, *J. Biomol. Struct. Dyn.*, 2013, **31**, 1034–1043.
48 A. Y. Tan and J. L. Manley, *J. Mol. Cell Biol.*, 2009, **1**, 82–92.
49 A. Aguzzi and L. Rajendran, *Neuron*, 2009, **64**, 783–790.
50 J. Couthouis, M. P. Hart, J. Shorter, M. DeJesus-Hernandez, R. Erion, R. Oristano, A. X. Liu, D. Ramos, N. Jethava, D. Hosangadi, J. Epstein, A. Chiang, Z. Diaz, T. Nakaya, F. Ibrahim, H.-J. Kim, J. A. Solski, K. L. Williams, J. Mojsilovic-Petrovic, C. Ingre, K. Boylan, N. R. Graff-Radford, D. W. Dickson, D. Clay-Falcone, L. Elman, L. McCluskey, R. Greene, R. G. Kalb, V. M.-Y. Lee, J. Q. Trojanowski, A. Ludolph, W. Robberecht, P. M. Andersen, G. A. Nicholson, I. P. Blair, O. D. King, N. M. Bonini, V. Van Deerlin, R. Rademakers, Z. Mourelatos and A. D. Gitler, *Proc. Natl. Acad. Sci.*, 2011, **108**, 20881–20890.

51 H. Doi, S. Koyano, Y. Suzuki, N. Nukina and Y. Kuroiwa, *Neurosci. Res.*, 2010, **66**, 131–133.
52 A. Crozat, P. Aman, N. Mandahl and D. Ron, *Nature*, 1993, **363**, 640–644.
53 T. J. Kwiatkowski, D. A. Bosco, A. L. Leclerc, E. Tamrazian, C. R. Vanderburg, C. Russ, A. Davis, J. Gilchrist, E. J. Kasarskis, T. Munsat, P. Valdmanis, G. A. Rouleau, B. A. Hosler, P. Cortelli, P. J. de Jong, Y. Yoshinaga, J. L. Haines, M. A. Pericak-Vance, J. Yan, N. Ticozzi, T. Siddique, D. McKenna-Yasek, P. C. Sapp, H. R. Horvitz, J. E. Landers and R. H. Brown, *Science*, 2009, **323**, 1205–1208.
54 C. Vance, B. Rogelj, T. Hortobágyi, K. J. De Vos, A. L. Nishimura, J. Sreedharan, X. Hu, B. Smith, D. Ruddy, P. Wright, J. Ganesalingam, K. L. Williams, V. Tripathi, S. Al-Saraj, A. Al-Chalabi, P. N. Leigh, I. P. Blair, G. Nicholson, J. de Belleroche, J.-M. Gallo, C. C. Miller and C. E. Shaw, *Science*, 2009, **323**, 1208–1211.
55 J. Jorda and A. V. Kajava, *Bioinforma. Oxf. Engl.*, 2009, **25**, 2632–2638.
56 Y. Kino, C. Washizu, E. Aquilanti, M. Okuno, M. Kurosawa, M. Yamada, H. Doi and N. Nukina, *Nucleic Acids Res.*, 2011, **39**, 2781–2798.
57 J.-F. Fisette, D. R. Montagna, M.-R. Mihailescu and M. S. Wolfe, *J. Neurochem.*, 2012, **121**, 763–773.
58 B. Rogelj, L. E. Easton, G. K. Bogu, L. W. Stanton, G. Rot, T. Curk, B. Zupan, Y. Sugimoto, M. Modic, N. Haberman, J. Tollervey, R. Fujii, T. Takumi, C. E. Shaw and J. Ule, *Sci. Rep.*, 2012, **2**, 603.
59 M. T. Bedford, *J. Cell Sci.*, 2007, **120**, 4243–4246.
60 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochim. Biophys. Acta*, 2010, **1804**, 996–1010.
61 M. J. W. VanGompel and E. Y. Xu, *Spermatogenesis*, 2011, **1**, 36–46.
62 S. Frey, R. P. Richter and D. Gorlich, *Science*, 2006, **314**, 815–817.
63 C. Ader, S. Frey, W. Maas, H. B. Schmidt, D. Gorlich and M. Baldus, *Proc. Natl. Acad. Sci.*, 2010, **107**, 6281–6285.
64 O. D. King, A. D. Gitler and J. Shorter, *Brain Res.*, 2012, **1462**, 61–80.
65 J. Shorter and S. Lindquist, *Science*, 2004, **304**, 1793–1797.
66 M. Y. Lobanov and O. V. Galzitskaya, *Phys. Biol.*, 2011, **8**, 035004.

Legends to Figures

Figure 1. Schematic representation of prions and prion-like domains that can lead to various conformational states and scenarios. Modified from[15].

Figure 2. Predictions of the residue status (ordered or disordered) for proteins with prion-like activity using the IsUnstruct program[47]: A)Sup35; B) URE2; and C) Rnq1p. The continuous line at 0.5 of the Y-axis is the threshold line for residues to be disordered. Prion-like domains are indicated by the light-green color.

Figure 3. Comparison of predictions using the two different programs (black circles correspond to IsUnstruct and white circles to PONDR-FIT) for the A) FUS and B) DAZ1 proteins.

Figure 4. Predictions of the residue status (ordered or disordered) with the IsUnstruct program [47] for the FET family: A) RNA-binding protein FUS; B) TATA-binding protein-associated factor 2N; and C) RNA-binding protein EWS. The continuous line at 0.5 of the Y-axis is the threshold line for residues to be disordered. Prion-like domains are indicated by the light-green color. SYGQ corresponds to a region rich in serine, tyrosine, glycine, and glutamine. RRM represents an RNA recognition motif. RGG corresponds to a region rich in arginine and glycine. Zn represents a zinc finger motif.

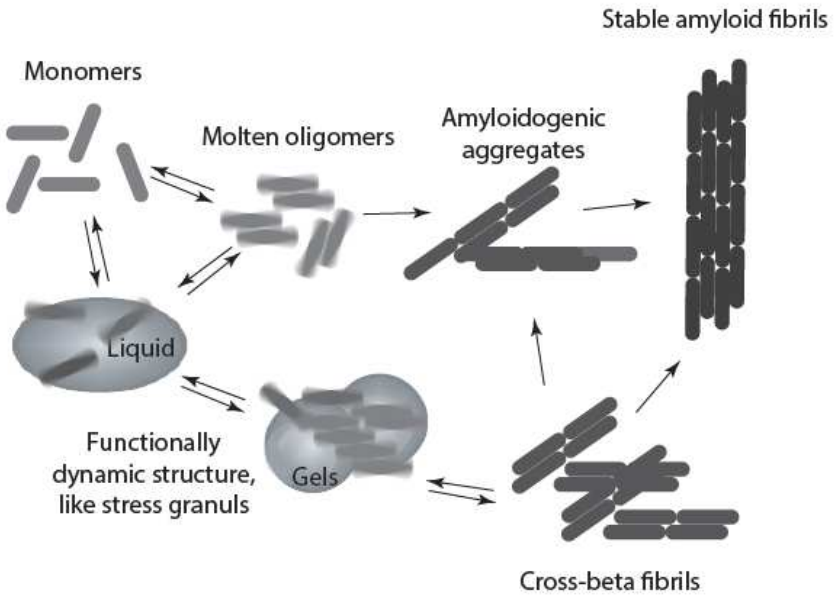


Figure 1. Schematic representation of prions and prion-like domains that can lead to various conformational states and scenarios. Modified from[15].

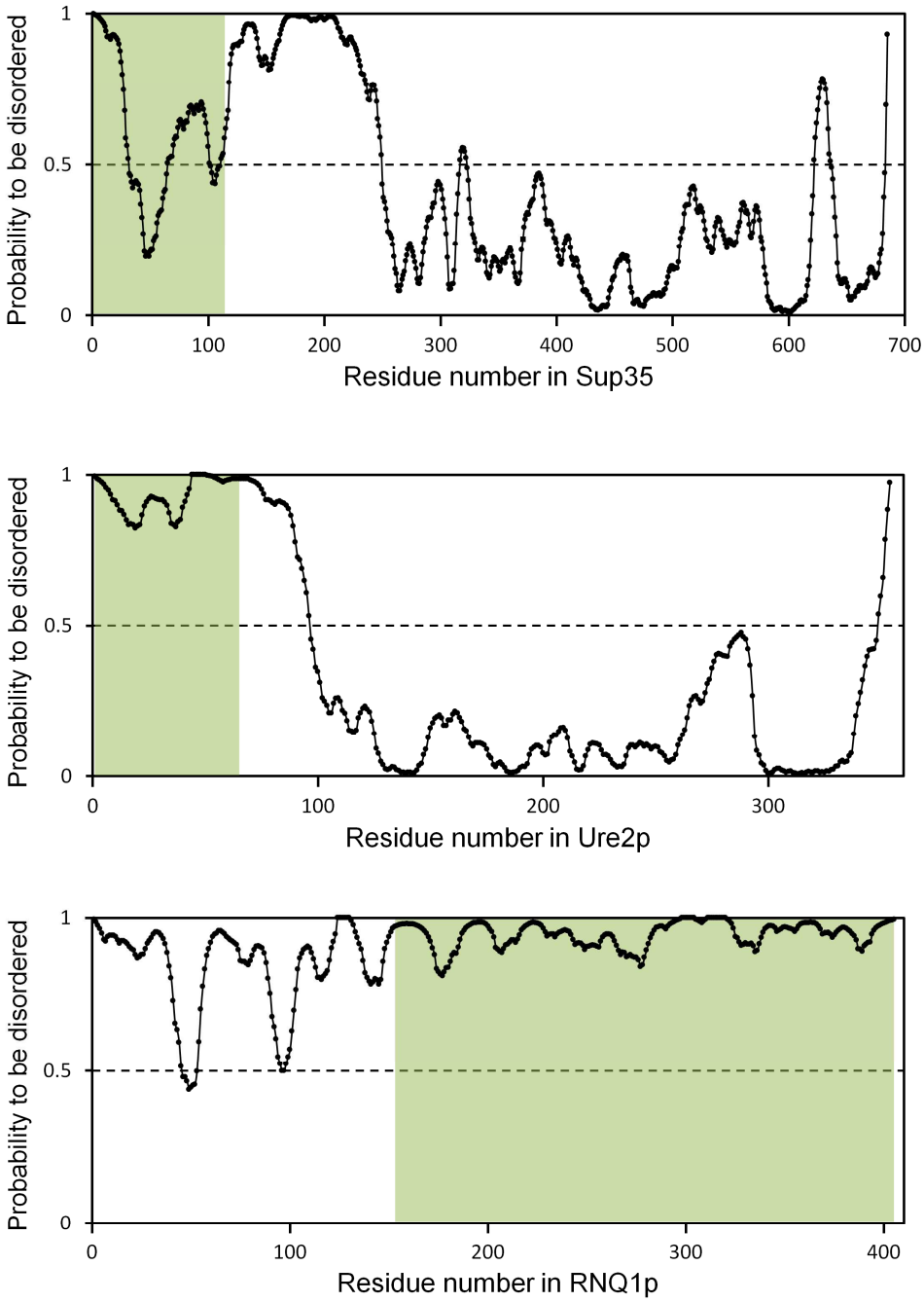


Figure 2. Predictions of the residue status (ordered or disordered) for proteins with prion-like activity using the IsUnstruct program[47]: A)Sup35; B) URE2; and C) Rnq1p. The continuous line at 0.5 of the Y-axis is the threshold line for residues to be disordered. Prion-like domains are indicated by the light-green color.

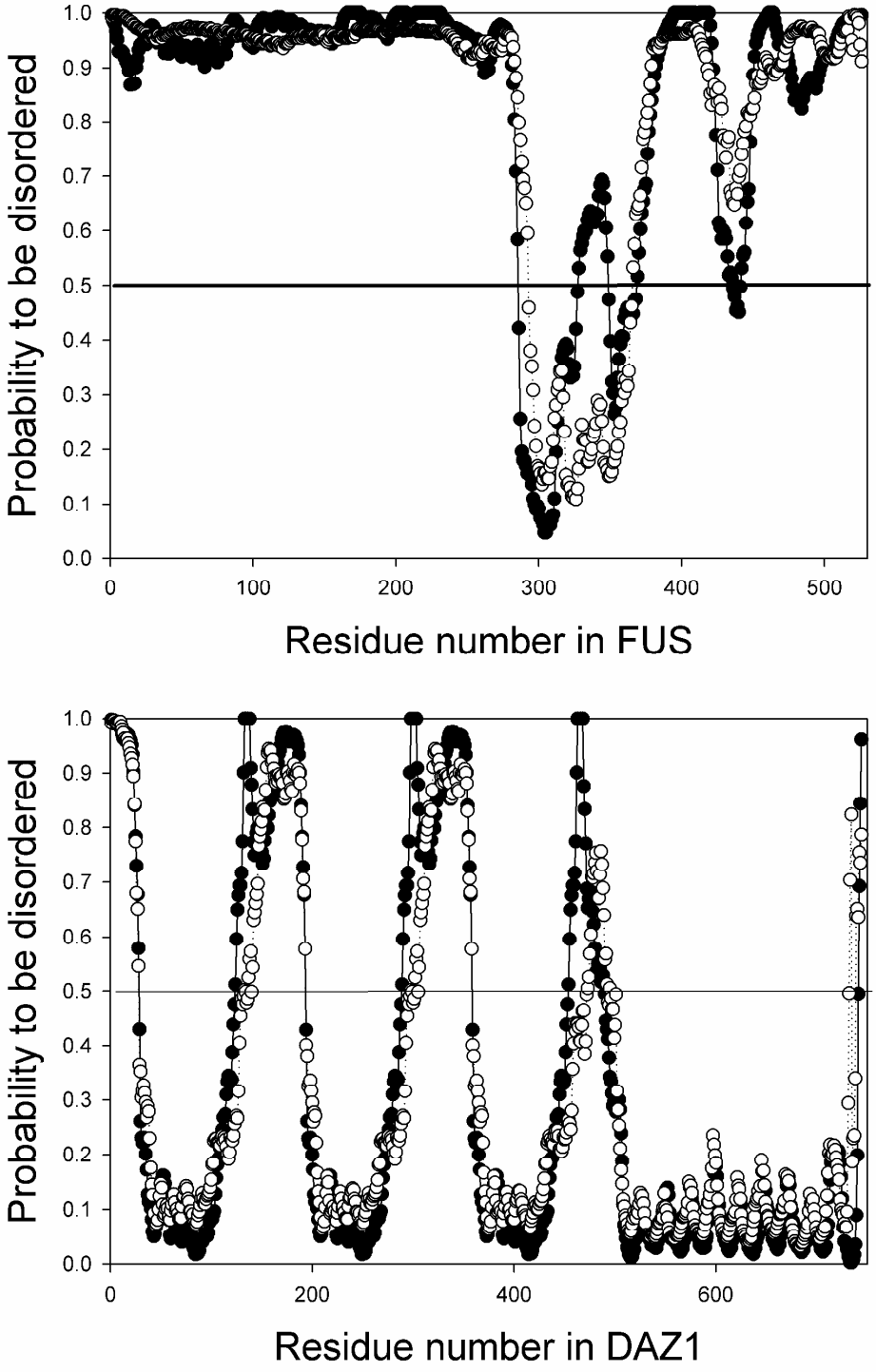


Figure 3. Comparison of predictions using the two different programs (black circles correspond to IsUnstruct and white circles to PONDR-FIT) for the A) FUS and B) DAZ1 proteins.

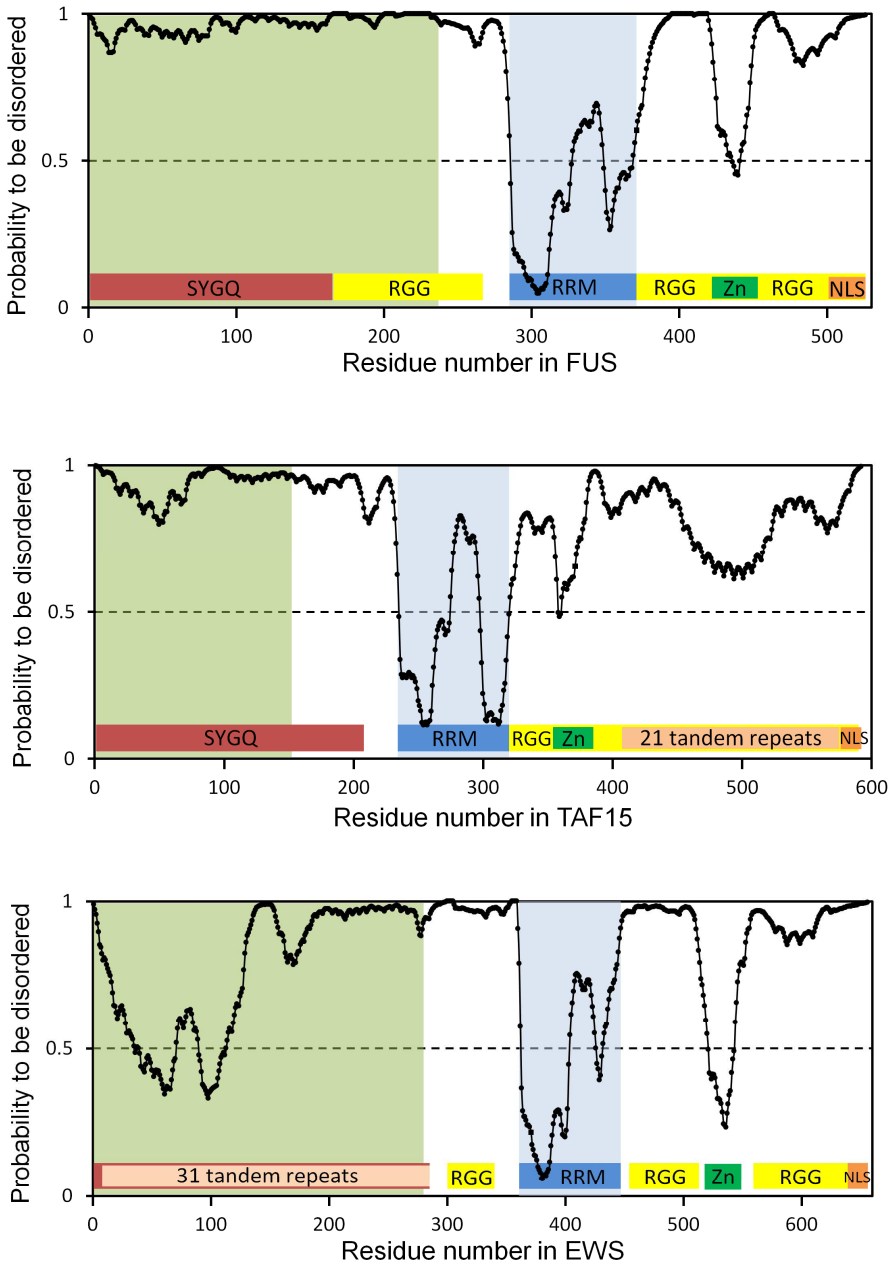


Figure 4. Predictions of the residue status (ordered or disordered) with the IsUnstruct program [47] for the FET family: A) RNA-binding protein FUS; B) TATA-binding protein-associated factor 2N; and C) RNA-binding protein EWS. The continuous line at 0.5 of the Y-axis is the threshold line for residues to be disordered. Prion-like domains are indicated by the light-green color. SYGQ corresponds to a region rich in serine, tyrosine, glycine, and glutamine. RRM represents an RNA recognition motif. RGG corresponds to a region rich in arginine and glycine. Zn represents a zinc finger motif.

Table 1. Repeats in the yeast and human prion proteins

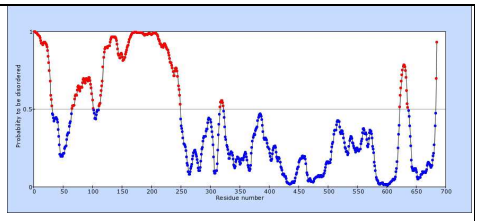
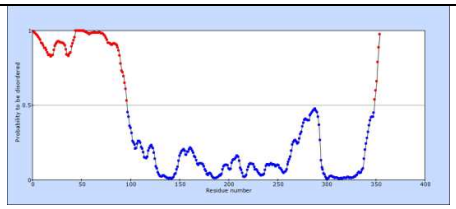
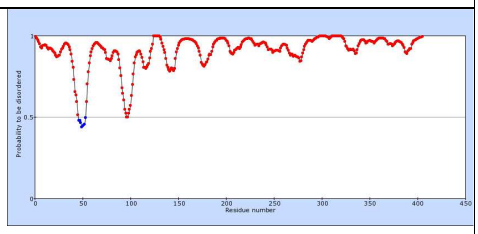
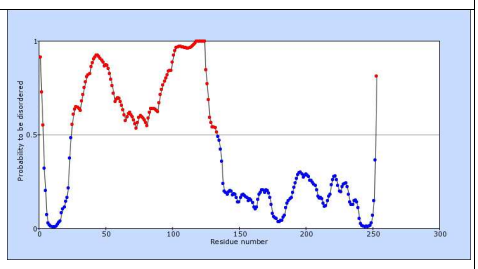
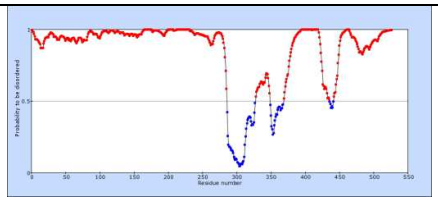
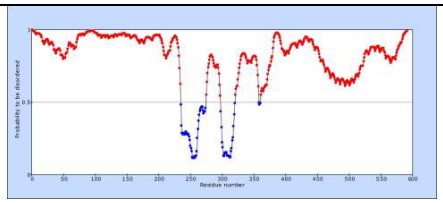
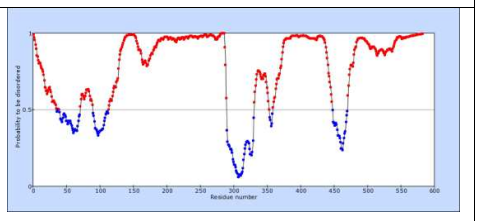
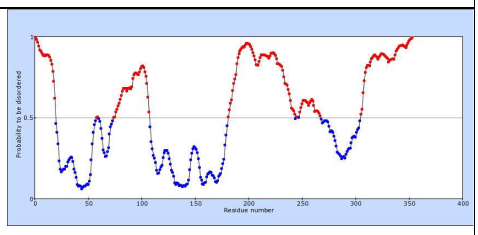
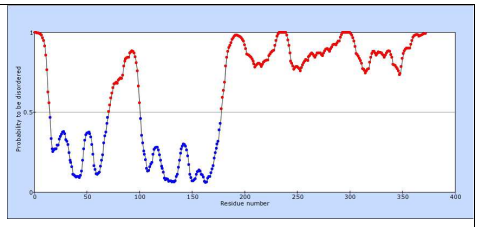
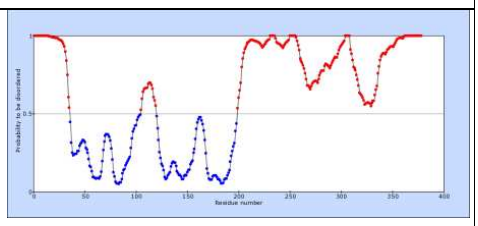
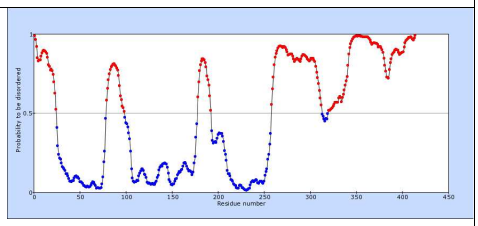
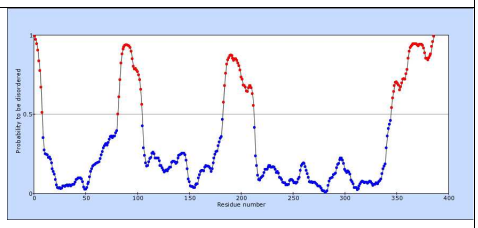
1	>tr|H9BH49|H9BH49_YEASX Sup35 OS=Saccharomyces cerevisiae GN=SUP35 PE=4 SV=1 MSDSNQGNNQQNYQQYSQNGNQQQGNNRYQGYQAYNAQAQPAGGYYQNYQGYSGYQ QGGYQQYNPDAGYQQQYNPQGGYQQYNPQGGYQQQFNPQGGRGNYKNFNYNNNLQG YQAGFQPQSQGMSLNDFQKQQKQAAPKPKKTLKLVSSSGIKLANATKKVDTKPAES DKKEEEKSAETKEPTKEPTKVEEPVKKEEKPVQTEEKKEEKSELPKVEDLKISEST HNTNNANVTSADALIKEQEEEVDDEVVNDMFGGKDHVSLIFMGHVDAGKSTMGGNL LYLTGSVDKRTIEKYEREAKDAGRQGWYLSWVMDTNKEERNDGKTIEVGKAYFETE KRRYTILDAPGHKMYVSEMIGGASQADVGVLVISARKGEYETGFERGGQTREHALL AKTQGVNKMAVVVNKMDDPTVNWSKERYDQCVSNVSNFLRAIGYNIKTDVVFMPVS GYSGANLKDHVDPKECPWYTGPTLLEYLDTMTHVDRHINAPFMLPIAAKMKDLGTI VEGKIESGHIKKGQSTLLMPNKTAVEIQNIYNETENEVDMAMCGEQVKLRIKGVEE EDISPGFVLTSPKNPIKSVTKFVAQIAIVELKSIIAAGFSCVMHVHTAIEEVHIVK LLHKLEKGTNRKSKKPPAFAKKGMKVIAVLETEAPVCVETYQDYPQLGRFTLRDQG TTIAIGKIVKIAE	
2	>tr|Q7LLZ5|Q7LLZ5_YEASX Ure2p OS=Saccharomyces cerevisiae GN=URE2 PE=3 SV=1 MMNNNGNQVSNLSNALRQVNIGSRNSNTTTDQSNINFEFSTGVNNNNNNNSSSNNN NVQNNNSGRNGSQNNDNENNIKNTLEQHRQQQQAFSDMSHVEYSRITKFFQEQPLE GYTLFSHRSAPNGFKVAIVLSELGFHYNTIFLDFNLGEHRAPEFVSVNPNARVPAL IDHGMDNLSIWESGAILLHLVNKYYKETGNPLLWSDDLADQSQINAWLFFQTSGHA PMIGQALHFRYFHSQKIASAVERYTDEVRRVYGVVEMALAERREALVMELDTENAA AYSAGTTPMSQSRFFDYPVWLVGDKLTIADLAFVPWNNVVDRIGINIKIEFPEVYK WTKHMMRRPAVIKALRGE	
3	>tr|K4I0B6|K4I0B6_YEASX Rnq1p OS=Saccharomyces cerevisiae GN=RNQ1 PE=4 SV=1 MDTDKLISEAESHFSQGNHAEAVAKLTSAAQSNPNDEQMSTIESLIQKIAGYVMDN RSGGSDASQDRAAGGGSSFMNTLMADSKGSSQTQLGKLALLATVMTHSSNKGSSNR GFDVGTVMSMLSGSGGGSQSMGASGLAALASQFFKSGNNSQGQGQGQGQGQGQGG QGQGSFTALASLASSFMNSNNNNQQGQNQSSGGSSFGALASMASSFMHSNNNQNSN NSQQGYNQSYQNGNQNSQGYNNQQYQGGNGGYQQQQGQSGGAFSSLASMAQSYLGG GQTQSNQQQYNQQGQNNQQQYQQQGQNYQHQQQGQQQQQGHSSSFSALASMASSYL GNNSNSNSSYGGQQQANEYGRPQQNGQQQSNEYGRPQYGGNQNSNGQHESFNFSGN FSQQNNNGNQNRY	
4	>sp|P04156|PRIO_HUMAN Major prion protein OS=Homo sapiens GN=PRNP PE=1 SV=1 MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGGGG WGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHM AGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDE YSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQA YYQRGSSMVLFSSPPVILLISFLIFLIVG	

Table 2. Human RNA-binding proteins with prion-like domains involved in neurodegenerative diseases

1	>sp|P35637|**FUS_HUMAN** RNA-binding protein FUS OS=Homo sapiens GN=FUS PE=1 SV=1, **N-terminal,1-237,ALS, FTLD** MASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQSYSGYSQSTDTSGYGQSSYS SYGQSQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQSSYPGYGQQPAPSSTSGS YGSSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYGQQNQYNSSSGGG GGGGGGGNYGQDQSSMSSGGGSGGGYGNQDQSGGGGSGGYGQQDRGGRGRGGSGGG GGGGGGGYNRSSGGYEPRGRGGGRGGRGGMGGSDRGGFNKFGGPRDQGSRHDSEQD NSDNNTIFVQGLGENVTIESVADYFKQIGIIKTNKKTGQPMINLYTDRETGKLKGE ATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGGGNGRGGRGRGGPM GRGGYGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCK APKPDGPGGGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGGDRG GFGPGKMDSRGEHRQDRRERPY	
2	>tr|Q86X94|Q86X94_HUMAN **TAF15 RNA polymerase II**, TATA box binding protein (TBP)-associated factor, 68kDa OS=Homo sapiens GN=TAF15 PE=2 SV=2, **N-terminal,1-152, ALS, FTLD** MSDSGSYGQSGGEQQSYSTYGNPGSQGYGQASQSYSGYGQTTDSSYGQNYSGYSSY GQSQSGYSQSYGGYENQKQSSYSQQPYNNQGQQQNMESSGSQGGRAPSYDQPDYGQ QDSYDQQSGYDQHQGSYDEQSNYDQQHDSYSQNQQSYHSQRENYSHHTQDDRRDVS RYGEDNRGYGGSQGGGRGRGGYDKDGRGPMTGSSGGDRGGFKNFGGHRDYGPRTDA DSESDNSDNNTIFVQGLGEGVSTDQVGEFFKQIGIIKTNKKTGKPMINLYTDKDTG KPKGEATVSFDDPPSAKAAIDWFDGKEFHGNIIKVSFATRRPEFMRGGGSGGGRRG RGGYRGRGGFQGRGGDPKSGDWVCPNPSCGNMNFARRNSCNQCNEPRPEDSRPSGG DFRGRGYGGERGYRGRGGRGGDRGGYGGDRSGGGYGGDRSSGGGYSGDRSGGGYGG DRSGGGYGGDRGGGYGGDRGGGYGGDRGGGYGGDRGGYGGDRGGYGGDRGGYGGD RGGYGGDRGGYGGDRGGYGGDRSRGGYGGDRGGGSGYGGDRSGGYGGDRSGGGYGG DRGGGYGGDRGGYGGKMGGRNDYRNDQRNRPY	
3	>sp|Q01844|**EWS**_HUMAN RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=1 SV=1, **N-terminal,1-280, ALS, FTLD, Ewing sarcoma (ES)** MASTDYSTYSQAAAQQGYSAYTAQPTQGYAQTTQAYGQQSYGTYGQPTDVSYTQAQ TTATYGQTAYATSYGQPPTGYTTPTAPQAYSQPVQGYGTGAYDTTTATVTTTQASY AAQSAYGTQPAYPAYGQQPAATAPTRPQDGNKPTETSQPQSSTGGYNQPSLGYGQS NYSYPQVPGSYPMQPVTAPPSYPPTSYSSTQPTSYDQSSYSQQNTYGQPSSYGQQS SYGQQSSYGQQPPTSYPPQTGSYSQAPSQYSQQSSSYGQQSSFRQDHPSSMGVYGQ ESGGFSGPGENRSMSGPDNRGRGRGGFDRGGMSRGGRGGGRGGMGSAGERGGFNKP GGPMDEGPDLDLGPPVDPDEDSDNSAIYVQGLNDSVTLDDLADFFKQCGVVKMNKR TGQPMIHIYLDKETGKPKGDATVSYEDPPTAKAAVEWFDGKDFQGSKLKVSLARKK PPMNSMRGGLPPREGRGMPPPLRGGPGGPGGPGGPMGRMGGRGGDRGGFPPRGPRG SRGNPSGGGNVQHRAGDWQCPNPGCGNQNFAWRTECNQCKAPKPEGFLPPPFPPPG GDRGRGGPGGMRGGRGGLMDRGGPGGMFRGGRGGDRGGFRGGRGMDRGGFGGGRRG GPGGPPGPLMEQMGGRRGGRGGPGKMDKGEHRQERRDRPY	
4	>sp|P22626|ROA2_HUMAN Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens GN=HNRNPA2B1 PE=1 SV=2, **C- terminal,235-327, IBMPFD** MEKTLETVPLERKKREKEQFRKLFIGGLSFETTEESLRNYYEQWGKLTDCVVMRDP ASKRSRGFGFVTFSSMAEVDAAMAARPHSIDGRVVEPKRAVAREESGKPGAHVTVK KLFVGGIKEDTEEHHLRDYFEEYGKIDTIEIITDRQSGKKRGFGFVTFDDHDPVDK IVLQKYHTINGHNAEVRKALSRQEMQEVQSSRSGRGGNFGFGDSRGGGGNFGPGPG SNFRGGSDGYGSGRGFGDGYNGYGGGPGGGNFGGSPGYGGGRGGYGGGGPGYGNQG GGYGGGYDNYGGGNYGSGNYNDFGNYNQQPSNYGPMKSGNFGGSRNMGGPYGGGNY GPGGSGGSGGYGGRSRY	

5	>sp|P09651|ROA1_HUMAN Heterogeneous nuclear ribonucleoprotein A1 OS=Homo sapiens GN=HNRNPA1 PE=1 SV=5, **C-terminal, 186-372,IBMPFD, ALS** MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRG FGFVTYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKIFVGGI KEDTEEHHLRDYFEQYGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYH TVNGHNCEVRKALSKQEMASASSSQRGRSGSGNFGGGRGGGFGGNDNFGRGGNFSG RGGFGGSRGGGGYGGSGDGYNGFGNDGGYGGGGPGYSGGSRGYGSGGQGYGNQGSG YGGSGSYDSYNNGGGGGFGGGSGSNFGGGGSYNDFGNYNNQSSNFGPMKGGNFGGR SSGPYGGGGQYFAKPRNQGGYGGSSSSSSYGSGRRF	
6	>sp|P51991|ROA3_HUMAN Heterogeneous nuclear ribonucleoprotein A3 OS=Homo sapiens GN=HNRNPA3 PE=1 SV=2, **C-terminal,207-378,C9orf72 ALS/FTLD** MEVKPPPGRPQPDSGRRRRRRGEEGHDPKEPEQLRKLFIGGLSFETTDDSLREHFE KWGTLTDCVVMRDPQTKRSRGFGFVTYSCVEEVDAAMCARPHKVDGRVVEPKRAVS REDSVKPGAHLTVKKIFVGGIKEDTEEYNLRDYFEKYGKIETIEVMEDRQSGKKRG FAFVTFDDHDTVDKIVVQKYHTINGHNCEVKKALSKQEMQSAGSQRGRGGGSGNFM GRGGNFGGGGGNFGRGGNFGGRGGYGGGGGGSRGSYGGGDGGYNGFGGDGGNYGGG PGYSSRGGYGGGGPGYGNQGGGYGGGGGYDGYNEGGNFGGGNYGGGGNYNDFGNYS GQQQSNYGPMKGGSFGGRSSGSPYGGGYGSGGGSGGYGSRRF	
7	>sp|Q13148|TADBP_HUMAN TAR DNA-binding protein 43 OS=Homo sapiens GN=TARDBP PE=1 SV=1, **C-terminal, 277-414, ALS, FTLD** MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRL VEGILHAPDAGWGNLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLP WKTTEQDLKEYFSTFGEVLMVQVKKDLKTGHSKGFGFVRFTEYETQVKVMSQRHMI DGRWCDCKLPNSKQSQDEPLRSRKVFVGRCTEDMTEDELREFFSQYGDVMDVFIPK PFRAFAFVTFADDQIAQSLCGEDLIIKGISVHISNAEPKHNSNRQLERSGRFGGNP GGFGNQGGFGNSRGGGAGLGNNQGSNMGGGMNFGAFSINPAMMAAAQAALQSSWGM MGMLASQQNQSGPSGNNQNQGNMQREPNQAFGSGNNSYSGSNSGAAIGWGSASNAG SGSGFNGGFGSSMDSKSSGWGM	
8	>sp|P31483|TIA1_HUMAN Nucleolysin TIA-1 isoform p40 OS=Homo sapiens GN=TIA1 PE=1 SV=3 **C-terminal,292-386,Welander distal myopathy** MEDEMPKTLYVGNLSRDVTEALILQLFSQIGPCKNCKMIMDTAGNDPYCFVEFHEH RHAAAALAAMNGRKIMGKEVKVNWATTPSSQKKDTSSSTVVSTQRSQDHFHVFVGD LSPEITTEDIKAAFAPFGRISDARVVKDMATGKSKGYGFVSFFNKWDAENAIQQMG GQWLGGRQIRTNWATRKPPAPKSTYESNTKQLSYDEVVNQSSPSNCTVYCGGVTSG LTEQLMRQTFSPFGQIMEIRVFPDKGYSFVRFNSHESAAHAIVSVNGTTIEGHVVK CYWGKETLDMINPVQQQNQIGYPQPYGQWGQWYGNAQQIGQYMPNGWQVPAYGMYG QAWNQQGFNQTQSSAPWMGPNYGVQPPQGQNGSMLPNQPSGYRVAGYETQ	

ALS - amyotrophic lateral sclerosis, FTLD - Frontotemporal lobar degeneration, ES - Ewing sarcoma, IBMPFD - Inclusion body myopathy with early-onset Paget disease with or without frontotemporal dementia, C9orf72 - chromosome 9 open reading frame 72