We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 6

Influence of Repeats in the Protein Chain on its Aggregation Capacity for ALS-Associated Proteins

Oxana V. Galzitskaya

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/63104

Abstract

Studies of diseases associated with pathological irreversible aggregation of proteins have become of special relevance and attracted the attention of researchers throughout the world because of the appearance of a new conceptual model based on the capacity of some proteins to self-assemble by the prion mechanism. Along with direct prion diseases, such as bovine rabies and Creutzfeldt-Jakob disease in humans, a great number of neurodegenerative disorders associated with the formation of aggregates through the prion mechanism are revealed. These disorders include Alzheimer's and Parkinson's diseases, amyotrophic lateral sclerosis, Huntington disease, and mucoviscidosis, some types of diabetes and hereditary cataracts. The listed diseases are caused by transition of a "healthy" protein or peptide molecule from the native conformation to a very stable "pathological" form. In this case, molecules in the "pathological" conformation aggregate specifically, forming amyloid fibrils that can multiply infinitely. An important result of studying the molecular mechanisms of prion diseases and different proteinopathies, associated with the formation of pathological aggregations by the prion mechanism, is the discovery of protein chain regions responsible for their aggregation. The ability to regulate aggregation (fibrillation) of proteins can be the focal tool for the drug development. Herein by the example of 29 RNA-binding proteins with prion-like domains, we demonstrate what role the amino acid repeats in prion-like domains can play. For these proteins, quite different repeats are revealed in the disordered part of the protein chain predicted with bioinformatics methods. Ten proteins of the 29 RNA-binding proteins are involved in the development of some diseases. The prion-like domains of FUS, TAF15, and EWS are critical for the aggregation of proteins associated with human neurodegenerative diseases. Proteins of this family are involved not only in neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), Huntington disease, spinocerebral ataxy, and dentatorubral pallidoluysian atrophy, but also in the formation of human mixoid liposarcoma. It can be suggested that the presence of a great number of repeats in prion-like domains of RNA-binding proteins can accelerate the formation of a dynamic beta-structure and

INTECH
open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

pathological aggregates, which are crucibles of amyotrophic lateral sclerosis (ALS) pathogenesis.

Keywords: disordered regions, repeats, motifs of low complexity, Alzheimer's disease, amyotrophic lateral sclerosis (ALS)

1. Introduction

Misfolded proteins leading to formation of protein aggregations are a reason for many diseases. There is no definite answer to the question what causes death of cells where such aggregations have been found: is this the cell defence mechanism or plain death? In a normal state, cells that accumulate such aggregations are usually programmed to death or apoptosis. Cell fragments subjected to apoptosis are removed by phagocytic cells. However, aggregations like amyloid fibrils are known to be resistant to the action of different proteases [1] that can impede the effective termination of efferocytosis and, as a result, accumulate in tissues. In any case, there is undoubtedly close association of the formation of aggregations and the development of many fatal diseases. There are several models describing the process of fibril formation. For example, in order to begin aggregation, proteins should be preliminarily unfolded or partially folded [2]. As known, the generation of fibrils is facilitated by denaturing conditions. At the same time, aggregation of peptides and proteins involved in pathogenesis of such types of amyloidosis as type II diabetes, Alzheimer's, and Parkinson's diseases does not necessitate preliminary unfolding of a protein molecule. But these data sooner support the general rule, because under physiological conditions most of these proteins have no definite structure, i.e., are natively unstructured [3]. However, most natively unfolded protein in vivo does not aggregate [4]. Moreover, unstructured proteins are resistant to denaturing conditions, i.e., to the factors bringing about stress, and, in the first place, high temperatures [5]. It was demonstrated that the absence of structure does not correlate with the aggregation capacity [6]. Therefore to avoid spontaneous self-assembly of a protein molecule, the evolutionary selection has led to an increased content of such amino acids as proline and glycine that inhibit protein aggregation [7] and to an increased content of charged amino acids [8]. On the contrary, due to a large number of amyloidogenic regions, globular proteins have developed capacity to avoid aggregation because of rapid folding into a globular structure. This shows that protein unfolding is necessary but not sufficient for activation of amyloid fibrils. It is most likely that there should be special motifs of amino acid sequences exposed to the solvent, which are more liable to aggregation than other regions of the amino acid sequences. Experimental data corroborate the hypothesis that there are small regions of a protein molecule responsible to the amyloidogenic behavior [9–11].

2. Role of repeats in protein aggregation

It has remained unclear what the mechanism of the earliest stage of initiation of the pathologic irreversible aggregation of proteins is and how this process is triggered in healthy organisms. It is supposed that the key role in the development of systematic amyloidosis belongs to the so-called primes or factors accelerating pathologic aggregation. Such primes can be infectious agents as well protein molecule regions containing motifs of low complexity, especially when these motifs are recurrent. It was shown that the more frequently the repeats occur in a protein sequence, the less structured the protein is. Generally, most homo-repeats are not structured [12–15]. Nevertheless, this is not characteristic of fibrillar proteins [16], the capacity of which to aggregation depends strongly on the amino acid sequence of the protein [17–19]. For example, if the identity of the amino acid sequence of immunoglobulin domains is lower than 30–40%, the proteins lose their capacity to co-aggregation [18]. Using the bioinformatics analysis, it has been established that in a large number of multidomain proteins the identity of amino acid sequences of their domains is below 40%. It suggests a conclusion that in this way the domains avoid mutual aggregation.

The existence of repeats in proteins and the clarification of their special roles are in the focus of attention of researchers. The role of different repeats is studied actively, including repeats such as PGMG (GPGM) and PNN upon biomineralization of PM27 proteins [20], NPNA (NANP) repeats in circumsporozoite protein of *Plasmodium falciparum*, YSPTSPS repeats in RNA-polymerase II, PHGGGWGQ repeats in prion protein, YGHGGG(N) and YNHGGG(G) repeats in plant proteins rich in glycine, PGQGQQ, PGQGQQGQQ, and GYYPTSOQQ repeats in wheat gluten, and FGGMGGGKGG repeat in bivalves (*Aequipecten abductin*). As a rule, conformational loops formed by these repeats are stabilized upon interaction with different cations; they are characterized by noncovalent interactions, particularly, interactions of aromatic groups. It was demonstrated that many tandem repeats add plasticity and mobility to the protein [21]. A leader on the occurrence of repeats in proteins is *P. falciparum* where 35% proteins of 5300 have repeats. Moreover, *P. falciparum* contains 24% proteins with prion-like domains rich in asparagine, whereas flies have only 3.4% of such proteins [22]. The occurrence frequency of asparagine repeats in our HRaP database (http://bioinfo.protres.ru/hrap/) [14] shows that *P. falciparum* is the leader. Of interest is the fact that one of the functions of these homorepeats is connected with the parasitic life [14]. For example, the sequence of the protein from *P. falciparum* (ID Q8IKW2_PLAF7 1304 aa 493.P_falciparum) has asparagine homorepeats of different lengths, the maximal length reaching 41 amino acids. The basic functions of this protein are associated with processes such as deacylation of proteins and "silence" of chromatin [14].

3. Cell stress and generation of stress granules

Cell stress can also be an important factor of initiation of aggregation even though each cell has developed an intricate defence system, because it is subjected to destructive stress action

at regular intervals. A striking example of waiting till stress is over in the formation of stress granules (SG), when the nontranslated mRNA and RNA-binding proteins are assembled in ribonucleoprotein (RNP) complexes in order to terminate protein synthesis and thereby maintain cell energy. In this case, only those proteins that are synthesized are required for cell survival [23]. This means that after termination of the stress action, SGs disintegrate rapidly and the "released" mRNA resumes its functioning. However, if due to some reasons, the residence time of such proteins in SGs increases or their concentration in SGs exceeds the norm, disintegration of SGs can be impeded, creating favorable conditions for generation of the "center of aggregation initiation" that may induce transition to irreversible pathologic protein aggregation [24]. A detailed analysis of the mechanism of assembly and disassembly of SGs can provide a new insight into the development of diseases associated with this process and suggest novel therapeutic approaches. Since the main function of SGs is protection of cells from stress, many investigations are conducted to reveal factors shifting the balance of reversible aggregation towards pathology after stress termination (**Figure 1**).

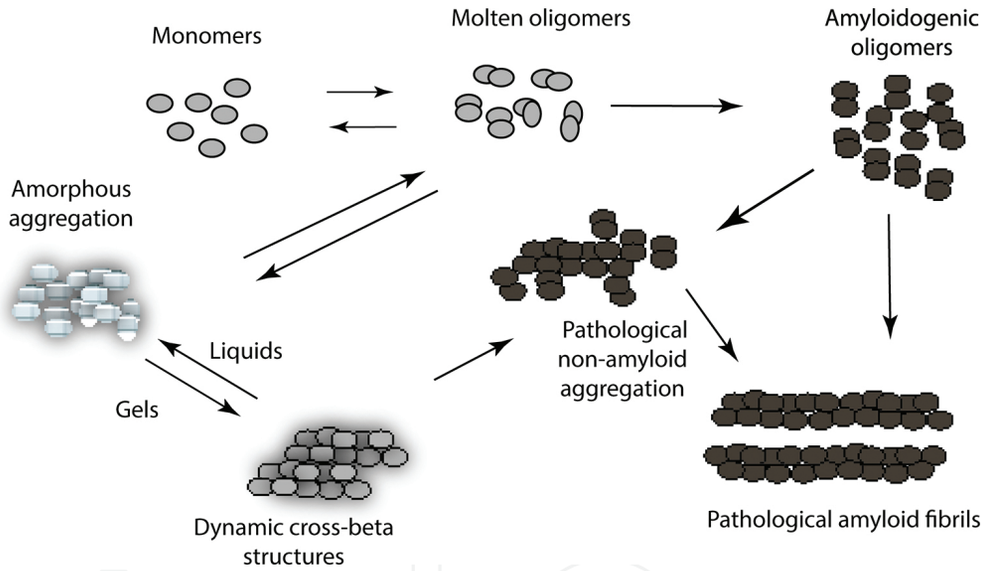


Figure 1. Schematic representation of different conformational states of self-assembly and disintegration of prion-like domains and a possible transition to irreversible pathologic aggregation. Modified and adapted from Li et al. [25].

It is assumed that due to self-assembly, RNA-binding proteins can facilitate formation of SGs using prion-like domains [26–28]. At any rate, it has been established that namely the structured part of RNA-binding proteins is responsible for the formation of hydrogel and binding to it [29]. Thus, protein FUS maintained its capacity to form gel even without the removal of the C-terminal region that corresponds to the RNA-binding domain and lost this capacity upon removal of the N-terminal unstructured region corresponding to the prion-like domain. The capacity to bind to hydrogel has been established for proteins such as hnRNPA2, RBM3 RNA-binding proteins, hnRNPA1, TIA1, CPEB2, FMRP, CIRBP, TDP43, and yeast Sup35. The formation of hydrogel was also demonstrated for the hnRNPA2 protein, and it was shown that proteins are retained to a different degree by hydrogel formed of different proteins [29].

4. The role of RNA-binding proteins with prion-like domains in diseases

Studies of the human genome have allowed isolating a set of RNA-binding proteins with a canonical RNA-binding motif. For example, prion-like domains were predicted for 29 of the 210 RNA-binding proteins for human diseases [30]. Ten of these 29 proteins are associated with neurodegenerative diseases, i.e., proteinopathies [31]. In this chapter, we focus our attention on isolation of repeats in well-studied proteins of the FET/TET family that are included in this group of 10. To elucidate the mechanism of pathologic aggregation of proteins, we have set ourselves the task to determine what repeats of amino acid residues in RNA-binding proteins can be responsible for reversible and irreversible aggregation.

The prion-like domains predicted for TDP-43 and FUS overlap with the region containing a large number of glycine residues. The prion-like domains are critical for aggregation of these proteins associated with human neurodegenerative diseases [32]. It was found that protein TAF15 is also involved in the development of disorder such as amyotrophic lateral sclerosis (ALS) or Lou Gering's disease [33]. At first, this protein was discovered using a bioinformatics analysis as a possible candidate and later it was revealed in ALS patients. It turned out that this protein has very similar properties to those of TDP-43 and FUS proteins involved in this disease. The only difference is that mutations of the protein associated with this disease are capable of higher aggregation in vitro than the wild-type protein, exert a stronger effect on the lifetime, and lead to incorrect localization of proteins in the spinal cord of mammals [33].

Among 29 candidates of RNA-binding proteins with predicted prion-like domains, the first and second places belong to FUS and TAF15, whereas TDP-43 is the 10th in this list [34]. Probably, it is expedient to pay attention to the first 10 proteins because the third place belongs to protein EWS [35]. They should also be detected in ALS patients.

5. Members of the FET family: FUS, EWS, TAF15

Members of the FET family are very similar RNA-binding proteins containing the following main domains in their structure: a SYGQ-rich N-terminal prion-like domain enriched with uncharged polar amino acids (asparagine, glutamine, serine, and tyrosine) and glycine [36], an RNA-binding motif (RNA-recognition motif, RRM), a "zinc finger" motif, and several RGG domains rich in glycine [37–39]. Three proteins (FUS, EWS, and TAF15) also have a nuclear location signal (NLS) that is recognized by the nuclear receptor transportin and is responsible for protein transport from the cytoplasm into the nucleus and back [27, 40–43]. Genes of the FET family are expressed practically everywhere. Proteins of this family are involved in regulation of different stages of gene expression, including transcription, pre-mRNA splicing, mRNA transport, and also take part in DNA repair [44–46]. The family name stems from three capital letters of its three members: proteins FUS/TLS (fused in sarcoma or translocated in liposarcoma), EWS (Ewing's sarcoma), and TAF15 (TATA-binding protein-associated factor 2N) [47]. These RNA-binding proteins participate in many cell processes, including transcription, pre-mRNA splicing, DNA repair, and mRNA transport in neurons [44–46]. As a rule,

under stress conditions such RNA-binding proteins shift from the nucleus into the cytoplasm and take part in the formation of SGs, after which they return to the nucleus. This cycle is multiply repeated during the cell lifetime, and it is not excluded that under such conditions errors may occur in its functioning. The mechanism of activity of FET proteins, their cellular localization, and determination of domains involved in different processes are to be clarified.

Prion-like domains of these proteins associated with human neurodegenerative diseases are critical for their aggregation [32]. Proteins of this family are involved not only in neurodegenerative diseases, such as ALS [33], Huntington's disease, spinocerebellar ataxia, and dentatorubral cerebellar atrophy [48], but also in the development of human mixoid liposarcoma [49–51].

5.1. FUS: Its functions, structural peculiarities of a prion-like domain, mutations associated with diseases

A member of the FET family is the RNA-binding protein FUS, for which a number of mutations associated with ALS have been revealed [27, 40–43, 52]. Under normal conditions, this protein is localized mostly in the nucleus, whereas under pathologic conditions its aggregations are accumulated in the cytoplasm. It was demonstrated on several cell models that FUS delocalized in the cytoplasm was accumulated in SGs. This permitted a conclusion that a high concentration of the protein facilitates generation of protein aggregations under FUS pathology [53, 54]. Nonetheless an exact mechanism by which this transformation occurs remains unclear. It is not excluded that prion-like domains affect aggregation-like prions or fibril formation analogous to amyloidogenesis. It is also possible that FUS, accumulated in large amounts, aggregates in the cytoplasm independently of its capacity to sequestration in SGs. The protein accumulation can also facilitate dysfunction of the protein intracellular degradation systems that frequently occurs parallel to neurodegenerative disorders [55]. It should be noted that under sporadic ALS diseases, the control of splicing of a large amount of genes is violated [56]. The FUS protein also regulates alternative splicing, as a rule binding to transcripts containing large introns. Therefore, accumulation of FUS in the cytoplasm can result in the loss of its function in the nucleus and consequently violation of the alternative splicing of a number of genes, e.g., coding proteins associated with the growth and development of axons and the cell cytoskeleton [57]. This version is in agreement with the results obtained on a drosophila model, when an increased FUS concentration in the cell was extremely toxic for organisms that can also cause dysfunction during splicing. In addition, enhanced expression of the gaz gene of the fus ortholog caused death at the pupal stage [58].

In transgenic *Caenorhabditis elegans* worms, in which serine 57 was removed in protein FUS, paralysis occurred much earlier when compared to transgenic species with a full-sized FUS protein [59]. Therewith, the sequence of the FUS protein itself corresponded to that of the FUS protein from the human proteome. This mutation (removal of serine 57) is associated with the sporadic form of ALS [60]. It has been shown that for aggregation and toxicity of yeasts a prion-like (1–239) and an RGG2 (374–422) domains are required, though FUS (1–359) is sufficient for aggregation in models of neuroblastoma cells [54]. Domain RGG3 does not affect aggregation,

and mutations in (502–526) lead only to the protein accumulation in the cytoplasm and its insertion in SGs. In yeasts, FUS was accumulated only in the cytoplasm [61]. Some components of SGs, such as translation initiation factors and poly-A-binding proteins, suppress toxicity of FUS aggregations. It is important that many proteins associated with RNA metabolism can affect the toxicity of FUS aggregations.

To verify the hypothesis that dislocation of FUS from the nucleus into the cytoplasm leads to the loss of its function in the nucleus, Murakami et al. created transgenic animals with double gene expression. One of these genes (fus) having the N-terminus, labeled with red fluorescent protein RFP, could not move into the cytoplasm under the action of stress, and the other gene (mutant GFP-fus-P525L) was uniformly distributed both over the nucleus and the cytoplasm. After heat shock, most of GFP-FUS-P525L moved into the SGs, whereas GFP-RUS remained in the cytoplasm. Therewith, the damaging effect was the same as in the experiment with expression of only GFP-fus-P525L. Therefore, the authors concluded that accumulation and aggregation of FUS in the cytoplasm are more neurotoxic than that when FUS lost its function in the nucleus [62].

In any case, formation of such aggregations is the reason for apoptosis, i.e., the process of programmed cell death. Apoptosis is characterized by retaining fragmentation of intracellular components with retention of the integrity of the plasmatic membrane that facilitates fast phagocytosis.

5.2. Role of FET proteins in the formation of stress granules

To understand the mechanism of assembly and disassembly of SGs, it is necessary to know what regions of the chain of RNA-binding proteins can perform the function of a prime and what the role of unstructured regions of simple complexity is. It is worth noting that in many respects protein FUS is a perfect model for studying processes involved in the formation of protein aggregations by the prion mechanism. To search for and reveal properties of prion-like domains, we have chosen three proteins from the FET/TET family of 29 RNA-binding proteins of the human proteome, the structures of which included prion-like domains [31]. The prediction was made using the algorithm developed by Alberti et al. [63], which is based on the choice of protein regions of 60 amino acid residues, similar in the amino acid content to the prion domains of yeast proteins, such as Sup35, Ure2p, and Rnq1p [64]. As a rule, these regions are rich in hydrophilic amino acid residues such as glutamine, asparagine, and tyrosine. In the range of proteins used in the prediction of prion-like domains, the first and second places belonged to FUS and TAF15 among 29 candidates of RNA-binding proteins, and the third was protein EWS [35].

The experiments, devoted to disclosing the capacity of protein FUS to aggregate, demonstrated that upon deletion of the most part of the predicted prion-like domain the protein lost its capacity to self-assemble; however, the formed aggregations did not reveal toxicity (**Figure 2**) [67]. Some components of SGs, such as translation initiation factors and poly-A-binding proteins, suppress toxicity of FUS aggregation [67]. It is important that proteins associated with RNA metabolism can affect the toxicity of FUS aggregations. Prion-like (1–239) and RGG2

(374–422) domains are also required for aggregation and toxicity of yeasts, although FUS (1–359) is sufficient for simulations on the neuroblastoma cell culture [68]. Domain RGG3 does not affect aggregation, and mutations at (502–526) result only in accumulation of the protein in the cytoplasm and its insertion in SGs [24, 69]. It should be mentioned that in contrast to mammalian cells, in yeast cells protein FUS is accumulated largely in the cytoplasm [61]. This may be connected with the fact that NLS FUS is not recognized by nuclear receptors of yeasts [65]. In mammalian cells, protein FUS is accumulated in the cytoplasm only when it has mutations distorting the reverse transport into the nucleus [69].

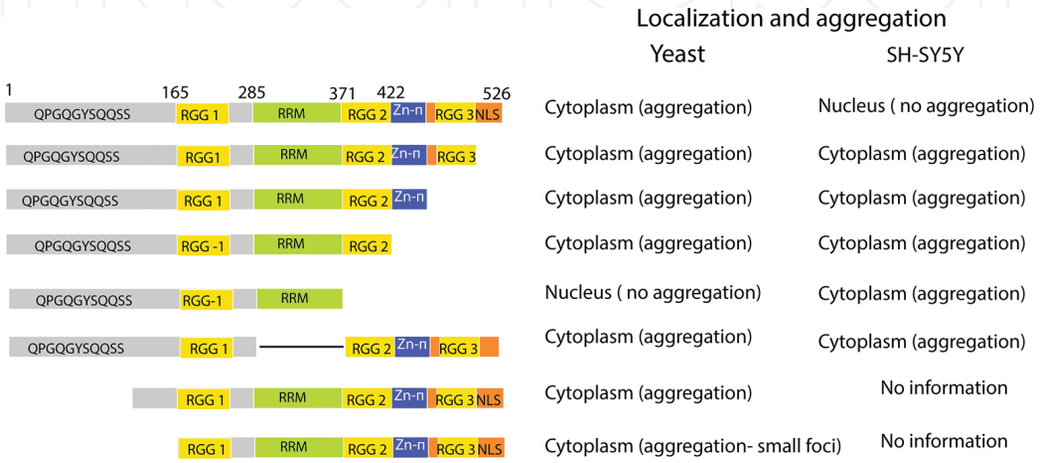


Figure 2. Effect of different constructions of FUS on aggregation and toxicity in yeasts and aggregation and localization in SGs in cell culture of the SH-SY5Y neuroblastoma [65, 66].

The N-terminal domain of FUS has 27 different variants of GYG, GYS, SYG, and SYS triplets (that can be designated as [G/S]Y[G/S] repeats) [70]. Four mutants were obtained with a different number of substituted tyrosine residues for serine ones to demonstrate that namely tyrosine residues are responsible for the formation of hydrogel. There were 5, 9, 15 substitutions and all 27. Neither of the mutants could form hydrogel, however, all of them could equally well bind to it. Mutants with substituted residues 5 and 9 could bind to hydrogel, but the remaining mutants could not [70].

5.3. Disordered regions in proteins of the FET family and search for partners for interactions with these proteins

For all three proteins, the IsUnstruct program [15] predicts the presence of unstructured domains, as a rule, at the N and C termini of the polypeptide chain. Unstructured proteins often play the role of hubs, i.e., have a capacity to concentrate a large number of partners around them (it is accepted that when there are more than five partners, it is a hub) [71]. To what extent is this role validated? To answer this question, it is necessary to determine the presence of functional sites in the protein considered. The search for the number of partners in the STRING database revealed that it exceeds 5 [72] (see **Figure 3**). Usually upon binding to a partner, natively unfolded proteins can acquire a structure that imparts a certain function to

http://dx.doi.org/10.5772/63104

them. In other words, the conformation of unstructured proteins is "dictated" by the interaction partners. This explains their capacity to perform different functions both in the cell and in the extracellular space. For example, the analyzed RNA-binding proteins have regions with large amounts of glycine in addition to large amounts of asparagine, glutamine, and tyrosine, which also facilitates their unfolded state and performance of various functions in the cell because these proteins are involved in the formation of RNP complexes, control of DNA transcription, pre-mRNA splicing, protein posttranslational modification, and many other vital functions [73]. According to the STRING database (version 10), the number of partners is 44 for TAF15, 132 for EWS (**Figure 3**), and 218 for FUS.

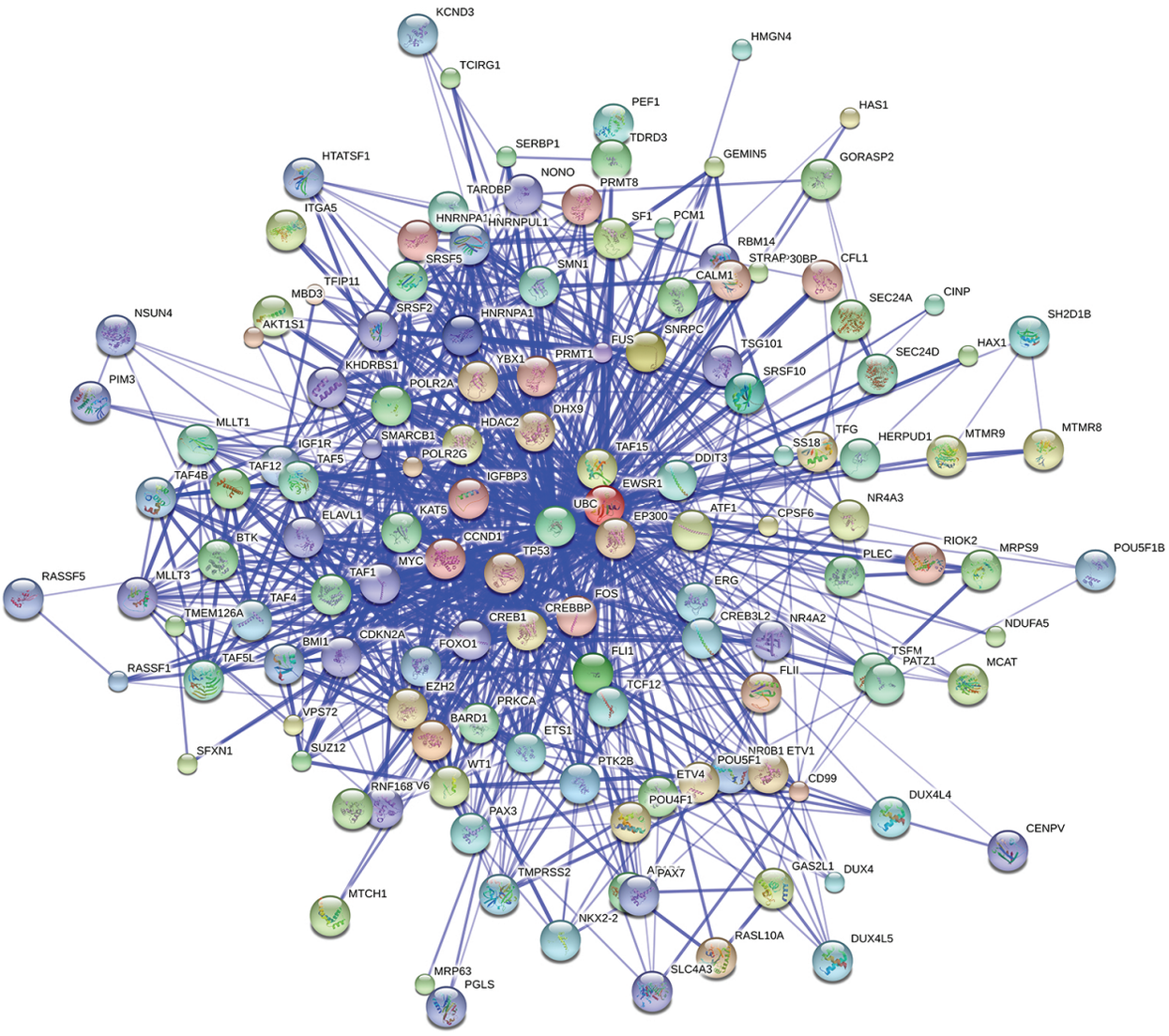


Figure 3. The list of partners for EWS has been obtained from the STRING database using the information from the data bases, data about homologies, possible co-expression, experimental confirmation about interactions, etc. The boundary condition for entry in the list of partners is probability of interactions equal to 0.4. Transparency of edges connecting vertices in the graph designates probability of interaction. The more transparent of edge, the less probability of interaction.

These proteins contain many motifs of simple complexity. As shown, the portion of proteins, included in periodic fragments or homorepeats, is an order of magnitude lower in eukaryotic proteomes than in bacterial ones [74, 75]. The proteins with periodic fragments are extremely nonuniformly distributed both over the kingdoms and over organisms within each kingdom [74, 75]. It is worth underlining that these repeated motifs can be located in the region not predicted as prion-like. It is known that protein Ure2 has regions in the carboxy-terminal domain affecting the capacity of the amino-terminal domain to become prion-like [76]. The available data for FUS allow us to conclude that the presence of both a prion-like domain and a C-terminal region corresponding to the RGG2 region (see **Figure 2**) is important for pathologic aggregation. It is most likely that the RNA-binding domain also contributes to the pathologic aggregation. The RNA-binding motif (RRM) in proteins RUS and TAF15 is highly identical and is retained from species to species. On the contrary, the RNA-binding motif (RRM) in protein EWS differs remarkably from other members of the family and does not preserve 100% identity in different species. For example, for FET proteins, the following amyloidogenic regions, revealed using program FoldAmyloid [77], can be indicated in the RNA-binding domain (RRM): AIYVQ/ADFFK/MIHIYL/VEWFD for EWS; TIFVQ/INLYT/IDWFD for TAF15; and TIFVQ/INLYT/IDWFD for FUS.

The prediction of unstructured regions in proteins of the FET/TET family using program IsUnstruct allowed us to isolate two unstructured domains at the N and C termini and one structured region, corresponding to the RNA-binding domain. **Figure 4** shows probability profiles for amino acid residues of the FET/TET proteins, which make it possible to determine whether they are structured or unstructured. Motifs in the amino acid sequence are shown by different colors. These proteins are characterized by the presence of a large number of homorepeats when one amino acid is recurrent many times. Generally, the longer are the repeats, the higher is the probability that the aggregated protein containing and it is associated with the development of a disease. For example, protein FUS is characterized by five unstructured patterns from the pattern library obtained by us from the Protein Data Bank: GSHM, GGGGSGG, GGGGG, GGSGGGGSGGG, and RGGGGSG. The occurrence of these patterns in the given protein in different organisms (human, monkey, pig, mouse, quicken and fish) can be found in our HRaP database, containing data on the occurrence of unstructured patterns and homorepeats in 122 proteomes [14]. For protein TAF15, the glycine-rich recurrent motif is well isolated at the C-terminus: DRGGGYGG/DRSSGGGYSG/DRGSRGGYGG, which is characteristic of many animals and fish [78]. We isolated 22 repeats, and the Uniprot program finds 21 repeats at the C terminus: GRGGRGG/DRGGYGG (**Figure 4**). As concerns EWS, we can observe 14 repeats as SYSQAPS in the prion-like domain (the N-terminal part) and 6 repeats (DRGRGGPGG) in the C-terminal part (**Figure 4**). It should be noted that 15 imperfect repeated motifs (QPGQGYSQQSS) are positioned in a prion-like region (the N-terminal part) and four repeats (DDRRGGRGGY) in the C-terminal one for FUS (**Figure 4**).

As known, protein regions enriched in glycine residues cannot have a rigid spatial structure; therefore, the main function characterizing this protein region is determined by a number of adjacent amino acid residues. It should be noted that in the mentioned proteins with high toxic aggregation, the glycine repeats adjoin arginine, serine, and tyrosine. Domains rich in glycine

http://dx.doi.org/10.5772/63104

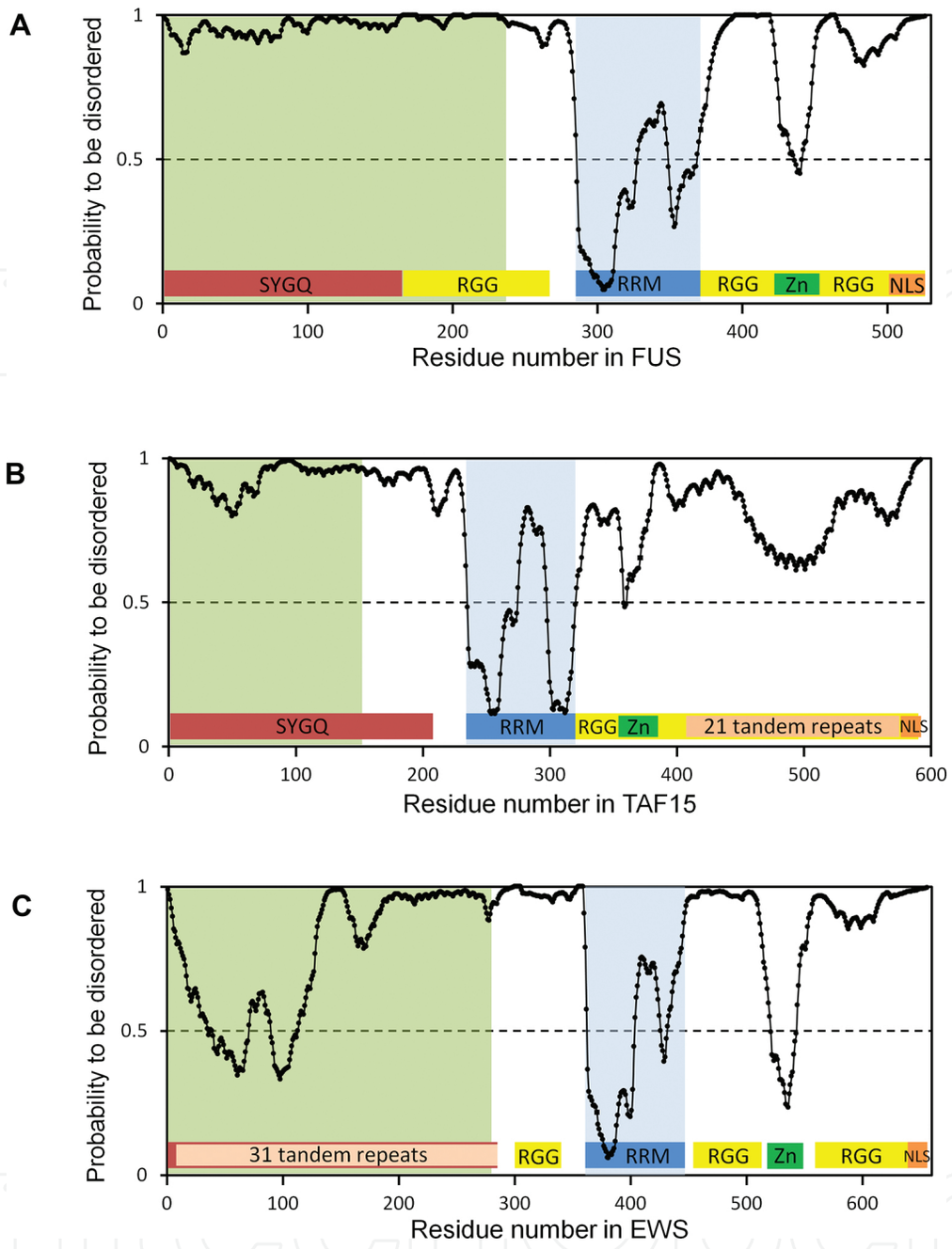


Figure 4. Probability profiles of amino acid residues in proteins of the FET/TET family (A for FUS, B for TAF15, and C for EWS) according to which it is possible to predict possible formation of the structure or its absence by the IsUnstruct program [15].

and arginine (RGG) are known to be responsible for the interaction of proteins with each other and with RNA. As a rule, these interactions are controlled by methylation of arginine [79], whereas phosphorylation of serine residues affects the direct mutual interaction of prion-like domains [45, 80]. Of interest is the fact that deletion of serine (S57), the mutation associated with sporadic form of ALS [81], induced paralysis in transgenic *C. elegans* species, that took place much faster than in transgenic species with a full-sized human FUS [82]. These data make

it possible to suggest that phosphorylation and dephosphorylation of serine residues are critical not only for self-assembly of prion domains but also for disassembly by aggregation when required, e.g., termination of stress action on the cell. It was assumed that the presence of tyrosine residues facilitates the formation of hydrogel. In this connection, it should be mentioned that in spite of high similarity of these proteins, relative to other members of the family, the cytoplasmic aggregation of FUS is more toxic that correlates with longer glycine repeats in the amino acid sequence of FUS [33].

5.4. Zinc-finger motif in proteins of the FET family

All proteins of the FET/TET family are characterized by the presence of a zinc-finger motif. The exclusion is the EWS protein in chickens, because here the zinc-finger motif has not been determined [78]. As known, a classic zinc-finger motif forms a loop, where two cysteine residues and two histidine residues bind zinc ions. The main function of a classic zinc-finger motif is the binding of DNA, which corresponds to its structure consisting of two to three beta-sheets in the N-terminal region of the protein and one alpha-helix in its C-terminal region. As for the FET/TET family of proteins, the amino acid sequence of the zinc-finger motif in them differs significantly from the classic consensus motif (Cys-$X_{2–4}$-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His) [83]. It should be noticed here that the amino acid sequence of this motif in proteins FUS and EWS is highly similar, which is preserved in all organisms studied by us. Our plots on the prediction of the structure demonstrated quite well the correspondence of this motif and the predicted structure (**Figure 4**: proteins FUS and EWS, second peak from the bottom, blue). On the contrary, in TAF15, this motif differs somewhat not only from FUS and EWS but also in the organisms studied by us; according to our prediction, it forms no structure (**Figure 4**). It is important that in proteins of the FET/TET family the zinc-finger motif occurs only once, contrary to the classic variant when it occurs as tandem repeats. As a rule, if the zinc-finger motif occurs once and its sequence differs considerably from the canonical one, the functions of this motif can differ remarkably from the classic motif. For example, it can both be bound to RNA and have no relation to the binding of nucleic acids [83]. The removal of this motif together with the terminal part of the FUS molecule did not affect the protein ability to aggregate and have toxicity either in yeasts or in the cell culture of the SH-SY5Y neuroblastoma [67]. Additional studies should be conducted to reveal the functions of the zinc-finger motif in proteins of the FET/TET family.

6. Conclusion: Repeats are a general characteristics of prion-like domains

In this study, we have analyzed RNA-binding proteins with prion-like domains, such as TAF15, FUS, and EWS. Using the FoldAmyloid program [77], we revealed the existence of amyloidogenic regions in the RRM domain in all three proteins of the FET family. This allows us to suggest that when the binding to RNA is violated, the proteins can aggregate spontaneously forming amyloid fibrils. In this case, protein FUS can aggregate both upon the removal of RRM and at point mutations in this domain [65, 66]. Further studies should be conducted to clarify whether aggregation of protein FUS in the presence of RRM does not lead to the

formation of irreversible pathologic aggregations and functions only by the mechanism of assembly and possible disassembly under favorable conditions analogous to the functioning of SGs. We have also found that in all members of the FET/TET family the zinc-finger motif does not correspond to the classical one because it concerns both the location of tandem repeats and the amino acid sequence. Moreover, according to our prediction, in the FUS and TAF15 proteins, it does not even form any structure. Based on this, we can propose that in proteins of the FET/TET family the zinc-finger motif is not responsible for the DNA binding, but performs other functions determination of which requires further investigations.

We have also predicted unstructured regions corresponding both to prion-like domains and to additional regions, where we have revealed several recurrent amino acid motifs—tandem repeats. In addition, these proteins are characterized by the occurrence of homorepeats when one amino acid is multiply recurrent. The length of homorepeats can affect not only the capacity to aggregate but also the toxicity in an aggregated state. Correspondingly, the larger is the number of repeats, the higher is the probability that the aggregated protein containing them is associated with the development of the disease. We have established that protein FUS contains five disordered patterns from the pattern library obtained from the protein data bank: GSHM, GGGGSGG, GGGGG, GGSGGGGSGGG, and RGGGGSG, which are not located in the prion-like domain predicted for FUS. For other members of this family, these patterns are not characteristic and, as known, are less toxic when compared to FUS. When a larger part of the prion-like domain in protein FUS was removed (100 residues of the 165), it retained its capacity to aggregate [65], but lost completely the ability to form gel and display toxicity. These results allow us to suggest that the presence of short repeats in the unstructured prion-like domain of RNA-binding proteins is required for fast formation of a dynamic cross-beta structure of SGs [84].

Acknowledgements

I thank T.B. Kuvshinkina, E.I. Leonova, I.V. Sokolovsky, and N.V. Dovidchenko for assistance in the chapter preparation. This study was supported by the Russian Science Foundation No. 14-14-00536.

Author details

Oxana V. Galzitskaya

Address all correspondence to: ogalzit@vega.protres.ru

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia

References

[1] Söderberg L, Dahlqvist C, Kakuyama H, Thyberg J, Ito A, Winblad B, et al. Collagenous Alzheimer amyloid plaque component assembles amyloid fibrils into protease resistant aggregates. FEBS J. 2005;272:2231–6.

[2] Dobson CM. Protein misfolding, evolution and disease. Trends Biochem. Sci. 1999;24:329–32.

[3] Rochet JC, Lansbury PT. Amyloid fibrillogenesis: themes and variations. Curr. Opin. Struct. Biol. 2000;10:60–8.

[4] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000;41:415–27.

[5] Tompa P. Intrinsically unstructured proteins. Trends Biochem. Sci. 2002;27:527–33.

[6] Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J. Mol. Biol. 2004;342:345–53.

[7] Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. Structure 2006. 14:1667–76.

[8] Kovacs E, Tompa P, Liliom K, Kalmar L. Dual coding in alternative reading frames correlates with intrinsic protein disorder. Proc. Natl. Acad. Sci. U. S. A. 2010;107:5429–34.

[9] Tenidis K, Waldner M, Bernhagen J, Fischle W, Bergmann M, Weber M, et al. Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. J. Mol. Biol. 2000;295:1055–71.

[10] von Bergen M, Friedhoff P, Biernat J, Heberle J, Mandelkow EM, Mandelkow E. Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming beta structure. Proc. Natl. Acad. Sci. U. S. A. 2000;97:5129–34.

[11] Ivanova MI, Sawaya MR, Gingery M, Attinger A, Eisenberg D. An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril. Proc. Natl. Acad. Sci. U. S. A. 2004;101:10584–9.

[12] Huntley MA, Golding GB. Simple sequences are rare in the Protein Data Bank. Proteins. 2002;48:134–40.

[13] Lobanov MY, Galzitskaya OV. Disordered patterns in clustered Protein Data Bank and in eukaryotic and bacterial proteomes. PLoS One. 2011;6:e27142.

[14] Lobanov MY, Sokolovskiy IV, Galzitskaya OV. HRaP: database of occurrence of HomoRepeats and patterns in proteomes. Nucleic Acids Res. 2014;42:D273–8.

[15] Lobanov MY, Sokolovskiy IV, Galzitskaya OV. IsUnstruct: prediction of the residue status to be ordered or disordered in the protein chain by a method based on the Ising model. J. Biomol. Struct. Dyn. 2013;31:1034–43.

[16] Jorda J, Xue B, Uversky VN, Kajava AV. Protein tandem repeats - the more perfect, the less structured. FEBS J. 2010;277:2673–82.

[17] Tanaka M, Chien P, Yonekura K, Weissman JS. Mechanism of cross-species prion transmission: an infectious conformation compatible with two highly divergent yeast prion proteins. Cell. 2005;121:49–62.

[18] Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. Nature. 2005;438:878–81.

[19] Krebs MRH, Morozova-Roche LA, Daniel K, Robinson CV, Dobson CM. Observation of sequence specificity in the seeding of protein amyloid fibrils. Protein Sci. Publ. Protein Soc. 2004;13:1933–8.

[20] Harkey MA, Klueg K, Sheppard P, Raff RA. Structure, expression, and extracellular targeting of PM27, a skeletal protein associated specifically with growth of the sea urchin larval spicule. Dev. Biol. 1995;168:549–66.

[21] Matsushima N, Yoshida H, Kumaki Y, Kamiya M, Tanaka T, Izumi Y, et al. Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. Curr. Protein Pept. Sci. 2008;9:591–610.

[22] Singh GP, Chandra BR, Bhattacharya A, Akhouri RR, Singh SK, Sharma A. Hyper-expansion of asparagines correlates with an abundance of proteins with prion-like domains in Plasmodium falciparum. Mol. Biochem. Parasitol. 2004;137:307–19.

[23] Lindquist S. Regulation of protein synthesis during heat shock. Nature. 1981;293:311–4.

[24] Bosco DA, Lemay N, Ko HK, Zhou H, Burke C, Kwiatkowski TJ, et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Hum. Mol. Genet. 2010;19:4160–75.

[25] Li YR, King OD, Shorter J, Gitler AD. Stress granules as crucibles of ALS pathogenesis. J. Cell Biol. 2013;201:361–72.

[26] Bosco DA, Lemay N, Ko HK, Zhou H, Burke C, Kwiatkowski TJ, et al. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Hum. Mol. Genet. 2010;19:4160–75.

[27] Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J. 2010;29:2841–57.

[28] Liu-Yesucevitz L, Bassell GJ, Gitler AD, Hart AC, Klann E, Richter JD, et al. Local RNA translation at the synapse and in disease. J. Neurosci. Off. J. Soc. Neurosci. 2011;31:16086–93.

[29] Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149:753–67.

[30] Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell. 2009;137:146–58.

[31] King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. 2012;1462:61–80.

[32] Aguzzi A, Rajendran L. The transcellular spread of cytosolic amyloids, prions, and prionoids. Neuron. 2009;64:783–90.

[33] Couthouis J, Hart MP, Shorter J, DeJesus-Hernandez M, Erion R, Oristano R, et al. A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. 2011;108:20881–90.

[34] King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. 2012;1462:61–80.

[35] Tan AY, Manley JL. The TET family of proteins: functions and roles in disease. J. Mol. Cell Biol. 2009;1:82–92.

[36] Zinszner H, Albalat R, Ron D. A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. Genes Dev. 1994;8:2513–26.

[37] Bentmann E, Neumann M, Tahirovic S, Rodde R, Dormann D, Haass C. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). J. Biol. Chem. 2012;287:23079–94.

[38] Lerga A, Hallier M, Delva L, Orvain C, Gallais I, Marie J, et al. Identification of an RNA binding specificity for the potential splicing factor TLS. J. Biol. Chem. 2001;276:6807–16.

[39] Iko Y, Kodama TS, Kasai N, Oyama T, Morita EH, Muto T, et al. Domain architectures and characterization of an RNA-binding protein, TLS. J. Biol. Chem. 2004;279:44834–40.

[40] Gal J, Zhang J, Kwinter DM, Zhai J, Jia H, Jia J, et al. Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. Neurobiol. Aging. 2011;32:27–40.

[41] Ito D, Seki M, Tsunoda Y, Uchiyama H, Suzuki N. Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. Ann. Neurol. 2011;69:152–62.

[42] Kino Y, Washizu C, Aquilanti E, Okuno M, Kurosawa M, Yamada M, et al. Intracellular localization and splicing regulation of FUS/TLS are variably affected by amyotrophic lateral sclerosis-linked mutations. Nucleic Acids Res. 2011;39:2781–98.

[43] Niu C, Zhang J, Gao F, Yang L, Jia M, Zhu H, et al. FUS-NLS/Transportin 1 complex structure provides insights into the nuclear targeting mechanism of FUS and the implications in ALS. PLoS One. 2012;7:e47056.

[44] Ishigaki S, Masuda A, Fujioka Y, Iguchi Y, Katsuno M, Shibata A, et al. Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. Sci. Rep. 2012;2:529.

[45] Rogelj B, Easton LE, Bogu GK, Stanton LW, Rot G, Curk T, et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. Sci. Rep. 2012;2:603.

[46] Fujii R, Okabe S, Urushido T, Inoue K, Yoshimura A, Tachibana T, et al. The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. Curr. Biol. 2005;15:587–93.

[47] Anderson P, Kedersha N. Stress granules: the Tao of RNA triage. Trends Biochem. Sci. 2008;33:141–50.

[48] Doi H, Koyano S, Suzuki Y, Nukina N, Kuroiwa Y. The RNA-binding protein FUS/TLS is a common aggregate-interacting protein in polyglutamine diseases. Neurosci. Res. 2010;66:131–3.

[49] Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science. 2009;323:1205–8.

[50] Vance C, Rogelj B, Hortobágyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009;323:1208–11.

[51] Crozat A, Aman P, Mandahl N, Ron D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature. 1993;363:640–4.

[52] Zhang ZC, Chook YM. Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). Proc. Natl. Acad. Sci. U. S. A. 2012;109:12017–21.

[53] Mitchell JC, McGoldrick P, Vance C, Hortobagyi T, Sreedharan J, Rogelj B, et al. Overexpression of human wild-type FUS causes progressive motor neuron degeneration in an age- and dose-dependent fashion. Acta Neuropathol. (Berl.). 2013;125:273–88.

[54] Shelkovnikova TA, Robinson HK, Connor-Robson N, Buchman VL. Recruitment into stress granules prevents irreversible aggregation of FUS protein mislocalized to the cytoplasm. Cell Cycle Georget. Tex. 2013;12:3194–202.

[55] Bossy-Wetzel E, Schwarzenbacher R, Lipton SA. Molecular pathways to neurodegeneration. Nat. Med. 2004;10 Suppl:S2–9.

[56] Rabin SJ, Kim JMH, Baughn M, Libby RT, Kim YJ, Fan Y, et al. Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. Hum. Mol. Genet. 2010;19:313–28.

[57] Orozco D, Edbauer D. FUS-mediated alternative splicing in the nervous system: consequences for ALS and FTLD. J. Mol. Med (Berl.) 2013;91:1343–54.

[58] Xia R, Liu Y, Yang L, Gal J, Zhu H, Jia J. Motor neuron apoptosis and neuromuscular junction perturbation are prominent features in a Drosophila model of Fus-mediated ALS. Mol. Neurodegener. 2012;7:10.

[59] Vaccaro A, Tauffenberger A, Aggad D, Rouleau G, Drapeau P, Parker JA. Mutant TDP-43 and FUS cause age-dependent paralysis and neurodegeneration in *C. elegans*. PLoS One. 2012;7:e31321.

[60] Belzil VV, Valdmanis PN, Dion PA, Daoud H, Kabashi E, Noreau A, et al. Mutations in FUS cause FALS and SALS in French and French Canadian populations. Neurology. 2009;73:1176–9.

[61] Ju S, Tardiff DF, Han H, Divya K, Zhong Q, Maquat LE, et al. A yeast model of FUS/TLS-dependent cytotoxicity. PLoS Biol. 2011;9:e1001052.

[62] Murakami T, Yang S-P, Xie L, Kawano T, Fu D, Mukai A, et al. ALS mutations in FUS cause neuronal dysfunction and death in *Caenorhabditis elegans* by a dominant gain-of-function mechanism. Hum. Mol. Genet. 2012;21:1–9.

[63] Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell. 2009;137:146–58.

[64] Shorter J, Lindquist S. Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers. Science. 2004;304:1793–7.

[65] Sun Z, Diaz Z, Fang X, Hart MP, Chesi A, Shorter J, et al. Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS Biol. 2011;9:e1000614.

[66] Shelkovnikova TA, Robinson HK, Connor-Robson N, Buchman VL. Recruitment into stress granules prevents irreversible aggregation of FUS protein mislocalized to the cytoplasm. Cell Cycle Georget. Tex. 2013;12:3194–202.

[67] Sun Z, Diaz Z, Fang X, Hart MP, Chesi A, Shorter J, et al. Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS Biol. 2011;9:e1000614.

[68] Shelkovnikova TA, Robinson HK, Connor-Robson N, Buchman VL. Recruitment into stress granules prevents irreversible aggregation of FUS protein mislocalized to the cytoplasm. Cell Cycle Georget. Tex. 2013;12:3194–202.

[69] Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J. 2010;29:2841–57.

[70] Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149:753–67.

[71] Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. Mol. Syst. Biol. 2008;4:179.

[72] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37:D412–6.

[73] Fisette J-F, Montagna DR, Mihailescu M-R, Wolfe MS. A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. J. Neurochem. 2012;121:763–73.

[74] Jorda J, Kajava AV. Protein homorepeats sequences, structures, evolution, and functions. Adv. Protein Chem. Struct. Biol. 2010;79:59–88.

[75] Lobanov MY, Galzitskaya OV. Occurrence of disordered patterns and homorepeats in eukaryotic and bacterial proteomes. Mol. Biosyst. 2012;8:327–37.

[76] Maddelein ML, Wickner RB. Two prion-inducing regions of Ure2p are nonoverlapping. Mol. Cell. Biol. 1999;19:4516–24.

[77] Garbuzynskiy SO, Lobanov MY, Galzitskaya OV. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics. 2010;26:326–32.

[78] Galzitskaya OV. Reversible and irreversible aggregation of proteins from the FET family: influence of repeats in protein chain on its aggregation capacity. Curr. Protein Pept Sci. 2016;17:319-331.

[79] Bedford MT. Arginine methylation at a glance. J. Cell Sci. 2007;120:4243–6.

[80] Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell. 2012;149:768–79.

[81] Belzil VV, Valdmanis PN, Dion PA, Daoud H, Kabashi E, Noreau A, et al. Mutations in FUS cause FALS and SALS in French and French Canadian populations. Neurology. 2009;73:1176–9.

[82] Vaccaro A, Tauffenberger A, Aggad D, Rouleau G, Drapeau P, Parker JA. Mutant TDP-43 and FUS cause age-dependent paralysis and neurodegeneration in C. elegans. PLoS One. 2012;7:e31321.

[83] Lewin B. Genes IX. 9th ed. Sudbury, Mass: Jones and Bartlett Publishers; 2008.

[84] Galzitskaya OV. Repeats are one of the main characteristics of RNA-binding proteins with prion-like domains. Mol. Biosyst. 2015;11:2210–8.

IntechOpen