ORGANELLE

PERSPECTIVE OPEN ACCESS

Intricate roles of spacers and stickers of Arg-rich C9ORF72 dipeptide repeat proteins: from toxicity to targeting to membraneless organelles

Tamami Miyagi[1] and Kohsuke Kanekura[1,*]

[1]Department of Pharmacology, Tokyo Medical University, 6-1-1 Shinjuku, Shinjuku-ku, Tokyo, 160-8402, Japan

*Correspondence to: Kohsuke Kanekura, M.D., Ph.D., Professor, Department of Pharmacology, Tokyo Medical University, 6-1-1, Shinjuku, Shinjuku-ku, Tokyo, 160-8402, Japan. Tel: +81-3-3351-6141, E mail: kanekura@tokyo-med.ac.jp

Accepted October 30, 2023

ONE SENTENCE SUMMARY

We summarize the intricate roles of sticker and spacer amino acids of C9ORF72-dipeptide repeat proteins in the toxicity and targeting to membraneless organelles.

ABSTRACT

C9ORF72, one of the most common genes implicated in amyotrophic lateral sclerosis and frontotemporal dementia, induces neurodegeneration through various pathways. The most notable is interference through liquid-liquid phase separation (LLPS). LLPS is a biophysical phenomenon involved in many fundamental biological processes, such as the formation of membraneless organelles (MLOs), transcription, and nucleocytoplasmic transport. The Arg-rich dipeptide repeat proteins (R-DPRs) produced from the aberrant *C9ORF72* gene are highly charged and are incorporated into the phase-separated MLOs, inhibiting their functions. However, the detailed molecular mechanism remains to be eluci-dated. Recently, we analyzed the structure-function relationship of R-DPRs and clarified the mechanism by which the sticker Arg and the spacer Pro/Gly regulate cytotoxicity and subcellular localization. Natural R-DPRs contribute to the localization of specific MLOs. In this review, we discuss the roles of the sticker and spacer of R-DPRs in the LLPS and how they regulate subcellular localization, protein-protein interaction, and neurotoxicity.

Key words: Amyotrophic lateral sclerosis (ALS); C9ORF72; membraneless organelle (MLO); Arg-rich dipeptide repeat protein; liquid-liquid phase separation (LLPS)

INTRODUCTION

Neurodegenerative diseases are characterized by progressive degeneration of specific types of neurons. Amyotrophic lateral sclerosis (ALS) is the most common motor neuron disease accompanied by the progressive loss of both upper and lower motor neurons. Frontotemporal dementia (FTD) causes cognitive dysfunction and personality disorders. Both diseases are categorized into the ALS-FTD spectrum because of the overlap of causative genes, similarities in pathology, and the existence of patients who develop both clinical manifestations[1]. Mutations in the *C9ORF72* gene account for approximately 40% of familial ALS and 25% of familial FTD cases and are present in approximately 5–10% of patients with sporadic ALS and 5% of patients with sporadic FTD[2–4]. Therefore, *C9ORF72* is one of the most important causative genes in ALS and FTD. All *C9ORF72*-related ALS/FTD (C9-ALS/FTD) patients have the same type of mutations, abnormal expansions of GGGGCC hexanucleotides (hexanucleotide repeat expansion: HRE) in intron 1 of the gene. While normal individuals have < 25 hexanucleotide repeats, C9-ALS/FTD patients have hundreds to thousands of HREs[3,4]. There is some debate on how many repeats are pathogenic, but longer than 200 repeats are considered definitely pathogenic[5].

Various hypotheses have been proposed to explain why HRE can lead to ALS/FTD:

1. Loss of function due to decreased translation of C9ORF72 protein by HRE[6].
2. HRE DNA/transcript toxicity due to the capture of DNA/RNA-binding proteins[7].
3. Neurotoxicity induced by dipeptide repeat proteins (DPRs) translated via repeat-associated non-ATG translation (RAN-T)[8,9].

In this review, we focused on the toxicity of DPRs. Among these C9-DPRs, Arg-containing poly(PR) and poly(GR) (Arg-rich DPRs, R-DPRs) are highly toxic *in vitro* and *in vivo*[8–12]. However, the detailed molecular mechanisms remain to be evaluated. We recently analyzed the structure-function relationships of R-DPRs and revealed the complicated roles of the sticker (Arg) and spacer (Pro/Gly); this is discussed in this review.

MLO: red

C9-R-DPRs	Localization	Affected Organelle
poly(PR)	nucleolus paraspeckle nuclear speckle stress granule	Cajal body stress granule nucleolus nuclear pore heterochromatin ribosome
poly(GR)	stress granule nucleolus mitochondria	stress granule ribosome mitochondria

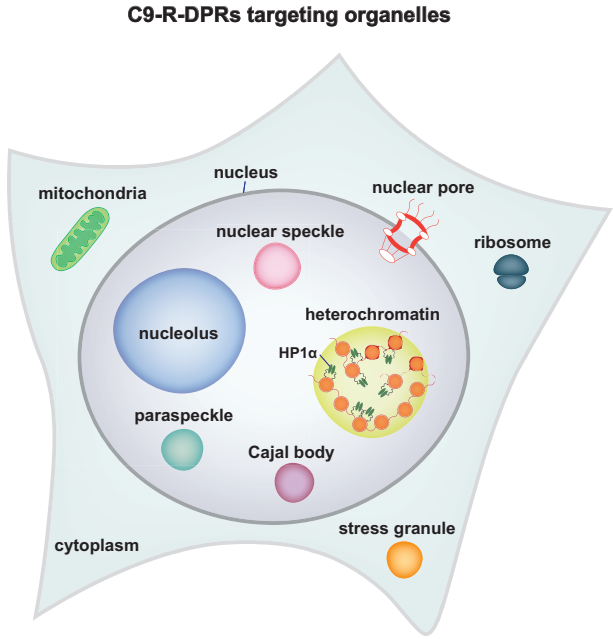

Figure 1. C9-R-DPRs localize to MLOs and impede their function

The left panel shows the subcellular localization of C9-DPRs and the organelles impeded by them, and the right panel shows a schematic of the affected intracellular organelles.

DPRs are produced from *C9ORF72* through RAN translation

Repeat expansions are reported in several neurodegenerative disease-causing genes, such as *Huntingtin* and *Ataxin-2*[13]. In most cases, repeat expansions are located in the introns or untranslated regions (UTR). However, CAG repeats sometimes exist in the coding region, resulting in the translation of the poly-Gln (poly-Q) sequence[14]. In addition to poly-Q translated from the in-frame (CAG) codon, poly-Ala (poly-A) and poly-Ser (poly-S), which correspond to the AGC (+1) and GCA (+2) frames, respectively, have been detected in these poly-Q diseases[15–17]. These were initially believed to be errors in the translation of poly-Q. However, in 2010, a novel translational machinery called repeat-associated non-ATG translation (RAN-T) was identified[18], revealing that these homopolymeric proteins are produced from repeat sequences in CAG triplet diseases via RAN-T[19]. The precise mechanism by which RAN-T proceeds independent of the start codon and is enhanced by the length of the repeats remains to be elucidated. The RAN-T of the *C9ORF72* repeat expansion depends on the phosphorylation level of eIF2α by double-stranded RNA-dependent protein kinase (PKR)[20] and is enhanced by the integrated stress response[21,22]. Since the discovery of RAN-T, various homopolymeric proteins have been shown to be translated via RAN-T, including *C9ORF72*[23].

Five DPRs, poly(Gly-Ala: GA), poly(Pro-Arg: PR), poly(Gly-Arg: GR), poly(Gly-Pro: GP), and poly(Pro-Ala: PA), are translated from the sense (GGGGCC)n and antisense (CCCCGG)n of the *C9ORF72* HRE. Mori et al. detected these C9-DPRs in the central nervous system of C9-ALS/FTD patients[23,24]. The tissue distribution profile of poly(GR), in particular, is consistent with the affected lesions in C9-ALS/FTD[25], and poly(GP) in the cerebrospinal fluid is a potential biomarker[26] that could be used in the evaluation of therapeutics in clinical trials[27,28]. RAN-T from elongated repeat sequences has been thought to occur only under pathological conditions; however, poly(Val-Arg: VR) and poly(Gly-Leu: GL) are translated from mammalian telomeric RNA (TTAGGG)n[29]. This suggests that RAN-T plays an important role in the production of unidentified homopolymeric proteins from endogenous microsatellites and recurrent repeat expansion detected in cancerous tissues[30].

C9-R-DPRs are toxic *in vitro* and *in vivo*

Although the physiological significance of C9-DPRs is unknown, Kwon et al. reported that synthetic $(PR)_{20}$ and $(GR)_{20}$ peptides penetrate the plasma membrane, accumulate in the nucleolus, inhibit RNA metabolism, and induce cell death[9]. Mizielinska et al. demonstrated that poly(PR) and poly(GR) of *C9ORF72*-derived DPRs induce retinal degeneration in a drosophila model[8]. Wen et al. examined the toxicity of these five DPRs and showed that R-DPRs, especially poly(PR), were highly toxic *in vitro* and *in vivo* using primary cortical neurons and a drosophila model[10]. In addition to R-DPRs, poly(GA) exerts modest toxicity *in vitro*[31] and *in vivo*[32] via the formation of detergent-insoluble aggregates[32], inhibition of proteasome[33], and DNA damage through the sequestration of phosphorylated ATM[31].

Interactome analysis showed that R-DPRs bind to RNA and RNA-binding proteins and that $(PR)_{20}$ and $(GR)_{20}$ peptides inhibit protein translation, using an *in vitro* translation assay and neuronal cells[34]. Lee et al. expressed GFP-$(PR)_{50}$ and GFP-$(GR)_{50}$ in HeLa cells and confirmed that both inhibited protein translation[35]. Zhang et al. found that poly(GR) impairs protein translation and stress granule dynamics in the brain of a poly(GR) transgenic mouse model[11]. Thus, the inhibition of protein translation is an important part of the cytotoxic mechanism of R-DPRs. However, the molecular mechanisms underlying the suppression of protein translation by R-DPR remain unclear. We previously demonstrated that R-DPRs suppress RNA transcription[36] and RNA processing[37]. R-DPRs are also reported to clog the ribosome tunnel[38] and induce ribosome stalling[39]. In addition to the inhibition of protein translation, R-DPRs exhibit a variety of toxic

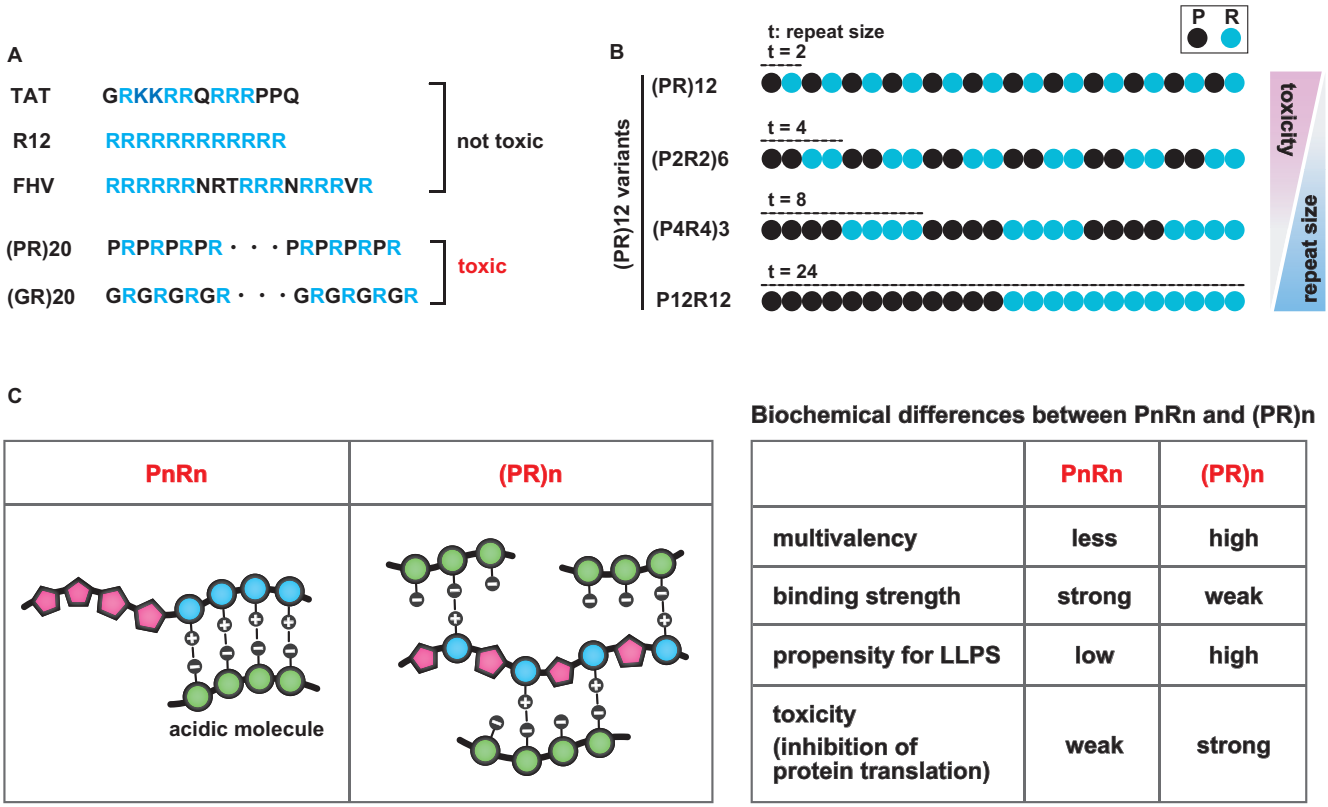

Figure 2. Toxicity of R-DPR depends on the alternate distribution of Arg

(A) Sequences of Arg-rich cell-penetrating peptides. Blue letters represent basic amino acids. Their toxicity was examined in NSC34 cells[52].

(B) Structures of $(PR)_{12}$ variants with different sizes of repeats (repeat size: $t = 2, 4, 8, 24$).

(C) The left panel is a schematic of molecular interaction between PnRn or (PR)n and acidic molecules. The right panel shows the biochemical features of PnRn and (PR)n.

effects on various processes, including axonal transport[40], RNA splicing[9], RNA editing[41], and mitochondrial respiratory chain complex function and ATP production[42,43]; in addition, they promote the aggregation of TAR DNA-binding protein (TDP-43)[44] and impair membraneless organelle (MLO) dynamics[35]. Owing to the positive charge of Arg, poly(PR) and poly(GR) are readily incorporated into MLOs that are rich in negatively charged nucleic acids, interfering with their function. The functions of a variety of MLOs are disrupted by C9-R-DPRs. Examples include the inhibition of phase separation of the heterochromatin component, HP1α[12], inhibition of nucleolus and stress granule dynamics[35,45], induction of nucleolar stress and activation of p53[46], failure of Cajal bodies,[35] and impaired nucleocytoplasmic transport due to disruption of the nuclear pore complex[47,48] (Fig. 1). This vast variety of toxic effects raises a fundamental question: "Why are poly(PR) and poly(GR) so toxic?". However, the molecular mechanisms underlying the toxicity of poly(PR) and poly(GR) have not yet been determined.

The distribution of Arg controls intermolecular interactions with acidic molecules and liquid-liquid phase separation and determines the toxicity of poly(PR)

The inhibition of protein translation by $(PR)_{20}$ and $(GR)_{20}$ suggests this could be a general property of Arg-rich peptides. Arg-rich peptides are studied as drug delivery systems (DDS) owing to their membrane permeability; not all Arg-rich peptides are toxic[49]. For example, the TAT peptide (GRKKRRQRRRPPQ) derived from the human immunodeficiency virus penetrates the plasma membrane without exhibiting toxicity; it is called a cell-penetrating peptide (CPP)[50]. All Arg-rich CPPs, including the TAT peptide, R_{12} peptide, Flock house virus peptide (FHV: RRRRRRNRTRRRNRRRVR)[51], and the $(PR)_{20}$ peptide, penetrate the plasma membrane[52]. However, among these CPPs, only $(PR)_{20}$ inhibits protein translation and decreases cell viability, indicating that not all Arg-rich peptides inhibit protein translation and that poly(PR) and poly(GR) are more likely to inhibit protein translation by obtaining a toxic property (Fig. 2A). The influence of the length of poly(PR) in suppressing protein translation was examined; the longer the poly(PR), the higher the toxicity. It was also found that $(PR)_{12}$, which contains 12 Arg residues, inhibits protein translation, whereas R_{12}, with exactly the same number of Arg residues, does not[36,52]. This suggests that $(PR)_{12}$ acquired the properties necessary to inhibit protein translation through the alternate insertion of Pro into R_{12}.

To elucidate the molecular mechanism, $(PR)_{12}$ mutants with different repeat sizes were synthesized, and their effects on protein translation were investigated[36] (Fig. 2B). $(PR)_{12}$ suppressed protein translation; however, the inhibitory effect weakened as the repeat size increased, and no protein translation suppression was observed with $P_{12}R_{12}$ (Fig. 2B). These results indicate that the distribution of Arg plays an important role in determining poly(PR) toxicity and that alternate insertion of Pro into contiguous Arg is important for the inhibition of protein translation. Although differences in their interactome may determine whether protein translation is inhibited, the abundant positive charge derived from Arg is the driving force

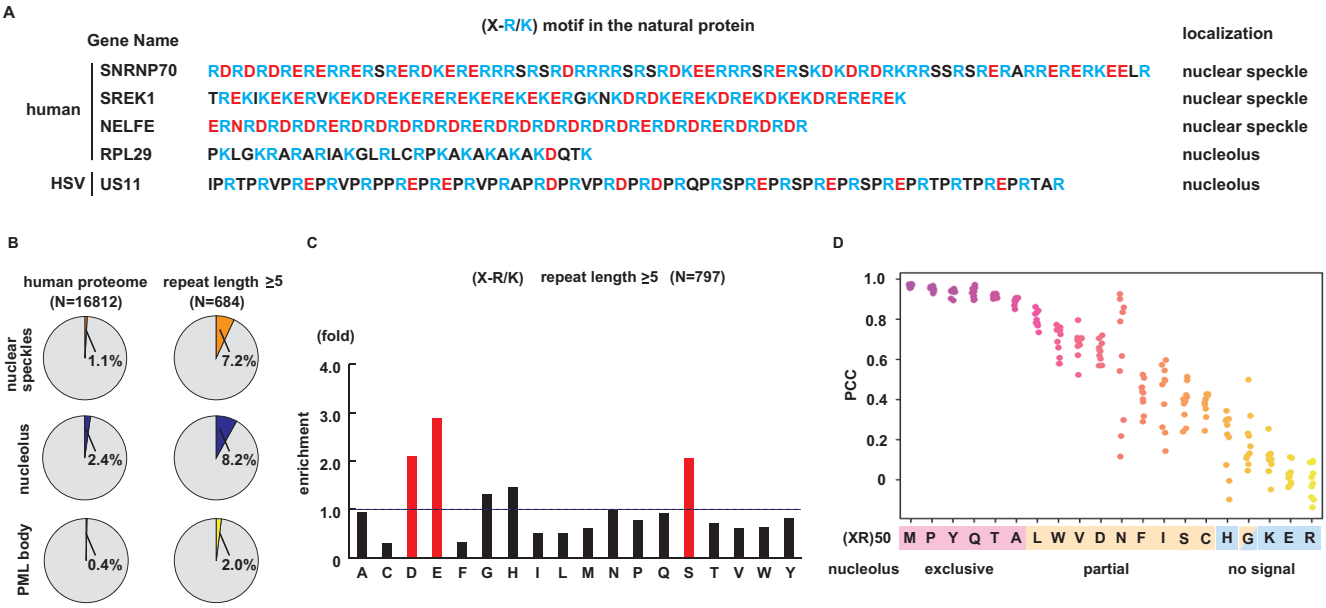

Figure 3. (X-R/K) repeats function as MLO-targeting signals

(A) Examples of (X-R/K) motif in human proteome. Red letters represent acidic amino acids and blue letters represent basic amino acids, respectively.

(B) Proteins containing five or longer (X-R/K) repeats are highly enriched in MLO proteome.

(C) Amino acid occurrence in five or longer (X-R/K) repeats in the human proteome.

(D) The Pearson correlation coefficient (PCC) of GFP-$(XR)_{50}$ and a nucleolar marker NPM1 (10 images/each)[37].

behind the protein-protein interactions, suggesting that the $(PR)_{12}$ interactome is qualitatively similar to the R_{12} interactome. Quantitative proteomic analysis identified approximately 2000 interacting proteins for each peptide, and the $(PR)_{12}$ interactome and R_{12} interactome were qualitatively similar, as expected. However, the insertion of Pro dramatically altered the signals of some interacting proteins, increasing the signal intensity up to thousands of fold compared to that of R_{12}. This quantitative increase in the interactome reflects the enhancement of multivalent protein-protein interactions, indicating that a single $(PR)_{12}$ peptide can quantitatively capture more interacting proteins than R_{12}. To elucidate this mechanism, we characterized the enriched interactome in the $(PR)_{12}$ group and found that it contained more acidic amino acids, Glu and Asp, and that more enriched proteins tended to have longer acidic stretches. The positively charged poly(PR) and the large number of negatively charged acidic amino acids in the interactome led us to speculate that poly(PR) and acidic proteins attract each other via electrostatic forces, resulting in liquid liquid phase sepration (LLPS). When $(PR)_{12}$ and the acidic peptide poly-E were mixed, $(PR)_{12}$ and poly E underwent LLPS, but $P_{12}R_{12}$ did not, indicating that LLPS was promoted by the alternating distribution of Arg charges. Interestingly, the fluorescence recovery after photobleaching (FRAP) showed that the FRAP recovery rate was higher for $(PR)_{12}$ droplets with alternating Arg distributions; greater periodicity resulted in lower FRAP recovery rates[36]. Therefore, the binding energies of $(PR)_{12}$ with alternating Arg distribution to acidic molecules were lower than those of $P_{12}R_{12}$ with a continuous Arg distribution. This was experimentally confirmed, and it was concluded that the alternating distribution of Arg is not advantageous for the binding energy but is advantageous for the formation of multivalent interactions. Although 1:1 binding is weak, multivalent binding promotes phase separation, which could allow $(PR)_{12}$ to incorporate more essential proteins into phase-separated droplets and suppress these functions. Therefore, the distribution of the sticker Arg is a critical determinant of the manner of LLPS with acidic molecules[36] (Fig. 2C).

Next, the roles of spacer amino acids were investigated[37]. There was a significant difference between the inhibitory effects of $(PR)_{12}$ and $(GR)_{12}$ on protein translation[52]. $(PR)_{12}$ strongly inhibits protein translation, whereas $(GR)_{12}$ does not. When the repeat length is increased, $(PR)_{20}$ suppresses protein translation more strongly, whereas $(GR)_{20}$ suppresses protein translation mildly[34,52]. We focused on quantitative differences between the interactomes of poly(PR) and poly(GR). The $(PR)_{12}$ and $(GR)_{12}$ interactomes are qualitatively similar but quantitatively very different[36]. In particular, the signal intensities of acidic proteins were hundreds of times higher in the $(PR)_{12}$ interactome than that in the $(GR)_{12}$ interactome. The interactome of $(GR)_{12}$ qualitatively and quantitatively resembled that of R_{12}. This difference in the interactome is caused by the different roles of the spacer amino acids, Gly and Pro. Gly has a very small hydrogen atom as its side chain and a small steric hindrance, while Pro has a five-membered pyrrolidine ring with greater steric hindrance and is structurally rigid. To investigate how these factors enhance the poly(PR) interactome, we synthesized $(XR)_{12}$ peptides (where X represents all amino acids) and performed immunoprecipitation analyses. Among the various spacer amino acids, only Pro promoted highly multivalent interactions, independent of the size of the side chain and hydrophobicity. Restriction of the free movement of Arg by Pro resulted in Arg being unable to align well with the contiguous acidic sequence. Consequently, Arg binds weakly to a large number of molecules. When the flexible structure of Gly is used as a spacer, the movement of Arg is not restricted, the alignment with the contiguous acidic sequence is well established, and poly(GR) binds tightly to fewer molecules. Thus, we conclude that the spacer amino acid determines the

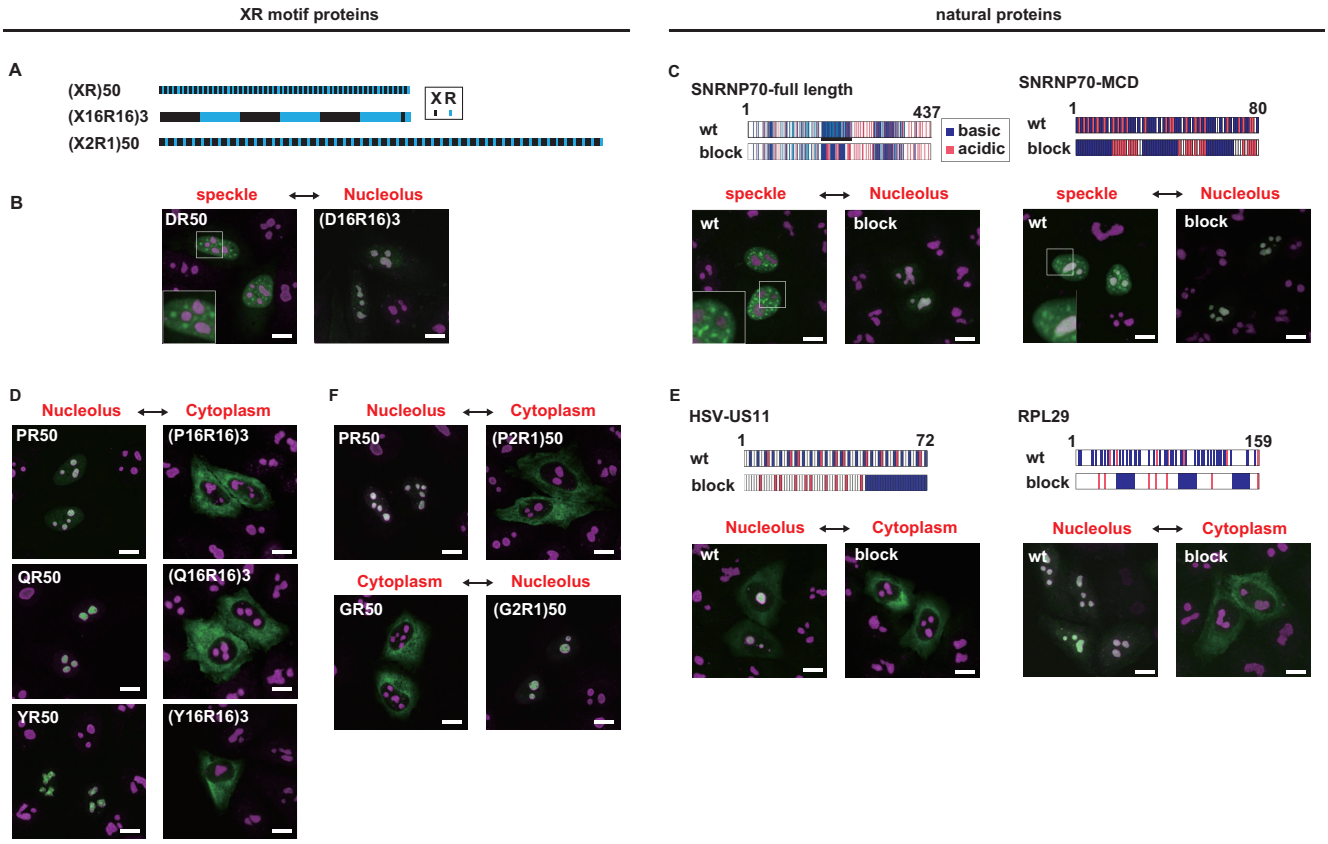

Figure 4. The charge distribution controls the protein targeting to specific MLOs

(A) Schematic diagrams of $(XR)_{50}$ mutants. X_2R_2 was added to the C-terminus of $(X_{16}R_{16})_3$ to adjust the number of Arg to 50.

(B) Subcellular localization of the GFP-$(DR)_{50}$ mutants (indicated in green) in HeLa cells. Nucleoli were visualized with CoraLite555-conjugated anti-NPM1 antibody (indicated in magenta). Scale bar: 10 μm.

(C) (Left) Subcellular localization of full-length human SNRNP70 harboring wild-type (wt) R-MCD or blocky R-MCD (block). The black line indicates the position of R-MCD. (Right) Subcellular localization of wt-SNRNP70-MCD and blocky SNRNP70-MCD. Nucleoli were visualized using CoraLite555-conjugated anti-NPM1 antibody (indicated in magenta).

(D) Subcellular localization of GFP-$(PR)_{50}$, GFP-$(QR)_{50}$, and GFP-$(YR)_{50}$ and their blocky mutants in HeLa cells. Nucleoli were visualized using CoraLite555-conjugated anti-NPM1 antibody (indicated in magenta).

(E) Subcellular localization of GFP-US11-wt and GFP-US11-block (left), and GFP-RPL29-wt and GFP-RPL29-block (right). Nucleoli were visualized using CoraLite555-conjugated anti-NPM1 antibody (indicated in magenta).

(F) Subcellular localization of proportional mutants of GFP-$(PR)_{50}$ and $(GR)_{50}$ in HeLa cells. Nucleoli were visualized using CoraLite555-conjugated anti-NPM1 antibody (indicated in magenta).

binding energy and multivalency of binding by controlling the movement of the sticker Arg[37].

Natural Arg-rich motifs promote protein-targeting to MLOs

MLOs are organelles that do not contain lipid membranes. Their main components are proteins and nucleic acids, particularly RNA-binding proteins and RNA[53]. The most important difference between a classical membranous organelle and an MLO is whether substances can move freely between the boundaries of the organelle. This is due to the presence or absence of a lipid membrane, and the mobility is controlled by interactions between the MLO components. Charged proteins can easily associate with MLOs because of the ease of intermolecular interactions[54]. Proteins containing GGR or GR motifs, often found in RNA-binding proteins, easily migrate to MLOs such as nucleolus[55].

To elucidate the role of natural Arg-rich sequences in the distribution of MLOs, the subcellular localization of proteins containing natural poly(X-R/K) repeats was analyzed[37]. Data mining from the UniProt database revealed that proteins harboring five or more (X-R/K) repeats tended to migrate to MLOs, such as the nucleolus, nuclear speckle, and nuclear promyelocytic leukemia (PML) bodies involved in the replication of DNA and cellular senescence[56], suggesting that sequences with alternating Arg function as a targeting signal for MLO (Fig. 3A and 3B). Amino acid occurrence in the natural (X-R/K) repeats showed that Asp, Glu, and Ser were highly enriched in sequences containing more than five repeats (Fig. 3C). To determine the role of the spacer amino acid of poly(XR) in MLO targeting, we generated $(XR)_{50}$ mutants (where X represents all amino acids) and investigated their subcellular localization. Most of the $(XR)_{50}$ mutants are localized to the nucleolus as well as $(PR)_{50}$ (Fig. 3D).

Poly(XR) tends to migrate to the nucleolus. What would happen if an acidic spacer was inserted to neutralize the Arg charge? Greig et al. expressed the Asp/Arg repeat-enriched fungal SPA-5 protein in mammalian cells and found that it localized to the nuclear speckle, prompting them to focus on the poly(RD) sequence[57]. The poly(RD) sequence has alternate acidic Asp and basic Arg, and its

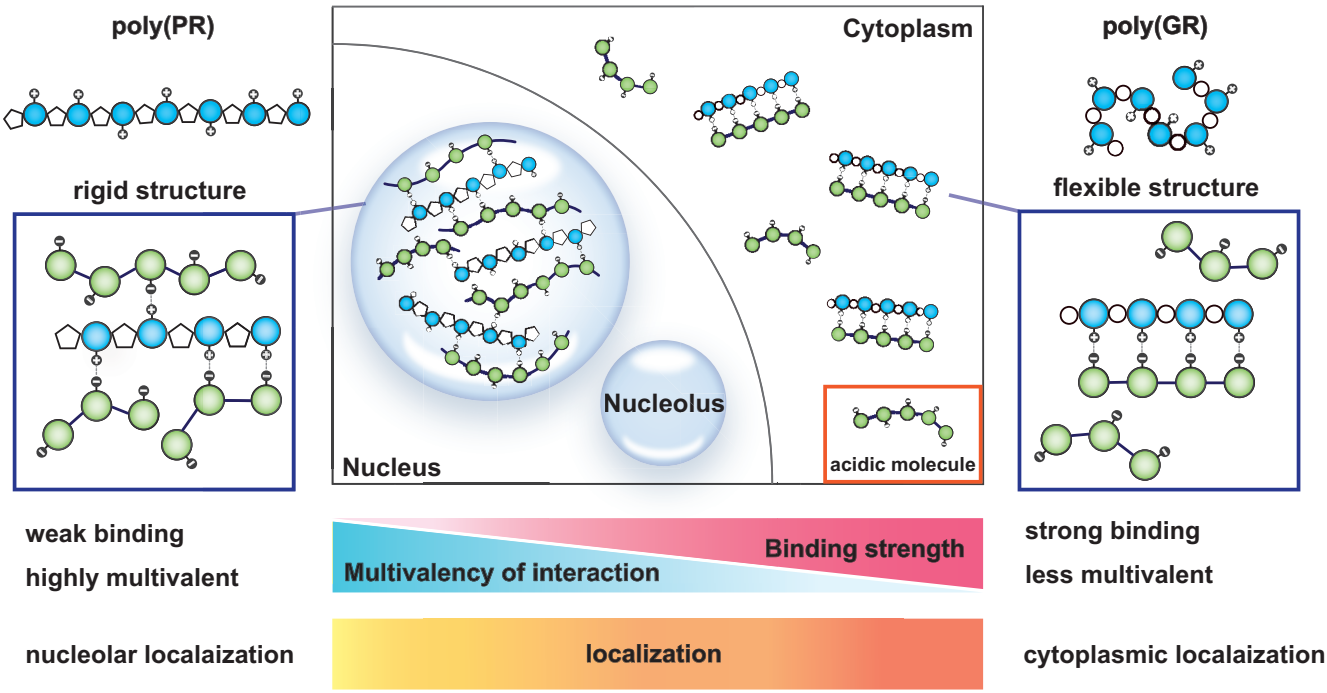

Figure 5. Differential toxicity and localization of poly(PR) and poly(GR) based on the sticker and spacer model.

overall charge is neutral. Poly(RD) accumulates in the nuclear speckle rather than in the nucleolus, and natural proteins containing poly(RD) migrate to the nuclear speckle (Fig. 3A). Interestingly, poly(ER), a sequence with alternate of Glu and Arg, which is another acidic amino acid, did not localize to the nuclear speckle, nor did poly(KD), a sequence with alternate Asp and Lys, which is a basic amino acid like Arg. It remains to be elucidated why Glu has a different effect than Asp. The difference between Arg and Lys is likely due to the characteristics of Arg. The guanidinium ion of Arg harbors three planar nitrogen groups that allow the simultaneous formation of electrostatic, pi-pi, and cation-pi interactions. Poly (RS)-containing alternate Ser, which can be phosphorylated in the cell, migrates to the nuclear speckles as poly(DR), and poly(SR) sequences are significantly enriched in nuclear speckle proteins. The reason for the accumulation of poly (DR) in nuclear speckles remains unknown. We created mutants of $(DR)_{50}$ with different charge distributions without changing the total number of Asp and Arg residues, and analyzed the localization of these mutants[58]. The $(DR)_{50}$ mutant, with a blocky charge distribution, shifted its localization from nuclear speckles to the nucleolus (Fig. 4A and 4B). This phenomenon was confirmed in natural nuclear speckle-associated proteins such as the small nuclear ribonucleoprotein U1 subunit 70 (SNRNP70-full length). Wild-type (wt) full-length SNRNP70 and SNRNP70-MCD (mixed-charge domain) were localized to nuclear speckles, but SNRNP70 with a blocky charge shifted its localization from nuclear speckles to the nucleolus (Fig. 4C). The $(DR)_{12}$ synthetic peptide did not phase-separate with RNA or polyE, whereas the $D_{12}R_{12}$ peptide phase-separated with RNA, polyE, and the nucleolar protein, NPM1. Both the nucleolus and nuclear speckles contain RNA; however, as indicated by the strong signal detected when intracellular RNA is stained, the nucleolus contains much more RNA than nuclear speckles, and the electrostatic force derived from the abundant RNA could strongly attract successive Arg. $(DR)_{50}$ could escape electrostatic attraction from RNA because Asp and Arg cancel each other's charges. However, the mechanism by which $(DR)_{50}$ specifically accumulates in the nucleus remains unclear. The zwitterionic structure may favor specific interactions with SON and SRRM2, which are scaffold proteins of nuclear speckles[59] or other components of the nuclear speckle. Differences in the intra-organelle environment, such as the hydrophobicity and density of aromatic compounds, could also have an effect.

When $(PR)_{50}$ and $(GR)_{50}$ were expressed in HeLa cells or motor neuronal NSC34 cells, $(PR)_{50}$ was almost exclusively confined to the nucleolus, whereas $(GR)_{50}$ localized mainly to the cytoplasm and nucleolus, although exclusive cytoplasmic localization was often observed[37]. The presence of clustered basic amino acids is important for nucleolar localization[60–62]; however, $(GR)_{50}$, despite having the same number of Arg residues as $(PR)_{50}$, localizes mainly to the cytoplasm rather than to the nucleolus, suggesting the existence of a mechanism other than the presence of basic amino acids that regulates its localization to the nucleolus.

We created a $(PR)_{50}$ mutant with the same positive charge but different charge distribution and examined its subcellular localization. Despite having the same net charge as $(PR)_{50}$, $(P_{16}R_{16})_3$ was exclusively localized in the cytoplasm (Fig. 4A and 4D). This phenomenon was also observed for $(QR)_{50}$ and $(YR)_{50}$, which were localized to the nucleolus, as well as for $(PR)_{50}$ (Fig. 4D).

We tested whether this rule could be applied to natural proteins. The US11 protein of herpes simplex virus (HSV) type 1 has 24 repeats of the (X-P-R) sequence and localizes to the nucleolus[63], but when we changed the distribution of Arg, the US11 mutant localized to the cytoplasm (Fig. 4E). Human ribosomal protein L29 (RPL29) contains repetitive basic amino acids and localizes to the nucleolus and cytoplasm[64]. However, by changing the distribution of basic amino acids without altering the net charge, the RPL29

mutant was localized to the cytoplasm (Fig. 4E). Therefore, we hypothesized that the spacer separates the charge of the consecutive Arg residues, thereby interfering with strong binding between successive Arg residues and cytoplasmic molecules, allowing Arg to migrate to the nucleolus. In the case of Gly, the small steric hindrance allows Arg to move freely; therefore, Gly cannot separate the Arg charge sufficiently, resulting in the cytoplasmic localization of $(GR)_{50}$. If there are two Gly residues between Arg residues, they separate the Arg charge well, and $(G_2R_1)_{50}$ localizes to the nucleolus[37] (Fig. 4A and 4F).

Sticker and spacer model for phase separation

Arg acts as a sticker, whereas Pro and Gly act as spacers. Arg in a contiguous arrangement can form a good alignment with acidic molecules, and its binding strength is strong, but less multivalent. Gly has a small side chain and causes less steric hindrance to the neighboring amino acids. Owing to this high binding energy, poly (GR) binds strongly to cytoplasmic molecules immediately after translation in the ribosome and cannot migrate to the nucleolus[37]. However, when Pro is inserted, its rigid nature inhibits alignment with nearby contiguous acidic sequences and each Arg could have more opportunities to bind to different acidic molecules, resulting in the formation of weak multivalent interactions. Poly(PR) could interact with more molecules via these multivalent interactions. Poly(PR) may be more likely to be involved in acquiring strong toxicity through this multivalent interaction by supplementing more molecules into the droplet. In other words, the balance between the sticker and spacer controls the multivalency and binding energy of the protein-protein interaction (Fig. 5). Furthermore, the spacer controls the distance between neighboring stickers and the degree of separation of the sticker Arg, thereby regulating the localization of R-DPRs to the nucleolus (Fig. 5). However, these results were obtained using relatively short R-DPRs, whose characteristics could differ from the physiological characteristics of R-DPRs with hundreds to thousands of repeats observed in ALS pathology[65]. In the case of R-DPRs with relatively short repeats (20- or 50-repeat), poly (PR) was more capable of forming multivalent protein interactions than poly(GR). However, in R-DPRs with >1000 repeats, the number of Arg residues is sufficiently large that the interaction could be as multivalent as that in poly(PR). Strong binding by poly(GR) could induce higher toxicity; further studies are warranted to elucidate this.

CONCLUSION

Several neurodegenerative disease-related proteins undergo LLPS; therefore, it is believed to play a critical role in the pathogenesis of neurodegenerative diseases. To understand LLPS, the sticker and spacer model was proposed[66]. It is clear that the sticker and spacer determine whether LLPS occurs; in addition, their position and type modulate protein-protein interactions and subcellular localization, eventually controlling the toxicity pathways. We hope that these studies on LLPS will help identify a cure for ALS.

AUTHOR CONTRIBUTIONS

T.M. and K.K. designed the structure of the article and wrote the manuscript.

FUNDING

This work was supported in part by grants from JSPS KAKENHI (grant numbers 20H03593 and 21H04634). This work was also supported in part by the Takeda Science Foundation, the Mochida Memorial Foundation for Medical and Pharmaceutical Research, and the Astellas Foundation for Research on Metabolic Disorders. This work was also sponsored by the ALS Foundation and the Japan ALS Association.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

We thank Editage for English language editing.

REFERENCES

1. Abramzon, Y.A., Fratta, P., Traynor, B.J., and Chia, R. (2020). The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. Front Neurosci. *14*, 42. doi:10.3389/fnins.2020.00042.
2. Brown, R.H., Jr., and Al-Chalabi, A. (2017). Amyotrophic Lateral Sclerosis. N Engl J Med. *377*, 1602. doi:10.1056/NEJMc1710379.
3. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron. *72*, 245–256. doi:10.1016/j.neuron.2011.09.011.
4. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. *72*, 257–268. doi:10.1016/j.neuron.2011.09.010.
5. van der Ende, E.L., Jackson, J.L., White, A., Seelaar, H., van Blitterswijk, M., and Van Swieten, J.C. (2021). Unravelling the clinical spectrum and the role of repeat length in C9ORF72 repeat expansions. J Neurol Neurosurg Psychiatry. *92*, 502–509. doi:10.1136/jnnp-2020-325377.
6. Shi, Y., Lin, S., Staats, K.A., Li, Y., Chang, W.-H., Hung, S.-T., Hendricks, E., Linares, G.R., Wang, Y., Son, E.Y., et al. (2018). Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. Nat Med. *24*, 313–325. doi:10.1038/nm.4490.
7. Donnelly, C.J., Zhang, P.W., Pham, J.T., Heusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., et al. (2013). RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron. *80*, 415–428. doi:10.1016/j.neuron.2013.10.015.
8. Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O.C., Pietrzyk, J., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science *345*, 1192–1194. doi:10.1126/science.1256800.
9. Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science *345*, 1139–1145. doi:10.1126/science.1254917.
10. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S.S., Shi, Y., Lin, S., Shneider, N.A., Monaghan, J., Pandey, U.B., et al. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate invitro and invivo neuronal death. Neuron. doi:10.1016/j.neuron.2014.12.010.
11. Zhang, Y.J., Gendron, T.F., Ebbert, M.T.W., O'Raw, A.D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C.N., et al. (2018). Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. Nat Med. *24*, 1136–1142. doi:10.1038/s41591-018-0071-1.
12. Zhang, Y.J., Guo, L., Gonzales, P.K., Gendron, T.F., Wu, Y., Jansen-West, K., O'Raw, A.D., Pickles, S.R., Prudencio, M., Carlomagno, Y., et al. (2019). Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. Science. *363*. doi:10.1126/science.aav2606.
13. Fujino, Y., and Nagai, Y. (2022). The molecular pathogenesis of repeat expansion diseases. Biochem Soc Trans. *50*, 119–134. doi:10.1042/BST20200143.

14. Costa, M.D., and Maciel, P. (2022). Modifier pathways in polyglutamine (PolyQ) diseases: from genetic screens to drug targets. Cell Mol Life Sci. *79*, 274. doi:10.1007/s00018-022-04280-8.

15. Davies, J.E., and Rubinsztein, D.C. (2006). Polyalanine and polyserine frameshift products in Huntington's disease. J Med Genet. *43*, 893–896. doi:10.1136/jmg.2006.044222.

16. Toulouse, A., Au-Yeung, F., Gaspar, C., Roussel, J., Dion, P., and Rouleau, G.A. (2005). Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. Hum Mol Genet. *14*, 2649–2660. doi:10.1093/hmg/ddi299.

17. Gaspar, C., Jannatipour, M., Dion, P., Laganiere, J., Sequeiros, J., Brais, B., and Rouleau, G.A. (2000). CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. Hum Mol Genet. *9*, 1957–1966. doi:10.1093/hmg/9.13.1957.

18. Zu, T., Gibbens, B., Doty, N.S., Gomes-pereira, M., Huguet, A., and Stone, M.D. (2010). Non-ATG – initiated translation directed by microsatellite expansions. Pnas. *108*, 260–265. doi:10.1073/pnas.1013343108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1013343108. doi:10.1073/pnas.1013343108.

19. Bañez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., Perez, B.A., Tusi, S.K., Pletnikova, O., Borchelt, D.R., Ross, C.A., Margolis, R.L., et al. (2015). RAN Translation in Huntington Disease. Neuron. *88*, 667–677. doi:10.1016/j.neuron.2015.10.038.

20. Zu, T., Guo, S., Bardhi, O., Ryskamp, D.A., Li, J., Khoramian Tusi, S., Engelbrecht, A., Klippel, K., Chakrabarty, P., Nguyen, L., et al. (2020). Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice. Proc Natl Acad Sci U S A. *117*, 18591–18599. doi:10.1073/pnas.2005748117.

21. Green, K.M., Glineburg, M.R., Kearse, M.G., Flores, B.N., Linsalata, A.E., Fedak, S.J., Goldstrohm, A.C., Barmada, S.J., and Todd, P.K. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nat Commun. *8*, 2005. doi:10.1038/s41467-017-02200-0.

22. Cheng, W., Wang, S., Mestre, A.A., Fu, C., Makarem, A., Xian, F., Hayes, L.R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., et al. (2018). C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. Nat Commun. *9*, 51. doi:10.1038/s41467-017-02495-z.

23. Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C., et al. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science. *339*, 1335–1338. doi:10.1126/science.1232927.

24. Mori, K., Arzberger, T., Grasser, F.A., Gijselinck, I., May, S., Rentzsch, K., Weng, S.M., Schludi, M.H., van der Zee, J., Cruts, M., et al. (2013). Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol. *126*, 881–893. doi:10.1007/s00401-013-1189-3.

25. Quaegebeur, A., Glaria, I., Lashley, T., and Isaacs, A.M. (2020). Soluble and insoluble dipeptide repeat protein measurements in C9orf72-frontotemporal dementia brains show regional differential solubility and correlation of poly-GR with clinical severity. Acta Neuropathol Commun. *8*, 184. doi:10.1186/s40478-020-01036-y.

26. Lehmer, C., Oeckl, P., Weishaupt, J.H., Volk, A.E., Diehl-Schmid, J., Schroeter, M.L., Lauer, M., Kornhuber, J., Levin, J., Fassbender, K., et al. (2017). Poly-GP in cerebrospinal fluid links C9orf72-associated dipeptide repeat expression to the asymptomatic phase of ALS/FTD. EMBO Mol Med. *9*, 859–868. doi:10.15252/emmm.201607486.

27. Querin, G., Biferi, M.G., and Pradat, P.F. (2022). Biomarkers for C9orf7-ALS in Symptomatic and Pre-symptomatic Patients: State-of-the-art in the New Era of Clinical Trials. J Neuromuscul Dis. *9*, 25–37. doi:10.3233/JND-210754.

28. Krishnan, G., Raitcheva, D., Bartlett, D., Prudencio, M., McKenna-Yasek, D.M., Douthwright, C., Oskarsson, B.E., Ladha, S., King, O.D., Barmada, S.J., et al. (2022). Poly(GR) and poly(GA) in cerebrospinal fluid as potential biomarkers for C9ORF72-ALS/FTD. Nat Commun. *13*, 2799. doi:10.1038/s41467-022-30387-4.

29. Al-Turki, T.M., and Griffith, J.D. (2023). Mammalian telomeric RNA (TERRA) can be translated to produce valine-arginine and glycine-leucine dipeptide repeat proteins. Proc Natl Acad Sci U S A. *120*, e2221529120. doi:10.1073/pnas.2221529120.

30. Erwin, G.S., Gursoy, G., Al-Abri, R., Suriyaprakash, A., Dolzhenko, E., Zhu, K., Hoerner, C.R., White, S.M., Ramirez, L., Vadlakonda, A., et al. (2023). Recurrent repeat expansions in human cancer genomes. Nature. *613*, 96–102. doi:10.1038/s41586-022-05515-1.

31. Nihei, Y., Mori, K., Werner, G., Arzberger, T., Zhou, Q., Khosravi, B., Japtok, J., Hermann, A., Sommacal, A., Weber, M., et al. (2020). Poly-glycine-alanine exacerbates C9orf72 repeat expansion-mediated DNA damage via sequestration of phosphorylated ATM and loss of nuclear hnRNPA3. Acta Neuropathol. *139*, 99–118. doi:10.1007/s00401-019-02082-0.

32. Zhang, Y.J., Gendron, T.F., Grima, J.C., Sasaguri, H., Jansen-West, K., Xu, Y.F., Katzman, R.B., Gass, J., Murray, M.E., Shinohara, M., et al. (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. Nat Neurosci. *19*, 668–677. doi:10.1038/nn.4272.

33. Guo, Q., Lehmer, C., Martinez-Sanchez, A., Rudack, T., Beck, F., Hartmann, H., Perez-Berlanga, M., Frottin, F., Hipp, M.S., Hartl, F.U., et al. (2018). In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. Cell 172., 696–705 e612. doi:10.1016/j.cell.2017.12.030.

34. Kanekura, K., Yagi, T., Cammack, A.J., Mahadevan, J., Kuroda, M., Harms, M.B., Miller, T.M., and Urano, F. (2016). Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. Hum Mol Genet. *25*, 1803–1813. doi:10.1093/hmg/ddw052.

35. Lee, K.-H., Zhang, P., Kim, H.J., Mitrea, D.M., Sarkar, M., Freibaum, B.D., Cika, J., Coughlin, M., Messing, J., Molliex, A., et al. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell. *167*, 774–788.e717. doi:10.1016/j.cell.2016.10.002.

36. Chen, C., Yamanaka, Y., Ueda, K., Li, P., Miyagi, T., Harada, Y., Tezuka, S., Narumi, S., Sugimoto, M., Kuroda, M., et al. (2021). Phase separation and toxicity of C9orf72 poly(PR) depends on alternate distribution of arginine. J Cell Biol. *220*. doi:10.1083/jcb.202103160.

37. Miyagi, T., Ueda, K., Sugimoto, M., Yagi, T., Ito, D., Yamazaki, R., Narumi, S., Hayamizu, Y., Uji, I.H., Kuroda, M., and Kanekura, K. (2023). Differential toxicity and localization of arginine-rich C9ORF72 dipeptide repeat proteins depend on declustering of positive charges. iScience. *26*, 106957. doi:10.1016/j.isci.2023.106957.

38. Loveland, A.B., Svidritskiy, E., Susorov, D., Lee, S., Park, A., Zvornicanin, S., Demo, G., Gao, F.B., and Korostelev, A.A. (2022). Ribosome inhibition by C9ORF72-ALS/FTD-associated poly-PR and poly-GR proteins revealed by cryo-EM. Nat Commun. *13*, 2776. doi:10.1038/s41467-022-30418-0.

39. Kriachkov, V., Ormsby, A.R., Kusnadi, E.P., McWilliam, H.E.G., Mintern, J.D., Amarasinghe, S.L., Ritchie, M.E., Furic, L., and Hatters, D.M. (2023). Arginine-rich C9ORF72 ALS proteins stall ribosomes in a manner distinct from a canonical ribosome-associated quality control substrate. J Biol Chem. *299*, 102774. doi:10.1016/j.jbc.2022.102774.

40. Fumagalli, L., Young, F.L., Boeynaems, S., De Decker, M., Mehta, A.R., Swijsen, A., Fazal, R., Guo, W., Moisse, M., Beckers, J., et al. (2021). C9orf72-derived arginine-containing dipeptide repeats associate with axonal transport machinery and impede microtubule-based motility. Sci Adv. *7*. doi:10.1126/sciadv.abg3013.

41. Suzuki, H., and Matsuoka, M. (2021). Proline-arginine poly-dipeptide encoded by the C9orf72 repeat expansion inhibits adenosine deaminase acting on RNA. J Neurochem. *158*, 753–765. doi:10.1111/jnc.15445.

42. Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Karydas, A., Tran, H., Yang, D., Petrucelli, L., Miller, B.L., Almeida, S., and Gao, F.-B. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron. *92*, 383–391. doi:10.1016/j.neuron.2016.09.015.

43. Li, S., Wu, Z., Li, Y., Tantray, I., De Stefani, D., Mattarei, A., Krishnan, G., Gao, F.B., Vogel, H., and Lu, B. (2020). Altered MICOS Morphology and Mitochondrial Ion Homeostasis Contribute to Poly(GR) Toxicity Associated with C9-ALS/FTD. Cell Rep. *32*, 107989. doi:10.1016/j.celrep.2020.107989.

44. Cook, C.N., Wu, Y., Odeh, H.M., Gendron, T.F., Jansen-West, K., Del Rosso, G., Yue, M., Jiang, P., Gomes, E., Tong, J., et al. (2020). C9orf72 poly(GR) aggregation induces TDP-43 proteinopathy. Sci Transl Med. *12*. doi:10.1126/scitranslmed.abb3774.

45. Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V.H., et al. (2017). Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. Mol Cell. *65*, 1044–1055 e1045. doi:10.1016/j.molcel.2017.02.013.

46. Maor-Nof, M., Shipony, Z., Lopez-Gonzalez, R., Nakayama, L., Zhang, Y.J., Couthouis, J., Blum, J.A., Castruita, P.A., Linares, G.R., Ruan, K., et al. (2021). p53 is a central regulator driving neurodegeneration caused by C9orf72 poly(PR). Cell. *184*, 689–708 e620. doi:10.1016/j.cell.2020.12.025.

47. Hayes, L.R., Duan, L., Bowen, K., Kalab, P., and Rothstein, J.D. (2020). C9orf72 arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. Elife. *9*. doi:10.7554/eLife.51685.

48. Jovičič, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S.B., Paul, J.W., Sun, S., Herdy, J.R., Bieri, G., et al. (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci. doi:10.1038/nn.4085.

49. Bechara, C., and Sagan, S. (2013). Cell-penetrating peptides: 20 years later, where do we stand? FEBS Lett. *587*, 1693–1702. doi:10.1016/j.febslet.2013.04.031.

50. Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K., and Sugiura, Y. (2001). Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. J Biol Chem. *276*, 5836–5840. doi:10.1074/jbc.M007540200.

51. Nakase, I., Hirose, H., Tanaka, G., Tadokoro, A., Kobayashi, S., Takeuchi, T., and Futaki, S. (2009). Cell-surface Accumulation of Flock House Virus-derived Peptide Leads to Efficient Internalization via Macropinocytosis. Mol Ther. *17*, 1868–1876. doi:10.1038/mt.2009.192.

52. Kanekura, K., Harada, Y., Fujimoto, M., Yagi, T., Hayamizu, Y., Nagaoka, K., and Kuroda, M. (2018). Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. Sci Rep. *8*, 12740. doi:10.1038/s41598-018-31096-z.

53. Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. *28*, 420–435. doi:10.1016/j.tcb.2018.02.004.

54. Boeynaems, S., Holehouse, A.S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P.S., Pappu, R.V., and Gitler, A. D. (2019). Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. Proceedings of the National Academy of Sciences. *116*, 7889–7898. doi:10.1073/pnas.1821038116.

55. Chong, P.A., Vernon, R.M., and Forman-Kay, J.D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. J Mol Biol. *430*, 4650–4665. doi:10.1016/j.jmb.2018.06.014.

56. Lallemand-Breitenbach, V., and de The, H. (2010). PML nuclear bodies. Cold Spring Harb Perspect Biol. *2*, a000661. doi:10.1101/cshperspect.a000661.

57. Greig, J.A., Nguyen, T.A., Lee, M., Holehouse, A.S., Posey, A.E., Pappu, R.V., and Jedd, G. (2020). Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. Mol Cell. *77*, 1237–1250 e1234. doi:10.1016/j.molcel.2020.01.025.

58. Miyagi, T., Yamazaki, R., Ueda, K., Narumi, S., Hayamizu, Y., Uji, I.H., Kuroda, M., and Kanekura, K. (2022). The Patterning and Proportion of Charged Residues in the Arginine-Rich Mixed-Charge Domain Determine the Membrane-Less Organelle Targeted by the Protein. Int J Mol Sci. *23*. doi:10.3390/ijms23147658.

59. Ilik, I.A., Malszycki, M., Lubke, A.K., Schade, C., Meierhofer, D., and Aktas, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. Elife. *9*, e60579. doi:10.7554/eLife.60579.

60. Scott, M.S., Troshin, P.V., and Barton, G.J. (2011). NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. BMC Bioinformatics. *12*, 317. doi:10.1186/1471-2105-12-317.

61. Martin, R.M., Ter-Avetisyan, G., Herce, H.D., Ludwig, A.K., Lattig-Tunnemann, G., and Cardoso, M.C. (2015). Principles of protein targeting to the nucleolus. Nucleus. *6*, 314–325. doi:10.1080/19491034.2015.1079680.

62. Musinova, Y.R., Kananykhina, E.Y., Potashnikova, D.M., Lisitsyna, O.M., and Sheval, E.V. (2015). A charge-dependent mechanism is responsible for the dynamic accumulation of proteins inside nucleoli. Biochim Biophys Acta. *1853*, 101–110. doi:10.1016/j.bbamcr.2014.10.007.

63. Xing, J., Wu, F., Pan, W., and Zheng, C. (2010). Molecular anatomy of subcellular localization of HSV-1 tegument protein US11 in living cells. Virus Res. *153*, 71–81. doi:10.1016/j.virusres.2010.07.009.

64. Hamidi, T., Singh, A.K., Veland, N., Vemulapalli, V., Chen, J., Hardikar, S., Bao, J., Fry, C.J., Yang, V., Lee, K.A., et al. (2018). Identification of Rpl29 as a major substrate of the lysine methyltransferase Set7/9. J Biol Chem. *293*, 12770–12780. doi:10.1074/jbc.RA118.002890.

65. Callister, J.B., Ryan, S., Sim, J., Rollinson, S., and Pickering-Brown, S.M. (2016). Modelling C9orf72 dipeptide repeat proteins of a physiologically relevant size. Hum Mol Genet. doi:10.1093/hmg/ddw327.

66. Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 367., 694–699. doi:10.1126/science.aaw8653.