

# The return of the rings: evolutionary role of aromatic residues in liquid-liquid phase separation

Wen-Lin Ho<sup>1</sup> and Jie-rong Huang<sup>1,2,3,\*</sup>

<sup>1</sup>*Institute of Biochemistry and Molecular Biology, National Yang Ming Chiao Tung University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan*

<sup>2</sup>*Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan*

<sup>3</sup>*Department of Life Sciences and Institute of Genome Sciences, National Yang Ming Chiao Tung University, No. 155 Section 2, Li-nong Street, Taipei, Taiwan*

To whom correspondence should be addressed: [jierongh@nycu.edu.tw](mailto:jierongh@nycu.edu.tw)

Keywords: intrinsically disordered proteins, liquid-liquid phase separation, RNA-binding proteins, biomolecular condensates, membraneless organelle

## **Abstract**

Aromatic residues appeared relatively late in the evolution of protein sequences. They stabilize the hydrophobic core of globular proteins and are typically absent from intrinsically disordered regions (IDRs). However, recent advances in protein liquid-liquid phase separation (LLPS) studies have shown that aromatic residues in IDRs often act as important “stickers”, promoting multivalent interactions and the formation of higher-order oligomers. To reconcile this apparent contradiction, we compared levels of sequence disorder in RNA binding proteins and the human proteome and found that aromatic residues appear more frequently than expected in the IDRs of RNA binding proteins, which are often found to undergo LLPS. Phylogenetic analysis shows that aromatic residues are highly conserved among chordates, highlighting their importance in LLPS-driven functional assembly. These results suggest therefore that aromatic residues have contributed twice to evolution: in stabilizing structured proteins and in the assembly of biomolecular condensates.

Amino acids with aromatic rings stabilize protein structure (Burley and Petsko 1985) and are accordingly rare in intrinsically disordered regions (IDRs)(Dunker, et al. 2001; Yan, et al. 2020). Interestingly however, aromatic residues in IDRs have recently been identified as being crucial in mediating liquid-liquid phase separation (LLPS) (Kwon, et al. 2013; Lin, et al. 2017; Li, Chiang, et al. 2018; Wang, et al. 2018). Here, through protein sequence analyses, we outline how aromatic residues may have contributed twice to evolution.

How did proteins emerge on Earth? Using ferredoxin as an example, Eck and Dayhoff (1996), proposed that early polypeptides were short, with simple compositions, but grew constantly by duplicating their sequences (Eck and Dayhoff 1966). New amino-acids appeared by mutation and the addition of cysteine provided sulfide bonding to ferrous sulfide, a catalyst used as a primitive energy source (Eck and Dayhoff 1966). The more lately incorporated amino acids, including hydrophobic and aromatic residues, enhanced the folding stability of these polypeptides (Burley and Petsko 1985). These polypeptides, or proteins, became the main workhorses of early cells. When lifeforms became more complex with the appearance of eukaryotes and multicellular organisms, proteins became more efficient and “moonlighted” in different functions (Jeffery 1999). The acquisition of IDRs substantially enlarged their functional repertoire (Oldfield and Dunker 2014). The flexibility of IDRs confers many functional advantages to the attached folded domains, notably by increasing the likelihood of interactions with a binding partner (Sugase, et al. 2007) or other molecules (Kriwacki, et al. 1996), in connecting multiple domains (Huang, et al. 2014), and by making the protein more accessible for posttranslational modifications (Li, et al. 2008). Posttranslational modifications alter the physicochemical properties of IDRs (e.g., phosphorylation adds negative charges) and proteins’ propensity to self-assemble (Monahan, et al. 2017; Lu, et al. 2018; Guo, et al. 2019; Saito, et al. 2019). The role of

IDRs in tunable and reversible assembly, also known as LLPS, the mechanism that drives the formation of many membraneless organelles (Alberti and Hyman 2021), has only recently been recognized.

The different building blocks of proteins appeared at different stages of evolution. Eck and Dayhoff found for example, that the earliest four amino acids to appear were Ala, Asp, Ser, and Gly (Eck and Dayhoff 1966), and based on multiple criteria, the order of appearance of the amino acids in life is now accepted to be (Trifonov 2004): Gly, Ala, Asp, Val, Pro, Ser, Glu, Leu/Thr, Arg, Ile/Gln/Asn, His, Lys, Cys, Tyr, Met, Trp. Interestingly, as many others have noticed, this suggests that the earliest peptides were disordered, since the most primitive amino acids are not structure-promoting (Zhu, et al. 2016; Kulkarni and Uversky 2018; Katsnelson 2020). These primitive disordered proteins may have initiated life on Earth and acquired aromatic residues later, thereby increasing their folding stability. The initial set of amino acids were implicated again in the evolution of IDRs for various functions. Although aromatic residues appear relatively late in protein evolution, recent studies have shown that they are one of the driving forces of LLPS in the later evolved IDRs (Kwon, et al. 2013; Lin, et al. 2017; Li, Chiang, et al. 2018; Wang, et al. 2018). This apparent contradiction, the appearance of “structure-promoting” amino acids in unstructured regions, raises interesting questions.

We are particularly interested in RNA-binding proteins (RBPs) because evidence is accumulating that the subcellular localization of RNA molecules is mediated by LLPS (Wiedner and Giudice 2021). Posttranscriptional gene control is dependent on RNA molecules, especially mRNA, being in precisely the right place at the right time (Kloc, et al. 2002; Martin and Ephrussi 2009). Accordingly, we analyzed a set of 1,542 RBPs, and a subgroup of 692 mRNA binding proteins (mRBPs) from a census study (Gerstberger, et al. 2014), in comparison with a set of 20,358 human proteins. We

collected protein sequences from the UniProt database (UniProt 2019) and analyzed their level of disorder using the PONDR server's VLXT, VL3, and VSL2 algorithms (Romero, et al. 1997; Radivojac, et al. 2003). We separated residues into ordered and disordered based on the annotation of these algorithms, and calculated the percentage of proteins with sequences of consecutive disordered residues longer than 30, 40, and 50 amino acids (fig. 1A; supplementary table S1, Supplementary Material online). Our analysis for the human proteome is similar to Dunker et al.'s classic study of eukaryotic organisms (Dunker, et al. 2000) and is in agreement with the observation that IDRs are more prevalent in RBPs (Varadi, et al. 2015; Zagrovic, et al. 2018). Our results also show that mRBPs have an even higher proportion of disordered residues (fig. 1A), with ~10 percentage points (pp) more residues than human proteins in consecutive sequences of 20 or more disordered residues, and ~20 pp more residues found in disordered regions longer than 60 amino acids (supplementary table S1, Supplementary Material online), suggesting that the mRBPs tend to have longer IDRs.

Next, we counted the numbers of each amino-acid type in the predicted disordered or folded domains (supplementary tables S2–S4, Supplementary Material online). For each amino acid, we calculated the log-odds ratio relative to the human proteome of occurring in a disordered region in RBPs or mRBPs (fig. 1B):

$$\text{Log-odds ratio of presence in IDRs} = \ln \left( \frac{\frac{\text{RBP}^{\text{disordered}}}{\text{RBP}^{\text{total}}}}{\frac{\text{Human}^{\text{disordered}}}{\text{Human}^{\text{total}}}} \right) \quad (1)$$

The values obtained are positive for all amino acids (except for cysteine with the VL3 and VSL2 algorithms; fig. 1B) and the average values (dashed lines in fig. 1B) confirm the above conclusion that IDRs are more prevalent in RBPs than in the human proteome. We also repeated the analysis 1,000 times for random selections of 1,542 or 692 proteins sequences (the numbers of RBPs and mRPBP sequences considered) from the human proteome. The averaged log-odds ratios obtained for the random selections are

around zero with standard deviations of mostly less than 0.05 and largest deviations no greater than  $\pm 0.1$  (dots in fig. 1B; supplementary table S5–S7 and figure S1, Supplementary Material online). This analysis of random selections indicates that the differences for RBPs and mRBPs (bars in fig. 1B) are significant.

The results from the different algorithms for each amino acid are consistent, except for cysteine. VL3 and VSL2 predict that the prevalence of cysteines in IDRs is lower in RBPs and mRBPs than in the human proteome, whereas VLXT predicts the opposite, but similar to the range of results of random selection. Although this may simply reflect the algorithm's use of different scoring functions, it is also possible that IDRs in human proteins have a higher portion of cysteines than those in RBPs do. Note also that asparagine is slightly more likely to be disordered in RBPs and mRBPs, whereas for glutamine the difference is similar to those obtained for the randomly selected pools (fig. 1B). Although it has been reported that the IDRs in RBPs are likely to be prion-like, i.e. rich in asparagine and glutamine (Wang, et al. 2018), these two amino acids are not obviously more prevalent than in the human proteome. Among all amino acids, the largest differences obtained with these three algorithms are for phenylalanine, tryptophan, and tyrosine, with values much higher than those obtained for random samples of human proteins, particularly for mRNA targeting proteins (fig. 1B). In other words, *structure-promoting* aromatic amino acids are relatively more abundant in *disordered* regions of RBPs than in the human proteome in general.

Functionally important residues are conserved during evolution (Capra and Singh 2007). We therefore aligned orthologue sequences for each RBP and calculated the Shannon sequence entropy of each residue (Sander and Schneider 1991) as a measure of sequence conservation (ranging from 1 for fully conserved to 0 for not conserved at all; fig. 2). The level of conservation is lower in the chordate phylum (dark blue lines in fig. 2) than in mammals (light blue lines). The analysis was repeated (gray lines) for

vertebrates (subphylum) and tetrapods (superclass). As expected, folded domains are conserved earlier in evolution than disordered regions as indicated by the dark blue lines being in the regions underlined with a red bar (sequences of more than 40 consecutive disordered residues). We also calculated the average level of conservation for each amino acid. A general increasing trend down the taxonomic ranks is expected. For structured domains, the trends are generally flat, indicating early conservation (examples are shown in figure S2, Supplementary Material online). For disordered regions on the contrary, increases are observed for most amino acids; however, the curves for the aromatic residues in many RBPs are flat, as observed for residues in structured domains (bottom panels in fig 2 and figure S2, Supplementary Material online).

In TDP-43 for example, one of the most extensively studied RBP undergoing LLPS (Conicella, et al. 2016; McGurk, et al. 2018; Babinchak, et al. 2019; Sun, et al. 2019), the aromatic residues in its IDRs are highly conserved (fig. 2A; the orange/purple/yellow circles on the Chordata line respectively indicate Phe/Trp/Tyr). The three tryptophans known to be key residues in driving LLPS (Li, Chiang, et al. 2018) are conserved in all chordates. The phenylalanines are conserved early but to a lesser extent, in keeping with the fact that they contribute less to LLPS (Li, Chiang, et al. 2018). This interpretation that sequence conservation reflects functional importance is further supported by the fact that the transient  $\alpha$ -helical region in TDP-43's IDR (fig. 2A, ~residues 320-340), which is involved in LLPS (Chen, et al. 2016; Conicella, et al. 2016; Li, Chen, et al. 2018) is also conserved.

Many other RBPs have highly conserved aromatic residues, whose functional importance has been demonstrated in several cases (fig. 2B). Tyrosines are known to be involved in the well-studied LLPS of the N-terminal domain of the RBP FUS (Lin, et al. 2017; Qamar, et al. 2018; Wang, et al. 2018) and a recent analysis of mammalian

FUS proteins found the same trend as observed here (Dasmeh and Wagner 2021). Phe-Gly repeats are a common feature of nucleoporins, which regulate nucleocytoplasmic transport in the nuclear pore complex (Milles, et al. 2015; Onischenko, et al. 2017; Hayama, et al. 2018). The nucleoporin NUP153 is also categorized as an RBP because its Phe-Gly repeats mediate mRNA trafficking (Bastos, et al. 1996). Phenylalanines are conserved in NUP153 (fig. 2B) and it is reasonable to suppose that this is also the case in other nucleoporins. The tryptophan-rich region in TNRC6 interacts with Ago2 to promote the phase separation of the miRNA induced silencing complex (Sheu-Gruttaduria and MacRae 2018); all but a few of these tryptophans are highly conserved in chordates (fig. 2B). Other examples include the LLPS of CPEB (Ford, et al. 2019) and HNRNPD (Batlle, et al. 2020). Although the role of aromatic residues has not been explicitly studied in these proteins, their conservation in chordates hints at a possible role in LLPS related functions (figure S3, Supplementary Material online). Many RBPs that have not been reported to undergo LLPS also have conserved aromatic patterns, including RBM19, DDX18, KHDRBS1, ABT1, RSL24D1, SMAD5, and TAF9 (figure S2B and S3, Supplementary Material online). We suggest these conserved aromatic residues in IDRs may be involved in LLPS-related functions.

Intrinsically disordered regions appeared late in proteins and RBPs evolved under the selective pressure to assemble precisely in specific cellular locations. This can be achieved by mimicking prion properties (Patel, et al. 2015) or by having short  $\alpha$ -helical motifs (Conicella, et al. 2016), blocked charged-pattern (Greig, et al. 2020), or coiled-coil domains (Fang, et al. 2019), features that provide multivalency and promote the formation of higher-order oligomers. Aromatic residues, with weak  $\pi$ - $\pi$  or cation- $\pi$  interactions, appear to have been selected in this role, affording the weak, reversible interactions required for biomolecular condensate formation. After contributing a first time in the evolution of folded proteins, the subsequent return of the (aromatic)



rings in RBPs appears to have been crucial to the emergence of liquid-liquid phase separation.

## **Methods**

All human protein sequences were retrieved from UniProt (UniProtKB\_2021\_01, download date: 2021/02/06, 20396 sequences in total). Disorder predictions were performed using the VLXT, VL3, and VSL2 algorithms on the PONDR web server (Romero, et al. 1997; Radivojac, et al. 2003), and analyzed using in-house scripts.

Human RBP orthologues were obtained from the Orthologous Matrix database (Altenhoff, et al. 2021) using the human protein's UniProt ID as input. The orthologues were filtered by taxonomy. The sequences of the orthologues were aligned using the Clustal Omega (v.1.2.4) (Sievers, et al. 2011) module in Biopython (Cock, et al. 2009). The aligned sequences were mapped to the human sequence's residue number and the level of conservation of each residue was quantified by its Shannon entropy (Sander and Schneider 1991).

## **Supplementary Material**

Supplementary data are available at Molecular Biology and Evolution online.

## **Data Availability**

All scripts and data generated or analyzed in this study are available in the repository: [http://github.com/allmwh/the\\_return\\_of\\_the\\_rings](http://github.com/allmwh/the_return_of_the_rings)

## **Acknowledgements**

This research was funded by the Ministry of Science and Technology of Taiwan, grant numbers 109-2113-M-010-003 and 110-2113-M-A49A-504-MY3.

## FIGURE LEGENDS

Fig. 1. Prevalence of disorder and disorder odds ratios relative to the human proteome by amino acid type for RNA-binding proteins. (A) Proportion of proteins with disordered regions longer than 30, 40, or 50 consecutive residues, as predicted using different algorithms, in the human proteome (left column), RNA-binding proteins (middle column), and mRNA binding proteins (right column). (B) Log-odds ratios relative to the human proteome of being in an IDR for amino acids in RBPs (left) and mRBPs (right). The dashed lines indicate the average value over all amino acid types. The dots represent the values obtained for randomly selected (negative control) subsamples (N = 1542 for RBPs and N = 689 for mRBP, the same numbers as considered in the main analysis) of the human proteome.

Fig. 2. Sequence conservation in example proteins with highly conserved aromatic residues. (A) TDP-43, (B) FUS, NUP153, and TNRC6A. (*Upper panels*) Levels of sequence conservation quantified by the Shannon entropy: a value of 1 means that all the residues aligned at a given position are identical. Levels of conservation in chordates (dark blue), vertebrates (gray), tetrapods (gray), and mammals (light blue), plotted versus the corresponding residue number in the human sequence. Predicted disordered regions longer than 40 residues are indicated with red bars. Aromatic residues are labeled on the chordate line: Phe (orange), Trp (purple), Tyr (yellow). (A) The three arrows indicate tryptophans experimentally identified as being crucial to liquid-liquid phase separation; the transient  $\alpha$ -helical region that also contributes to self-assembly is also labeled. (B) The colored horizontal curly brackets indicate highly conserved aromatic-residue-rich regions. (*Lower panels*) Average levels of conservation as a function of decreasing taxonomic rank for amino-acid types in regions predicted to be disordered (indicated by red bars in the upper panel).

## References:

- Alberti S, Hyman AA. 2021. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat Rev Mol Cell Biol* 22:196-213.
- Altenhoff AM, Train CM, Gilbert KJ, Mediratta I, Mendes de Farias T, Moi D, Nevers Y, Radoykova HS, Rossier V, Warwick Vesztröcy A, et al. 2021. OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* 49:D373-D379.
- Babinchak WM, Haider R, Dumm BK, Sarkar P, Surewicz K, Choi JK, Surewicz WK. 2019. The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J Biol Chem* 294:6306-6317.
- Bastos R, Lin A, Enarson M, Burke B. 1996. Targeting and function in mRNA export of

- nuclear pore complex protein Nup153. *J Cell Biol* 134:1141-1156.
- Battle C, Yang P, Coughlin M, Messing J, Pesarrodonna M, Szulc E, Salvatella X, Kim HJ, Taylor JP, Ventura S. 2020. hnRNPD Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. *Cell Rep* 30:1117-1128 e1115.
- Burley SK, Petsko GA. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 229:23-28.
- Capra JA, Singh M. 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics* 23:1875-1882.
- Chen TC, Hsiao CL, Huang SJ, Huang JR. 2016. The Nearest-Neighbor Effect on Random-Coil NMR Chemical Shifts Demonstrated Using a Low-Complexity Amino-Acid Sequence. *Protein Pept Lett* 23:967-975.
- Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422-1423.
- Conicella AE, Zerze GH, Mittal J, Fawzi NL. 2016. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* 24:1537-1549.
- Dasmeh P, Wagner A. 2021. Natural Selection on the Phase-Separation Properties of FUS during 160 My of Mammalian Evolution. *Mol Biol Evol* 38:940-951.
- Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al. 2001. Intrinsically disordered protein. *J Mol Graph Model* 19:26-59.
- Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-171.
- Eck RV, Dayhoff MO. 1966. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* 152:363-366.
- Fang X, Wang L, Ishikawa R, Li Y, Fiedler M, Liu F, Calder G, Rowan B, Weigel D, Li P, et al. 2019. Arabidopsis FLL2 promotes liquid-liquid phase separation of polyadenylation complexes. *Nature* 569:265-269.
- Ford L, Ling E, Kandel ER, Fioriti L. 2019. CPEB3 inhibits translation of mRNA targets by localizing them to P bodies. *Proc Natl Acad Sci U S A* 116:18078-18087.
- Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. *Nat Rev Genet* 15:829-845.
- Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G. 2020. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol Cell* 77:1237-1250 e1234.

- Guo YE, Manteiga JC, Henninger JE, Sabari BR, Dall'Agnese A, Hannett NM, Spille JH, Afeyan LK, Zamudio AV, Shrinivas K, et al. 2019. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572:543-548.
- Hayama R, Sparks S, Hecht LM, Dutta K, Karp JM, Cabana CM, Rout MP, Cowburn D. 2018. Thermodynamic characterization of the multivalent interactions underlying rapid and selective translocation through the nuclear pore complex. *J Biol Chem* 293:4555-4563.
- Huang JR, Warner LR, Sanchez C, Gabel F, Madl T, Mackereth CD, Sattler M, Blackledge M. 2014. Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study. *J Am Chem Soc* 136:7068-7076.
- Jeffery CJ. 1999. Moonlighting proteins. *Trends Biochem Sci* 24:8-11.
- Katsnelson A. 2020. Did Disordered Proteins Help Launch Life on Earth? *ACS Cent Sci* 6:1854-1857.
- Kloc M, Zearfoss NR, Etkin LD. 2002. Mechanisms of subcellular mRNA localization. *Cell* 108:533-544.
- Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* 93:11504-11509.
- Kulkarni P, Uversky VN. 2018. Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *Proteomics* 18:e1800061.
- Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL. 2013. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155:1049-1060.
- Li HR, Chen TC, Hsiao CL, Shi L, Chou CY, Huang JR. 2018. The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43. *Biochim Biophys Acta* 1866:214-223.
- Li HR, Chiang WC, Chou PC, Wang WJ, Huang JR. 2018. TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *J Biol Chem* 293:6090-6098.
- Li P, Martins IR, Amarasinghe GK, Rosen MK. 2008. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nat Struct Mol Biol* 15:613-618.
- Lin Y, Currie SL, Rosen MK. 2017. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J Biol Chem*.
- Lu H, Yu D, Hansen AS, Ganguly S, Liu R, Heckert A, Darzacq X, Zhou Q. 2018. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II.

Nature 558:318-323.

Martin KC, Ephrussi A. 2009. mRNA localization: gene expression in the spatial dimension. *Cell* 136:719-730.

McGurk L, Gomes E, Guo L, Mojsilovic-Petrovic J, Tran V, Kalb RG, Shorter J, Bonini NM. 2018. Poly(ADP-Ribose) Prevents Pathological Phase Separation of TDP-43 by Promoting Liquid Demixing and Stress Granule Localization. *Mol Cell* 71:703-717 e709.

Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, Tyagi S, Clarke J, Shammass SL, Blackledge M, et al. 2015. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* 163:734-745.

Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, O'Meally R, Dignon GL, Conicella AE, Zheng W, et al. 2017. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J* 36:2951-2967.

Oldfield CJ, Dunker AK. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* 83:553-584.

Onischenko E, Tang JH, Andersen KR, Knockenhauer KE, Vallotton P, Derrer CP, Kralt A, Mugler CF, Chan LY, Schwartz TU, et al. 2017. Natively Unfolded FG Repeats Stabilize the Structure of the Nuclear Pore Complex. *Cell* 171:904-917 e919.

Patel A, Lee HO, Jawerth L, Maharana S, Jahnle M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, et al. 2015. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162:1066-1077.

Qamar S, Wang G, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, Phillips EC, Miyashita A, Williams D, Strohl F, et al. 2018. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-pi Interactions. *Cell* 173:720-734 e715.

Radivojac P, Obradovic Z, Brown CJ, Dunker AK. 2003. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*:216-227.

Romero, Obradovic, Dunker K. 1997. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome Inform Ser Workshop Genome Inform* 8:110-124.

Saito M, Hess D, Eglinger J, Fritsch AW, Kreysing M, Weinert BT, Choudhary C, Matthias P. 2019. Acetylation of intrinsically disordered regions regulates phase separation. *Nat Chem Biol* 15:51-61.

Sander C, Schneider R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9:56-68.

Sheu-Gruttadauria J, MacRae IJ. 2018. Phase Transitions in the Assembly and Function of Human miRISC. *Cell* 173:946-957 e916.

- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
- Sugase K, Dyson HJ, Wright PE. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021-1025.
- Sun Y, Medina Cruz A, Hadley KC, Galant NJ, Law R, Vernon RM, Morris VK, Robertson J, Chakrabartty A. 2019. Physiologically Important Electrolytes as Regulators of TDP-43 Aggregation and Droplet-Phase Behavior. *Biochemistry* 58:590-607.
- Trifonov EN. 2004. The triplet code from first principles. *J Biomol Struct Dyn* 22:1-11.
- UniProt C. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506-D515.
- Varadi M, Zsolyomi F, Guharoy M, Tompa P. 2015. Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. *PLoS One* 10:e0139731.
- Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovskiy A, Drechsel D, et al. 2018. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174:688-699 e616.
- Wiedner HJ, Giudice J. 2021. It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat Struct Mol Biol* 28:465-473.
- Yan J, Cheng J, Kurgan L, Uversky VN. 2020. Structural and functional analysis of "non-smelly" proteins. *Cell Mol Life Sci* 77:2423-2440.
- Zagrovic B, Bartonek L, Polyansky AA. 2018. RNA-protein interactions in an unstructured context. *FEBS Lett* 592:2901-2916.
- Zhu H, Sepulveda E, Hartmann MD, Kogenaru M, Ursinus A, Sulz E, Albrecht R, Coles M, Martin J, Lupas AN. 2016. Origin of a folded repeat protein from an intrinsically disordered ancestor. *Elife* 5.

Fig. 1

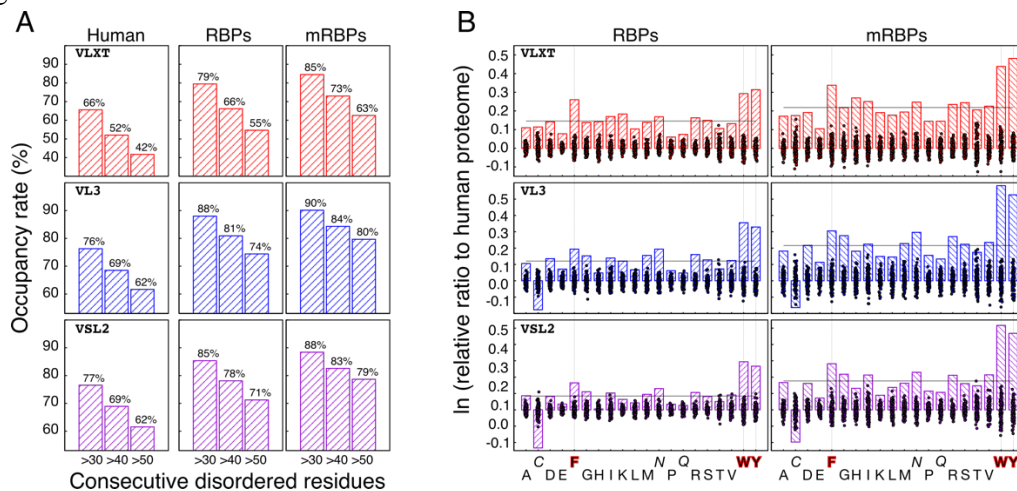


Fig. 2

