REVIEW
Bioinformatics

Proteomics
www.proteomics-journal.com

Intrinsically Disordered Proteome of Human Membrane-Less Organelles

*April L. Darling, Yun Liu, Christopher J. Oldfield, and Vladimir N. Uversky**

It is recognized now that various proteinaceous membrane-less organelles (PMLOs) are commonly found in cytoplasm, nucleus, and mitochondria of various eukaryotic cells (as well as in the chloroplasts of plant cells). Being different from the "traditional" membrane-encapsulated organelles, such as chloroplasts, endoplasmic reticulum, Golgi apparatus, lysosomes, mitochondria, nucleus, and vacuoles, PMLOs solve the cellular need to facilitate and regulate molecular interactions via reversible and controllable isolation of target molecules in specialized compartments. PMLOs possess liquid-like behavior and are believed to be formed as a result of biological liquid-liquid phase transitions (LLPTs, also known as liquid-liquid phase separation), where an intricate interplay between RNA and intrinsically disordered proteins (IDPs) or hybrid proteins containing ordered domains and intrinsically disordered protein regions (IDPRs) may play an important role. This review analyzes the prevalence of intrinsic disorder in proteins associated with various PMLOs found in human cells and considers some of the functional roles of IDPs/IDPRs in biogenesis of these organelles.

1. Diversity of Proteinaceous Membrane-Less Organelles

The intracellular space of a typical eukaryotic cell is crowded and inhomogeneous, containing both the well-known membrane-encapsulated organelles, such as chloroplasts, endoplasmic reticulum, Golgi apparatus, lysosomes, mitochondria, nucleus, and vacuoles, and numerous proteinaceous membrane-less organelles (PMLOs), which were neglected for ages, despite being observed for the first time more than 150 years ago.[1] Such PMLOs are numerous (see **Figure 1**). They are cell-size dependent and highly dynamic, often optically observed as spherical micron-sized droplets,[2] have unique morphologies, and specific distribution patterns within a cell, contain specific sets of resident proteins and typically have RNA, and therefore are commonly known as ribonucleoprotein (RNP) granules/bodies or RNP droplets.[3] Physically, PMLOs are characterized by liquid-like behavior, being able to drip, fuse, wet, and relax to spherical structures upon fusion.[4–7] Although such fluidity is determined by the lack of membrane encapsulation, the structural integrity and biogenesis of PMLOs are supported by dynamic protein–RNA, protein–protein, and protein–DNA interactions.[8] Furthermore, due to the lack of membranes, PMLO are highly dynamic and their interior and components are engaged in direct contacts with the cytoplasm, nucleoplasm, mitochondrial matrix, or stroma of the chloroplasts.[9,10] Biophysical characterization suggested that PMLOs might represent a different liquid state of cytoplasm/nucleoplasm/matrix/stroma, whose major biophysical properties are rather similar to those of these intracellular fluids.[3] In fact, the density of PMLOs is only slightly higher than that of the surrounding intracellular fluids.[11,12] As a result, these organelles are characterized by high internal dynamics, and classified as liquid-droplet phases of the cytoplasm/matrix/nucleoplasm/stroma.[4–7,13,14]

Figure 1 illustrates the diversity of PMLOs found in eukaryotic cells by schematically showing various cytoplasmic, nuclear, mitochondrial, and chloroplast PMLOs. Figure 1 indicates that in cytoplasm of a eukaryotic cell, one can find centrosomes,[15] germline P-granules (germ cell granules or nuage),[4,16] neuronal RNA granules,[17] processing bodies or P-bodies,[18] and stress granules (SGs).[7] Mitochondria and chloroplasts have only one PMLO type, mitochondrial RNA granules,[65] whereas nuclear PMLOs are more numerous and diversified, including Cajal bodies (CBs[19]), chromatin,[20] cleavage bodies,[21] histone locus bodies (HLBs),[22] nuclear gems (gemini of coiled of

A. L. Darling and Dr. V. N. Uversky
Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute
Morsani College of Medicine
University of South Florida
Tampa, FL, USA
E-mail: vuversky@health.usf.edu

Dr. Y. Liu
Guangdong Provincial Key Laboratory for Plant Epigenetics
Shenzhen Key Laboratory of Microbial Genetic Engineering
College of Life Sciences and Oceanography
Shenzhen University
Shenzhen, Guangdong, P. R. China

Dr. C. J. Oldfield
Department of Computer Science
Virginia Commonwealth University
Richmond, VA, USA

Dr. V. N. Uversky
Institute for Biological Instrumentation
Russian Academy of Sciences
Moscow Region, Russia

DOI: 10.1002/pmic.201700193

Significance of the study

Eukaryotic cells contain different proteinaceous membrane-less organelles (PMLOs) that possess liquid-like behavior and are formed as a result of the biological liquid–liquid phase transitions. PMLOs are highly dynamic entities that facilitate and regulate molecular interactions via reversible and controllable isolation of target molecules in specialized compartments. Fluidity of PMLOs is determined by the lack of membrane encapsulation, but their structural integrity and biogenesis are supported by dynamic protein–RNA, protein–protein, and protein–DNA interactions. One of the important features of proteins engaged in the formation of PMLOs is the presence of high levels of intrinsic disorder. We analyze here the prevalence of intrinsic disorder in proteins associated with various PMLOs found in human cells and describe some of the functional roles of intrinsically disordered proteins in biogenesis of these organelles.

CBs),[23,24] nuclear pores,[25] nuclear speckles or interchromatin granule clusters,[26] nuclear stress bodies (nSBs),[27,28] nucleoli,[29] Oct1/PTF/ transcription (OPT) domains,[30] paraspeckles,[31] PcG bodies (polycomb bodies, subnuclear organelles containing polycomb group proteins),[32] perinucleolar compartment (PNC),[33] promyelocytic leukaemia nuclear bodies (PML nuclear bodies) or PML oncogenic domains (PODs),[34] and the Sam68 nuclear bodies (SNBs).[33] Detailed description of these PMLOs and illustrative examples are given elsewhere.[35] PMLOs are many, different, and are present in cytoplasm, chloroplasts, mitochondria, and nucleus of eukaryotic cells. Although these sub-nuclear organelles are diverse, have very different functions, possess rather different morphologies, have divergent cellular distribution, and typically have highly dissimilar composition, they all have something in common, being membrane-less, highly dynamic, and always containing proteins and often including RNA (or, in some cases, DNA). Therefore, assembly/disassembly cycles, dynamics, morphology, and structure of PMLOs are all critically dependent on proteins, which thereby serve as a common denominator. Curiously, contents of only a few PMLOs partially overlap, raising an important question on what define the capability of different proteins located in different cellular regions to regulate the biogenesis of different PMLOs. Importantly, the presence of significant levels of intrinsic disorder in some proteins associated with PMLOs was pointed out,[35–46] indicating that the intrinsic disorder phenomenon might be a part of an answer to that question, and, therefore, it clearly requires careful consideration.

Phase separation can typically take place only when a specific concentration threshold of the macromolecule undergoing liquid–liquid phase transition (LLPT) is reached.[47,48] Many different proteins and nucleic acids can be present at high enough concentrations to promote LLPTs, thereby defining the capability of living cells to simultaneously have several coexisting liquid phases.[47,49–51]

2. Phenomenon of Intrinsic Disorder in Proteins

Recent years witnessed important decoupling of protein functionality from the presence of unique structure. In fact, the lack of specific tertiary structure in many biologically active proteins is associated now with a wide spectrum of crucial functions.[52–58] Such structure-less proteins or domains are known as intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs). Therefore, crudely, the universe of functional proteins can be split into four general categories: globular proteins, fibrous proteins, membrane proteins, and IDPs, with IDPs/IDPRs being very common in various proteomes.[55,59–65] Obviously, many natural proteins are order-disorder hybrids containing ordered domains and IDPRs. There are multiple levels that differentiate such IDPs/IDPRs from their structural

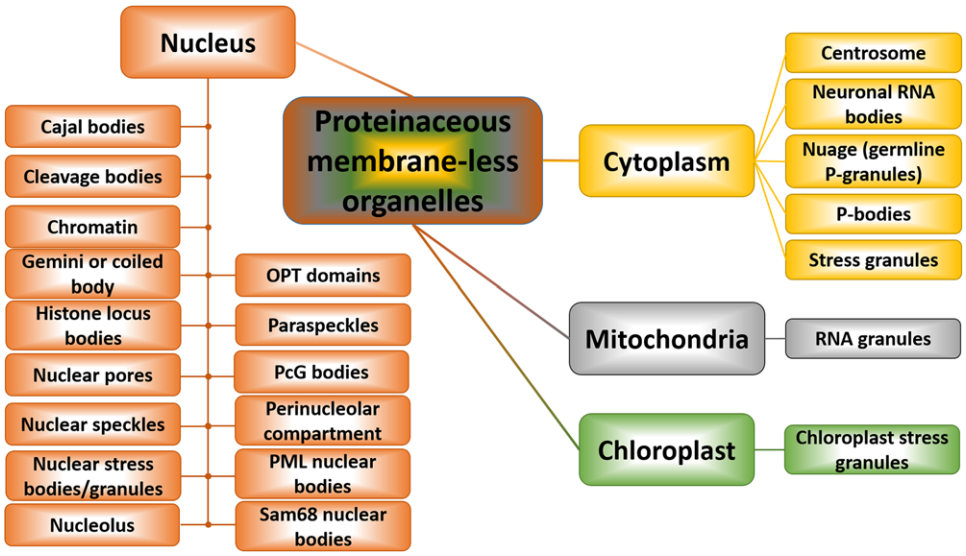


Figure 1. Diversity of PMLOs found in eukaryotic cells. Schematic representation of the multitude of cytoplasmic, nuclear, mitochondrial, and chloroplast PMLOs.

counterparts, ranging from amino acid composition to charge, sequence complexity, hydrophobicity, and flexibility. For example, typical IDPs/IDPRs are critically enriched in several disorder-promoting amino acids (Ala, Arg, Gln, Glu, Gly, Lys, Pro, and Ser), being substantially depleted in Asn, Cys, Ile, Leu, Trp, Tyr, Phe, and Val, which are, therefore, considered as order-promoting residues.[55,66–68]

Although IDPs/IDPRs can be rather accurately predicted based on the specific sequence features,[69] a common criticism of the existing disorder predictions is their bias towards X-ray and NMR-derived experimental input data. However, multiple biophysical methods are currently available for the multifaceted analysis of IDPs/IDPRs,[70,71] and there is a plethora of recently emerging orthogonal experimental approaches to discover, validate, and describe intrinsic disorder both in vivo and in in vitro, and also as a function of space and time.[[72,73]]

The most suitable structural description of an IDP/IDPR involves consideration of the disordered structure as a highly dynamic conformational ensemble containing multiple forms that interconvert on a number of timescales. This defines high structural heterogeneity and plasticity of IDPs/IDPRs, which can be compact or extended, and their various parts can be heterogeneous as well. In other words, such dynamic conformational ensemble representation emphasizes the spatiotemporal heterogeneity of IDPs/IDPRs, where different parts of a molecule are ordered (or disordered) to a different degree, and this distribution is changing with time.[74] Therefore, typical IDP/IDPR is characterized by a mosaic structure that represents a combinations of different foldons (independent foldable units of a protein), inducible foldons (disordered regions that can fold at least in part due to the interaction with binding partners), non-foldons (non-foldable protein regions), semi-foldons (regions which are always in semi-folded state), and unfoldons (regions that have to undergo an order-to-disorder transition to become functional).[74,75]

Indirect support of the functional importance of IDPs/IDPRs follows from their high natural abundance that increases with the increase in the organism complexity.[55,59–65] Also, proteins without unique 3D structures have unique functionality, playing determining roles in control of various signaling pathways, recognition, and regulation.[76–78] Obviously, this functional repertoire complements catalytic and transport functions of ordered proteins,[68,79–81] emphasizing importance of both order and disorder for endless and highly diversified biological activities ascribed to proteins. Just a few illustrative examples of disorder-based or disorder-related advantages include "hubness" of IDPs/IDPRs (i.e., their ability to serve as highly connected nodes in protein–protein interaction networks),[77,82–87] the ability of IDPs/IDPRs to be engaged in specific but weak interactions[88] allowing them to serve as dynamic and sensitive "on-off" switches,[89] to contain molecular recognition features (MoRFs; i.e., disordered regions that fold at interaction with a partner),[90–93] to be promiscuous binders interacting with numerous, often unrelated partners[94] and to adopt different structures upon binding to different partners,[52,94–99] to form fuzzy complexes preserving significant disorder in the bound state.[89,100–104]

Being natural regulators and controllers of various biological processes, IDPs/IDPRs are tightly regulated and controlled themselves. For example, alternative splicing (AS), which is a process by which two or more mature mRNAs are produced from a single precursor pre-mRNA by the inclusion and omission of different segments,[105,106] has an intimate link to intrinsic disorder, since mRNA regions affected by AS typically encodes for IDPR in a protein.[107] It is believed that being mostly found in multicellular eukaryotes,[108] AS provides an important mechanism for enhancing protein diversity in multicellular eukaryotes[109] by expanding various protein functions, such as protein–protein interactions, ligand binding, regulation, recognition, and enzymatic activity.[110–112] Furthermore, IDPRs (or their close proximity) often contain sites of various posttranslational modifications (PTMs, such as acetylation, hydroxylation, ubiquitination, methylation, phosphorylation, etc.) and proteolytic attack.[113–115] On the other hand, disruption of such disorder-based PTM sites by mutations often causes diseases.[116] More generally, pathogenesis of various human maladies, such as amyloidoses, cardiovascular disease, cancer, diabetes, and neurodegenerative diseases is commonly linked to dysfunctions of IDPs/IDPRs.[76,117–123]

3. Prevalence of Intrinsic Disorder in Proteins Associated with Human Membrane-Less Organelles

As it was already emphasized, all PMLOs contain specific sets of resident proteins. Several previous studies clearly indicated that the presence of significant levels of intrinsic disorder represents a characteristic feature of some of the proteins associated with PMLOs.[35–46] However, to the best of our knowledge, no systematic analysis of the intrinsic disorder predisposition was conducted so far for the PMLO proteome. To fill this gap, we discuss below results of a systematic bioinformatics analysis of the disorder status of 4796 human proteins from 20 PMLOs. These proteins were retrieved mostly using the outputs of the QuikGO tool (https://www.ebi.ac.uk/QuickGO) complemented with some literature search. The analyzed proteins were distributed among the human PMLOs as follows: Nucleolus (1626) > Chromatin (1350) > Nuclear speckles (650) > Centrosome (530) > Mitochondrial RNA granules (229) > PML bodies (104) > SGs (57) > Perinuclear compartment (55) > CBs (54) > PcG bodies (48) > P-granules (19) > Nuage (18) > Cleavage bodies (14) > Gemini (10) > SAM68 bodies (8) > Paraspeckles (6) > Nuclear SGs (5) = OPT domain (5) > HLB (4) = Neuronal RNP granules (4). One should keep in mind that, by no means, this list is exhaustive and contains all human PMLO-related proteins. In fact, a recently designed Cell Atlas representing a comprehensive image-based map of the subcellular protein distribution identified localization of 12 003 human proteins to 30 subcellular structures assembled into 13 major organelles, such as nucleus (1922 proteins, together with nucleoplasm, nuclear speckles, and nuclear bodies containing 3739, 444 and 482 proteins, respectively), nucleoli (1016 proteins, together with fibrillar center (254 proteins) and rim of nucleoli), nuclear membrane (272 proteins), Golgi apparatus (959 proteins), endoplasmic reticulum (430 proteins), vesicles (1806 proteins, together with lipid droplets containing 35 proteins), plasma membrane (1466 proteins, together with cell junctions containing 285 proteins), mitochondria (1070 proteins), cytosol (4279 proteins, together with cytoplasmic bodies (48 protein), aggresomes (17 proteins), and rods and rings (18

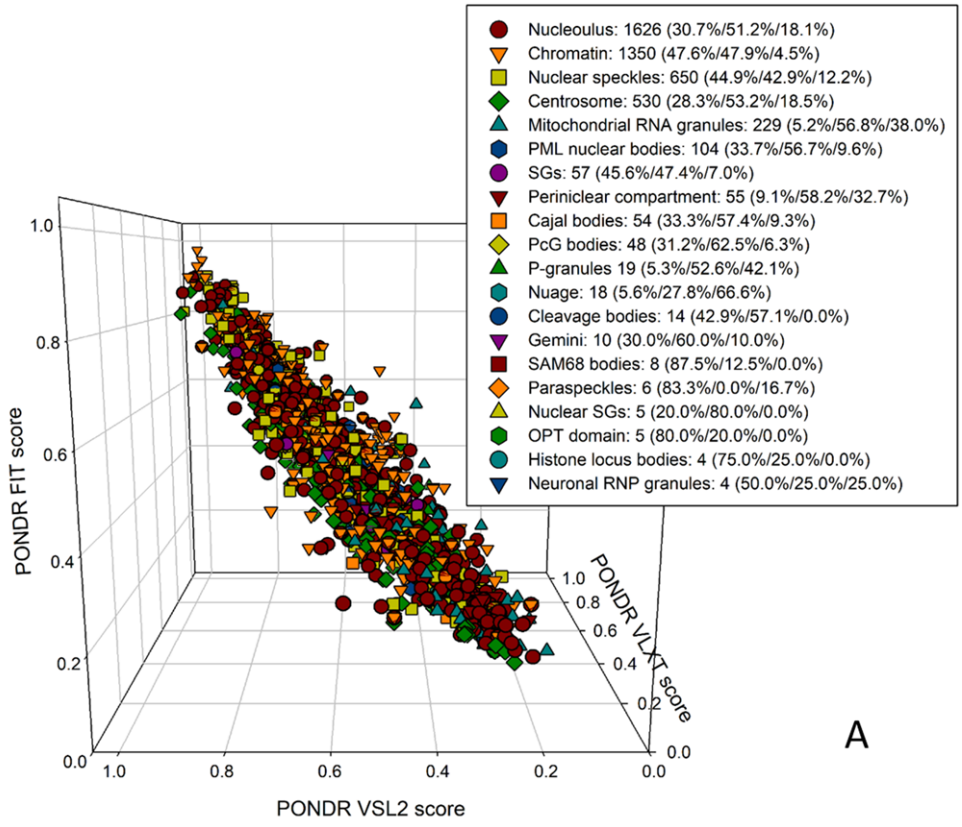
proteins)), intermediate filaments (179 protein), microtubules (263 proteins, together with microtubule ends (four proteins), cytokinetic bridge (88 proteins), mitotic spindle (17 proteins), midbody (36 proteins), and midbody ring (12 proteins)), centrosome (336 proteins, together with the microtubule organizing center containing 132 proteins), and actin filaments (223 proteins, together with focal adhesions containing 133 proteins).[124]

Figures 2, **3**, **4**, and **5** represent the result of a global analysis of the intrinsic disorder predisposition of 4796 human proteins associated with different PMLOs. To this end, first, we looked at their overall disorder levels (protein-average disorder scores, PADS) evaluated by three members of the PONDR family of disorder predictors, PONDR® FIT,[125] PONDR® VLXT,[66] and PONDR® VSL2.[114,126,127] Results of this analysis are summarized in Figure 2A, where correlation between the outputs of these three predictors are shown as 3D plot, and which illustrate that significant fractions of proteins in the majority of human PMLOs are noticeably disordered. This conclusion is based on the consideration of the outputs of this analysis using the criteria of accepted classification, where two arbitrary cutoffs for the levels of intrinsic disorder are used to classify proteins as highly ordered ($PADS_{ho} < 0.25$), moderately disordered ($0.25 \leq PADS_{md} < 0.5$), and highly disordered ($PADS_{hd} \geq 0.5\%$).[128] Figure 2A shows that according to this classification, the majority of proteins in almost all PMLOs are highly or moderately disordered. By their disorderedness degree ($PADS_{md} + PADS_{hd}$) evaluated by PONDR® FIT, human PMLOs can be ranked as follows: nuage (33.4%) < P-granules (57.9%) < mitochondrial RNA granules (62.0%) < perinuclear compartment (67.3%) < neuronal RNP granules (75.0%) < centrosome (81.5%) < nucleolus (81.9%) < paraspeckles (83.3%) < nuclear speckles (87.8%) < gemini (90.0%) < PML bodies (90.3%) < CBs (90.7%) < SGs (93.0%) < PcG bodies (93.7%) < chromatin (95.5%) < cleavage bodies (100.0%) = SAM68 bodies (100.0%) = nuclear SGs (100.0%) = OPT domain (100.0%) = HLB (100.0%). Analogous analysis conducted for the entire human proteome (20 228 proteins retrieved from the Consensus Coding Sequence database[129–131]) revealed that this set contains 3969 (19.6%), 5164 (25.5%), and 11 095 (54.9%) highly disordered ($PADS_{hd} \geq 0.5\%$), moderately disordered ($0.25 \leq PADS_{md} < 0.5$), and highly ordered ($PADS_{ho} < 0.25$) proteins, respectively (see Figure 2B). In other words, according to their overall disorderedness degree ($PADS_{md} + PADS_{hd}$) evaluated by PONDR® FIT (45.1%), proteins of human proteome are noticeably less disordered than proteins in all human PMLOs (except for nuages). These data clearly show that the proteome of PMLOs is, in general, highly intrinsically disordered.

Figure 3A provides further illustration of the fact that PMLOs are characterized by a wide range of protein-average disorder scores, with the PMLO-average disorder scores ranging from 0.72–0.67–0.64 in the OPT domain to 0.40–0.30–0.27 in nuage as evaluated by PONDR® VSL2 PONDR® VLXT, and PONDR® FIT, respectively. Importantly, Figure 3B shows that the vast majority of human PMLO-related proteins are characterized by the presence of at least one long IDPR (identified as a protein region containing at least 30 disordered residues), and several proteins (e.g., from chromatin, nuclear speckles, PcG bodies, PML bodies, CBs, nucleolus, and centrosome) contain more than ten such regions each.

Although the presence of high levels of structural disorder in human PMLO-associated proteins, the existence of disorder-based binding regions is at least as important, since those regions emphasize potential functionality of intrinsic disorder. Therefore, we applied three well-established computational tools for finding potential disorder-based protein binding sites in human PMLO-associated proteins. These three algorithms are α-MoRF-Pred I,[90] α-MoRF-Pred II,[92] and ANCHOR.[132,133] The first two tools are specifically designed for finding α-helix-forming MoRFs (i.e., disordered regions that undergo transition to α-helix at interaction with binding partner,[90,92]). Whereas ANCHOR algorithm utilizes the pair-wise energy estimation approach originally used by IUPred.[134,135] This approach acts on the hypothesis that long regions of disorder include localized potential binding sites which are not capable of folding on their own due to not being able to form enough favorable intrachain interactions, but can obtain the energy to stabilize via interaction with a globular protein partner.[132,133] Results of these analyses are summarized in Figure 4 that represents the abundance of MoRFs in proteins from various PMLOs. In fact, according to these analyses, all proteins have at least one MoRF predicted by at least one computational tool (see Figures 4A, C, and E). Centrosome, Sam68 nuclear body, and SG proteins have the highest rates of disorder-based binding site prediction, whereas chromatin proteins show the lowest rates (with none predicted by α-MoRF-Pred I). According to α-MoRF-Pred I/ α-MoRF-Pred II/ANCHOR, the median numbers of MoRFs per protein ranges from 0/1/2 in chromatin to 6/11/19 in centrosome. Figures 4B, D, and F show that very significant fraction of residues in human proteins associated with various PMLOs can be potentially engaged in disorder-based interactions with other proteins. These observations indicate that intrinsic disorder is not only common in human PMLO-related proteins, but is systematically used for protein–protein interactions.

Next, we utilized binary disorder predictors that evaluate the predisposition of a query protein to be ordered or disordered as a whole. The outputs of two such tools, the charge-hydropathy (CH) plot[54,62] and the cumulative distribution function (CDF) plot,[62,136] were combined to generate the CH-CDF plot.[136–138] Although binary predictors operate at the level of whole proteins and, therefore, generate results at a much lower resolution than the residue-level predictors considered in the previous section, these two types of disorder prediction (generated by binary and per-residue predictors) produce complementary data. As a matter of fact, although binary predictors evaluate the overall disorder propensity of query proteins at the whole molecule level, by combining outputs of such predictors, proteins can be grouped into different structural classes (see below), whereas such classification is not possible with the per-residue predictors. In fact, in a CH-CDF plot,[136–138] the coordinates of a query protein are calculated as following: Y-coordinate corresponded to the distance of the point representing this protein in the CH-plot from the boundary (ΔCH), whereas the X-coordinate was an average distance of the respective CDF curve from the CDF boundary (ΔCDF). In the resulting CH-CDF plot, positive and negative Y-values correspond to proteins predicted by CH-plot to be extended or compact, respectively. However, positive and negative X-values correspond to proteins predicted to be ordered or intrinsically disordered by CDF analysis. The CH-CDF phase space



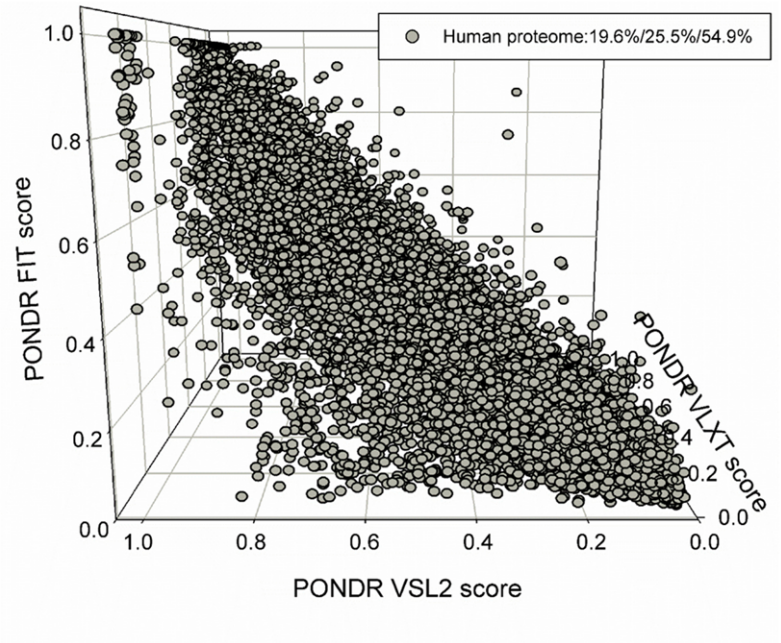


Figure 2. Abundance of intrinsic disorder in human PMLO-related proteins A) and in human proteome B). In both plots, consensus 3D PONDR® FIT vs. PONDR® VSL2 *vs.* PONDR® VLXT plot representing the correlation between the disorder content in human PMLO-related proteins A) and in human proteome B) evaluated by PONDR® VLXT (X-axis), PONDR® VSL2 (Y-axis), and PONDR® FIT (Z-axis). Following the accepted practice, two arbitrary cutoffs for the levels of intrinsic disorder are used to classify proteins as highly ordered ($PADS_{ho} < 0.25$), moderately disordered ($0.25 \leq PADS_{md} < 0.5$) and highly disordered ($PADS_{hd} \geq 0.5\%$).[128] The values in brackets show the content of highly disordered, moderately disordered, and highly ordered proteins in each PMLO A) or in human proteome B). Data for the human proteome (20 228 proteins) were retrieved from the Consensus Coding Sequence database.[129]

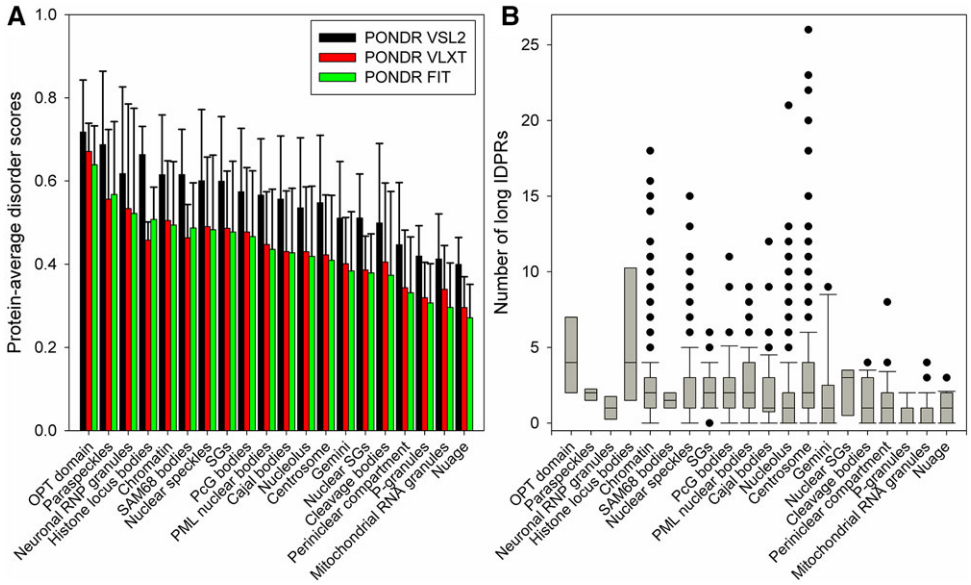


Figure 3. Evaluation of the overall disorder levels and peculiarities of disorder distribution in human proteins associated with PMLOs. A) Spread of the protein-average disorder scores in individual PMLOs evaluated by PONDR® VSL2 (black bars), PONDR® VLXT (red bars), and PONDR® FIT (green bars). Bars show mean protein-average disorder scores in corresponding PMLOs, whereas error bars reflect the corresponding standard deviations calculated by SigmaPlot software. B) Box-and-whisker plot representing statistical analysis of the commonness of long IDPRs in human proteins associated with various PMLOs. In this plot, the top of each box indicates the third quartile, a horizontal line near the middle of the box indicates the median, and the bottom of the box indicates the first quartile. A vertical line extending from the top of the box indicates the maximum value, whereas a vertical line extending from the bottom of the box indicates the minimum value. Black circles represent outliers. Boxes without whiskers correspond to the PMLOs with small number of proteins (4–8).

provides specific expectations for the disorder status of a protein depending on its position within four quadrants. Here, the upper-right quadrant Q1 contains proteins predicted to be disordered by CH-plot but ordered by CDF; the lower-right quadrant Q2 is occupied by ordered proteins; the lower-left quadrant Q3 includes proteins that are predicted as disordered by CDF but compact by CH-plot (i.e., native molten globules or hybrid proteins containing comparable quantities of order and disorder); whereas the upper-left quadrant Q4 contains proteins with extended disorder, such as native coils and native pre-molten globules.[137]

Figure 5A represents the results of the CH-CDF analysis of human proteome that revealed the following distribution of 20 848 proteins between the quadrants: Q1 ($\Delta CH > 0$ and $\Delta CDF > 0$): 167 (0.8%); Q2 ($\Delta CH < 0$ and $\Delta CDF > 0$): 11 380 (54.6%); Q3 ($\Delta CH < 0$ and $\Delta CDF < 0$): 6237 (29.9%); and Q4 ($\Delta CH > 0$ and $\Delta CDF < 0$): 3064 (14.7%). In other words, mostly disordered proteins in human proteome (e.g., proteins located in quadrants Q3 and Q4 of the CH-CDF plot) account for 44.6%. Similar analysis of 4796 human proteins associated with various PMLOs shows that 67 (1.4%), 1642 (34.2%), 1863 (38.9%), and 1224 (25.5%) such proteins are found in quadrants Q1, Q2, Q3, and Q4, respectively (see Figure 5B), indicating 64.4% of these proteins are mostly disordered. In other words, according to this analysis, PMLO-associated proteins are noticeably more disordered than the proteins in human proteome. Furthermore, by their disorder level (percentage of proteins in quadrants Q3 and Q4), human PMLOs can be ranked as follows: mitochondrial RNA granules (18.0%) < nuage (22.0%) < P-granules (25.3%) < perinuclear compartment (30.9%) < nuclear SGs (40.0%) < cleavage bodies (42.8%) < nucleolus (56.4%) < gemini (60.0%) < PML bodies (61.2%) < centrosome (62.7%) < CBs 66.7%) = PcG bodies (66.7%) < SGs (68.4%) < nuclear speckles (72.3%) < neuronal RNP granules (75.0%) < chromatin (80.3%) < paraspeckles (83.3%) < SAM68 bodies (100.0%) = OPT domain (100.0%) = HLB (100.0%). This means that the proteomes of 16 of 20 human PMLOs are noticeably more disordered than human proteome in general, indicating potential functional importance of intrinsic disorder for biogenesis and functionality of PMLOs.

4. Intrinsic Disorder in Membrane-Less Organelles: What is it for?

The findings described in the previous section are in line with known data on the prevalence of intrinsic disorder in some PMLOs. In fact, although current literature contains rather limited information on proteins experimentally shown to undergo LLPTs in aqueous solutions alone or in mixtures with other proteins, nucleic acids, or polysaccharides, but all such proteins are either IDPs or hybrid proteins containing ordered domains and long IDPRs.[35] Furthermore, previous studies indicated that some specific IDPs can be related to the biogenesis of nuages,[2] nucleolus,[139] P-granules,[38] and RNA granules.[39] Several illustrative examples, where LLPTs were experimentally observed and analyzed in solutions containing just one protein (e.g., nuage-related Ddx4 protein,[2] LAF-1 protein associated with the biogenesis of P-granules,[140] and SG-related TIA-1 protein,[141,142] or

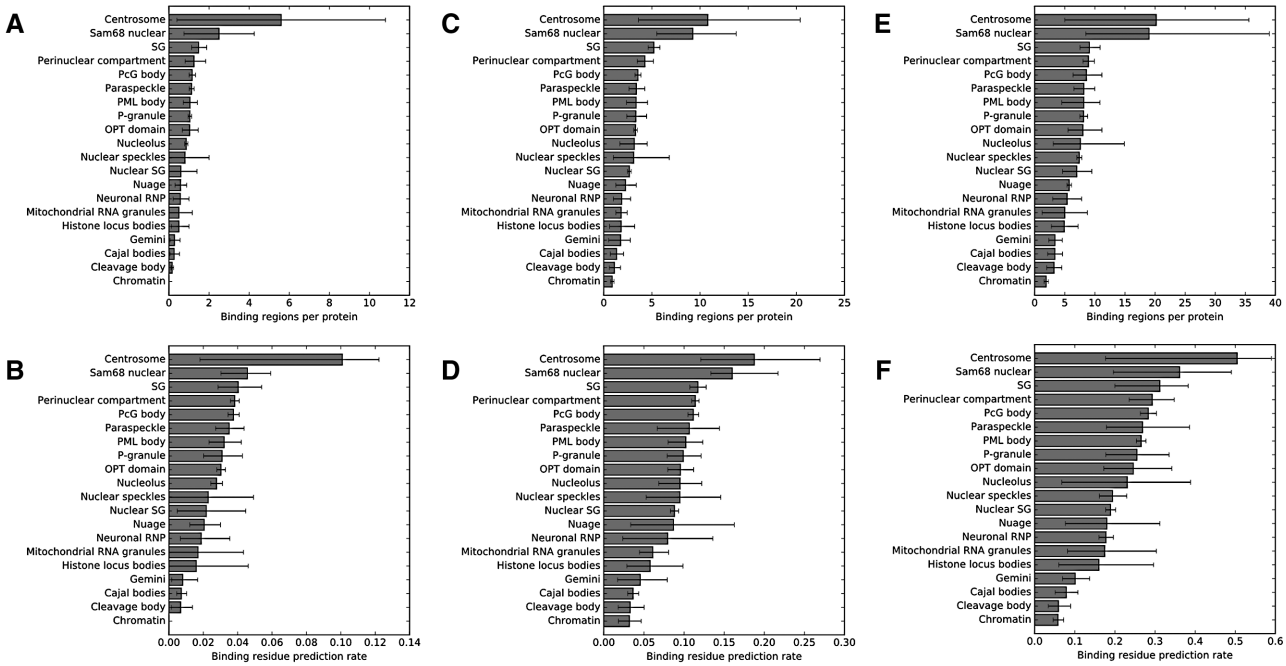


Figure 4. Finding potential disorder-based protein–protein interactions sites in human PMLO-related proteins using the α-MoRF-Pred I[90] (A and B), α-MoRF-Pred II[92] (C and D), and ANCHOR algorithms[132,133] (E and F). For all three tools, data are aggregated to show predicted binding regions per protein (plots A, C, and E) and predicted binding residues per residue (plots B, D, and F). In these plots, horizontal bars correspond to median values, whereas error bars show 2.5 percentile and 97.5 percentile, all estimated by 100 000 bootstrap iterations. Between 2.5 and 97.5 is the 95% confidence interval on the median.

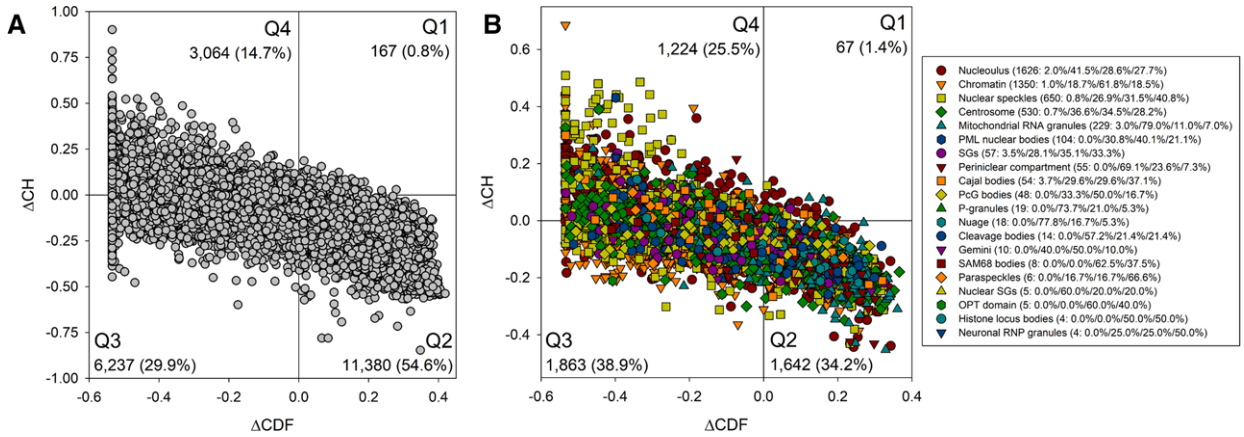


Figure 5. Evaluating global intrinsic disorder predisposition of all human proteins A) and human proteins associated with various PMLOs B) by combining the outputs of binary disorder classifiers, CH-plot[54] and CDF.[54,62,158] Here, the coordinates of each point were calculated as a distance of the corresponding protein in the CH-plot from the boundary (Y-coordinate) and an average distance of the respective CDF curve from the CDF boundary (X-coordinate). The four quadrants correspond to the following predictions: Q1, proteins predicted to be disordered by CH-plots, but ordered by CDFs; Q2, ordered proteins; Q3, proteins predicted to be disordered by CDFs, but compact by CH-plots (i.e., putative molten globules or hybrid proteins); and Q4, proteins predicted to be disordered by both methods.

IDPRs of several RNA-binding proteins associated with the assembly of different RNP granules, such as $Pub1_{IDPR}$, $eIF4GII_{IDPR}$, $Lsm4_{IDPR}$, TIA-1_{IDPR}, Fus_{IDPR}, and $hnRNPA1_{IDPR}$[39]), or systems containing a protein and RNA (e.g., aforementioned $Pub1_{IDPR}$, $eIF4GII_{IDPR}$, $Lsm4_{IDPR}$, TIA-1_{IDPR}, Fus_{IDPR}, and $hnRNPA1_{IDPR}$ mixed with RNA at physiological conditions,[39] solution of the Whi3 protein and *CLN3* mRNA[140]), as well as systems, where at least two proteins are needed for LLPT (e.g., maternal-effect germline defective proteins (MEG-1 (maternal-effect germline defective protein 1), MEG-2, MEG-3, and MEG-4) interacting with PGL-1 (P-granule abnormality protein 1)[143]), or more complex multi-protein (e.g., mixtures of synthetic multivalent signaling proteins[13] and mRNA decapping machinery related to the formation of P-bodies[144]) or multi-protein-RNA systems undergoing phase separation (e.g., nucleolus-related nucleophosmin (NPM1) interacting with RNA-binding proteins containing

numerous arginine-rich linear motifs (R-motifs) and rRNA[139]), were discussed in some detail in [35].

Computationally, the idea on the commonness of intrinsic disorder in PMLO-related proteins was validated for 3005 mouse proteins localized in several nuclear PMLOs,[40,145] thereby supporting data for human proteins reported in the current study. This important observation indicates that high level of intrinsic disorder is present in PMLO-related proteins from different organisms, suggesting evolutionary conservation of the disorder-PMLO link. Furthermore, it was hypothesized that the critical dependence on the intrinsic disorder might represent a global feature of the mechanisms underlining the formation and disintegration of not only some of the "assemblages",[37,146] but of all PMLOs and complex biological coacervates.[35,41,46] A very specific feature of PMLOs is that biological LLPTs leading to their formation are reversible, highly controllable, and do happen under the physiological conditions of living cells. Why, then, are IDPs crucial for such biologically relevant LLPTs and, thus, for PMLO biogenesis? What does define the LLPT compatibility of these proteins? Why would IDPs/IDPRs serve as abundant constituents and crucial players of the formation and disassembly of PMLOs?

As it was already emphasized, although phase separation is controlled by environmental alterations, such as changes in pH, solution ionic strength, and temperature among other factors, one of the most important facets contributing to the phase separation (if not the most important one) is concentrations of macromolecules.[147] In fact, phase separation can typically occur only when a specific concentration threshold is reached.[47,48] Therefore, the high abundance of IDPs/IDPRs in eukaryotic cells represents an important contributing factor for the commonness of such proteins in PMLOs. It was also pointed out that proteins undergoing LLPTs, and thereby commonly found in PMLOs, are typically characterized by the presence of repetitive units, multivalency, flexibility, enrichment in some specific residues, and accessibility to PTMs.[35,46] All these features are characteristic for proteins with intrinsic disorder. Therefore, the lack of fixed structure in IDPs/IDPRs, capability to participate in a wide spectrum of interactions of different physico-chemical nature and strength, and ability to serve as common targets for various PTMs represent crucial elements of the IDP-PMLO engagement. Overall, it is likely that the lack of stable structure represents one of the crucial determinants defining the ability of proteins to form cyto- and nucleoplasmic PMLOs via reversible and highly controlled LLPTs. Let us take a closer look at some of these factors linking PMLOs with IDPs/IDPRs.

4.1. High Conformational Flexibility of IDPs is Needed for the Fluidity of Resulting PMLOs

Physically, PMLOs are characterized as liquid droplets found in the cytoplasm, nucleoplasm, mitochondrial matrix, and stroma of chloroplasts.[4–7,13,14] Their physico-chemical properties are generally rather similar to those of the surrounding fluids,[3] and their density is only slightly higher than that of the surrounding intracellular fluids.[11,12] As a result, PMLOs are able to drip, fuse, wet, and relax to spherical structures upon fusion.[4–7] It is likely that the physical fluidity of PMLOs (which typically are RNP droplets) is determined by the conformational "fluidity" of their principle constituents, flexible RNAs and intrinsically disordered RNA-binding proteins. However, one should also keep in mind that in addition to the fluid droplets, some flexible polymers (despite their structural flexibility) can form strong gels or glassy domains via weak interactions (a few kT per link),[148,149] indicating that assembly of flexible constituents can generate physically different conglomerations.

4.2. Specific but Weak Disorder-Based Interactions Contribute to the PMLO Biogenesis

The strength of interactions between the constituents defines both the stability of the complexes and the reversibility of their formation. Obviously, only when interactions are specific but weak, the resulting assemblages might have liquid-like properties and might rapidly disintegrate in response to changes. Since many IDPs/IDPRs are known to be engaged in multiple specific, but weak interactions, this can give a possible explanation for the critical involvement of intrinsic disorder in biological LLPTs and formation of various PMLOs in a highly controllable manner.

It was pointed out[35] that an illustrative example providing support to the hypothesis that multiple weak interactions can hold partners together is given by a "polyelectrostatic model". This model was developed to explain the mechanism of formation of a highly dynamic binary complex between an ordered protein, the SCF ubiquitin ligase subunit Cdc4, possessing a single receptor site recognizing phosphorylated serines and threonines and an IDP, the cyclin-dependent kinase inhibitor Sic1, that can be phosphorylated at nine sites or phosphodegrons each forming suboptimal binding motifs.[150] In the resulting Sic1-Cdc4 complex, Sic1 utilizes all its phosphorylated sites, regardless of their location relative to the receptor site of Cdc4, for interaction with Cdc4 via the spatially long-range polyelectrostatic interactions.[150] Importantly, such polyelectrostatic interaction mode defines the ultrasensitivity of the resulting complex to the Sic1 phosphorylation degree, where the strength of Sic1-Cdc4 interaction increases proportionally to the increase in the number of phosphorylated sites.[150]

It was also pointed out that the polyelectrostatic model can provide a mechanistic description of the PMLO assembly, although these decently sized liquid droplets are not typical proteinaceous complexes.[35] In fact, it is likely that similar to the case of Sic1-Cdc4,[151] highly flexible members of the conformational ensembles of some of the PMLO-forming IDPs/IDPRs, instead of presenting discrete charges, create mean electrostatic fields that are utilized in polyelectrostatic attraction.[35]

4.3. Roles of Intrinsic Disorder in the Resilience and Stability of PMLOs

Despite the fact that PMLOs devoid membranes and their constituents are in free and constant exposure and exchange with the environment, these organelles are characterized by remarkable resilience and stability. In fact, once formed, PMLOs stay

assembled for as long as it is required. Since IDPs/IDPRs are important for the biogenesis of the phase-separated liquid droplets, these structure-less proteins should have some specific features that are helping them to hold fluid assemblages together. Obviously, this is rooted in a principle mechanistic difference between the ordered proteins and IDPs in formation of proteinaceous assemblages. In fact, ordered proteins, with their limited number of specific interfaces giving rise to the limited number of specific high affinity interactions, resemble rigid building blocks or bricks within a wall. Here, a resulting assembly ("a brick wall") is held together by the ability of each "brick" to specifically fit into a defined and well-ordered niche, shape of which is determined by the shape of the said brick that is complementary to the shapes of other bricks. Despite being "sturdy", such rigid assemblage, however, can be easily damaged or even completely destroyed if a few or even just one brick is taken out. This is different from a liquid droplet, formation of which is determined by multiple weak interactions between IDPs/IDPRs. Here, a resulting assembly ("a bowl of noodles") is formed by flexible constituents that are never engaged in real bonding, but form a multitude of transient contacts, where molecules are constantly touching each other utilizing a swarm of their own binding motifs to act on a host of binding motifs of partners. As a result, PMLOs, these fluid complexes made of flexible constituents, have a resilience akin to a bowl of noodles, which remains to be a bowl of noodles even when many noodles are eaten.[35,46]

4.4. Roles of Disorder-Based Flexible Polyvalency

Biogenesis and fluidity of PMLOs rely on flexible polyvalency of their constituents. Such flexible polyvalency can be defined by the presence of repetitive units of various physico-chemical nature in many PMLO-related IDPs. These could be alternating blocks (or clusters) of opposite charges spread over the IDPRs of proteins undergoing unimolecular LLPTs (e.g., N-terminal tail of protein LAF-1[140] or disordered tails of orthologous Ddx4 proteins[2]), or some other repetitive units found in more complex cases (e.g., multiple RNA recognition motifs (RRMs) in the RNA-binding protein TIA-1[142]; multiple PRMs (proline-rich motifs) in WASP protein[152]; leucine-rich motifs in Dcp2[144]; R-motifs in NPM1,[139] and a polyQ tract in the Whi3 protein.[140])

4.5. Roles of PTMs in the PMLO biogenesis

Since PTMs can affect physico-chemical properties of target proteins, and since the presence of multiple sites of various PTMs represents one of the characteristic features of many IDPs/IDPRs,[113–115] it is not too surprising to see that the efficiency of LLPT and PMLO formation can be affected by the PTMs.[13] In fact, several cases are known showing a critical dependence of the efficiency of an LLPT on the PTM status of participating constituents. For example, methylation of several arginine residues of Ddx4 protein noticeably destabilizes the Ddx4–based organelles indicating that PMLO biogenesis can be controlled by the methylation of this protein.[2] The processes of P-granule assembly and disassembly are regulated by phosphorylation and dephosphorylation of MEG-1, MEG-2, MEG-3, and MEG-4 proteins.[143] The Ubp3/USP10-driven deubiquitination of several constituent proteins is required for the efficient SG formation in *Saccharomyces cerevisiae*[153] and in mammalian cells.[154] Phosphorylation degree of nephrin fragment defines the efficiency of the LLPT in the Nck/N-WASP binary system[13,155]).

To provide an oversimplified but illustrative representation of the physical mechanism that leads to this correlation between PTMs and PMLO biogenesis, let us consider the "polyelectrostatic model" once again. In the aforementioned Sic1-Cdc4 system, the phosphorylation-dependent ultrasensitive binding of Sic1 to Cdc4 is defined by the presence of multiple phosphorylatable suboptimal binding motifs in Sic1.[150] In fact, it is known that phosphorylation of any six of its nine sites is required for Sic1 to bind to Cdc4. Furthermore, the Sic1 phosphorylation degree (that obviously affects the net charge of this protein) is used to control the strength of Sic1-Cdc4 interaction, which increases proportionally to the increase in the number of phosphorylated sites from six to nine. Since no Sic1-Cdc4 complex is formed when Sic1 contains less than six phosphorylated sites and since binding strength is proportional to the extent of Sic1 phosphorylation, one can conclude that the degree of Sic1 phosphorylation in this system serves both as an "on-off switch" controlling the formation and dissociation of the Sic1-Cdc4 complex or as an "rheostat" regulating the strength of Sic1-Cdc4 interaction.[150]

4.6. Effects of External Factors in Regulation of PMLO Formation

There are multiple external factors that can trigger LLPTs and, therefore, control the formation of PMLOs. Among those cues are changes in concentrations of macromolecular constituents undergoing LLPTs, changes in the concentrations of specific small molecules interacting with constituents involved in the formation of PMLOs, as well as changes in the ionic strength, pH, and temperature of the solution. Although well-folded proteins are characterized by the funnel-like energy landscapes with a well-defined global energy minimum corresponding to their folded conformation,[156,157] the energy landscape of an IDP is relatively flat, lacks such a deep energy minimum, but possesses numerous local energy minima, due to which protein tend to behave as a highly frustrated system without any stable well-folded conformation.[74] Because of such peculiar 'topology' of their energy landscapes IDPs/IDPRs are characterized by conformational plasticity and are exceptionally sensitive to local environment, being much more sensitive than ordered proteins with the relatively robust funnel-like energy landscapes.[74] As a result, any changes in the IDP/IDPR surroundings might have a very strong effect on their structures. Furthermore, different environmental factors might differently affect the energy landscape of an IDP/IDPR defining it's the ability to fold and/or interact differently, depending on the peculiarities of environmental conditions.[74]

Abbreviations

AS, alternative splicing; CB, Cajal body; Cdc4, cell division control protein 4; CDF, cumulative distribution function; CH-plot, charge-hydropathy plot; HLB, histone locus body; IDP, intrinsically disordered protein; IDPR,

intrinsically disordered protein region; LLPT, liquid–liquid phase transition; MEG-1, maternal-effect germline defective protein 1; MoRF, molecular recognition feature; NPM1, nucleophosmin; nSB, nuclear stress body; OPT, Oct1/PTF/ transcription; PADS, protein-average disorder score; PcG, polycomb; PGL-1, P-granule abnormality protein 1; PML, promyelocytic leukaemia; PMLO, proteinaceous membrane-less organelle; PNC, perinucleolar compartment; POD, PML oncogenic domain; PRM, proline-rich motif; PTM, posttranslational modification; R-motif, arginine-rich linear motif; RNP, ribonucleoprotein; RRM, RNA recognition motif; SG, stress granule; Sic1, cyclin-dependent kinase inhibitor; SNB, Sam68 nuclear body

Acknowledgements

A.L.D. and Y.L. contributed equally to this work. We are thankful to Prof. Bin Xue (University of South Florida) for his help with bioinformatics analysis of human proteome and human PMLO-related proteins.

Conflict of Interest

The authors have declared no conflict of interest.

Keywords

intrinsically disordered proteins, liquid-liquid phase transition, phase separation, proteinaceous membrane-less organelle, protein-nucleic acid interactions

Received: August 29, 2017
Revised: February 10, 2017

[1] E. Metschnikoff, *Arch. Naturgesch.* **1865**, *31*, 304.
[2] T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, A. J. Baldwin, *Mol. Cell* **2015**, *57*, 936.
[3] C. P. Brangwynne, *J. Cell Biol.* **2013**, *203*, 871.
[4] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, A. A. Hyman, *Science* **2009**, *324*, 1729.
[5] C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 4334.
[6] M. Feric, C. P. Brangwynne, *Nat. Cell Biol.* **2013**, *15*, 1253.
[7] F. Wippich, B. Bodenmiller, M. G. Trajkovska, S. Wanka, R. Aebersold, L. Pelkmans, *Cell* **2013**, *152*, 791.
[8] M. Dundr, T. Misteli, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000711.
[9] R. D. Phair, T. Misteli, *Nature* **2000**, *404*, 604.
[10] T. Pederson, *Cell* **2001**, *104*, 635
[11] K. E. Handwerger, J. A. Cordero, J. G. Gall, *Mol. Biol. Cell* **2005**, *16*, 202
[12] D. L. Updike, S. J. Hachey, J. Kreher, S. Strome, *J. Cell Biol.* **2011**, *192*, 939
[13] P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. Jiang, B. T. Nixon, M. K. Rosen, *Nature* **2012**, *483*, 336
[14] S. Aggarwal, N. Snaidero, G. Pahler, S. Frey, P. Sánchez, M. Zweckstetter, A. Janshoff, A. Schneider, M. T. Weil, I. A. Schaap, D. Görlich, M. Simons, *PLoS Biol.* **2013**, *11*, e1001577.
[15] M. Decker, S. Jaensch, A. Pozniakovsky, A. Zinke, K. F. O'Connell, W. Zachariae, E. Myers, A. A. Hyman, *Curr. Biol.* **2011**, *21*, 1259.
[16] S. Chuma, M. Hosokawa, T. Tanaka, N. Nakatsuji, *Mol. Cell. Endocrinol.* **2009**, *306*, 17.
[17] M. A. Kiebler, G. J. Bassell, *Neuron* **2006**, *51*, 685.
[18] C. J. Decker, D. Teixeira, R. Parker, *J. Cell Biol.* **2007**, *179*, 437
[19] M. Strzelecka, S. Trowitzsch, G. Weber, R. Luhrmann, A. C. Oates, K. M. Neugebauer, *Nat. Struct. Mol. Biol.* **2010**, *17*, 403.
[20] B. Li, M. Carey, J. L. Workman, *Cell* **2007**, *128*, 707.
[21] L. Li, K. Roy, S. Katyal, X. Sun, S. Bléoo, R. Godbout, *Mol. Biol. Cell* **2006**, *17*, 1126.
[22] Z. Nizami, S. Deryusheva, J. G. Gall, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000653.
[23] A. G. Matera, M. R. Frey, *Am. J. Hum. Genet.* **1998**, *63*, 317.
[24] A. K. Gubitz, W. Feng, G. Dreyfuss, *Exp. Cell Res.* **2004**, *296*, 51.
[25] E. Grossman, O. Medalia, M. Zwerger, *Annu. Rev. Biophys.* **2012**, *41*, 557.
[26] A. I. Lamond, D. L. Spector, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 605.
[27] G. Biamonti, C. Vourc'h, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000695.
[28] G. Biamonti, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 493.
[29] Y. Shav-Tal, J. Blechman, X. Darzacq, C. Montagna, B. T. Dye, J. G. Patton, R. H. Singer, D. Zipori, *Mol. Biol. Cell* **2005**, *16*, 2395.
[30] J. A. Harrigan, R. Belotserkovskaya, J. Coates, D. S. Dimitrova, S. E. Polo, C. R. Bradshaw, P. Fraser, S. P. Jackson, *J. Cell Biol.* **2011**, *193*, 97.
[31] A. H. Fox, Y. W. Lam, A. K. Leung, C. E. Lyon, J. Andersen, M. Mann, A. I. Lamond, *Curr. Biol.* **2002**, *12*, 13.
[32] V. Pirrotta, H. B. Li, *Curr. Opin. Genet. Dev.* **2012**, *22*, 101.
[33] S. Huang, *J. Struct. Biol.* **2000**, *129*, 233.
[34] G. G. Maul, D. Negorev, P. Bell, A. M. Ishov, *J. Struct. Biol.* **2000**, *129*, 278.
[35] V. N. Uversky, *Adv. Colloid Interface Sci.* **2017**, *239*, 97.
[36] N. Kedersha, P. Ivanov, P. Anderson, *Trends Biochem. Sci.* **2013**, *38*, 494.
[37] J. A. Toretsky, P. E. Wright, *J. Cell Biol.* **2014**, *206*, 579.
[38] S. Elbaum-Garfinkle, Y. Kim, K. Szczepaniak, C. C. Chen, C. R. Eckmann, S. Myong, C. P. Brangwynne, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7189.
[39] Y. Lin, D. S. Protter, M. K. Rosen, R. Parker, *Mol. Cell* **2015**, *60*, 208.
[40] F. Meng, I. Na, L. Kurgan, V. N. Uversky, *Int. J. Mol. Sci.* **2015**, *17*.
[41] V. N. Uversky, I. M. Kuznetsova, K. K. Turoverov, B. Zaslavsky, *FEBS Lett.* **2015**, *589*, 15.
[42] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, M. K. Rosen, *Mol. Cell.* **2016**, *63*, 72.
[43] M. D. Panas, P. Ivanov, P. Anderson, *J. Cell Biol.* **2016**, *215*, 313.
[44] D. S. Protter, R. Parker, *Trends Cell Biol.* **2016**, *26*, 668.
[45] J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. Lopez, *Nat. Chem.* **2017**, *9*, 509.
[46] V. N. Uversky, *Curr. Opin. Struct. Biol.* **2017**, *44*, 18.
[47] C. D. Keating, *Acc. Chem. Res.* **2012**, *45*, 2114.
[48] V. Tolstoguzov, *Nahrung* **2000**, *44*, 299.
[49] H. Walter, D. E. Brooks, *FEBS Lett.* **1995**, *361*, 135.
[50] H. Walter, *Int. Rev. Cytol.* **2000**, *192*, 331.
[51] D. E. Brooks, *Int. Rev. Cytol.* **2000**, *192*, 321.
[52] A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger, J. E. Villafranca, *Pac. Symp. Biocomput.* **1998**, 473.
[53] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321.
[54] V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins* **2000**, *41*, 415.

[55] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, *J. Mol. Graph Model* **2001**, *19*, 26.
[56] P. Tompa, *Trends Biochem. Sci.* **2002**, *27*, 527.
[57] G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese, A. K. Dunker, in: J. Buchner, T. Kiefhaber (Eds.), *Handbook of Protein Folding*, Wiley-VCH, Weinheim, Germany **2005**, 271.
[58] V. N. Uversky, A. K. Dunker, *Biochim. Biophys. Acta* **2010**, *1804*, 1231.
[59] A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, C. J. Brown, *Genome Inform. Ser. Workshop Genome Inform.* **2000**, *11*, 161.
[60] P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, A. K. Dunker, *Pac. Symp. Biocomput.* **1998**, 437.
[61] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, *J. Mol. Biol.* **2004**, *337*, 635.
[62] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, A. K. Dunker, *Biochemistry* **2005**, *44*, 1989.
[63] B. Xue, A. K. Dunker, V. N. Uversky, *J. Biomol. Struct. Dyn.* **2012**, *30*, 137.
[64] M. E. Oates, P. Romero, T. Ishida, M. Ghalwash, M. J. Mizianty, B. Xue, Z. Dosztányi, V. N. Uversky, Z. Obradovic, L. Kurgan, A. K. Dunker, J. Gough, *Nucleic Acids Res.* **2013**, *41*, D508.
[65] Z. Peng, J. Yan, X. Fan, M. J. Mizianty, B. Xue, K. Wang, G. Hu, V. N. Uversky, L. Kurgan, *Cell Mol. Life Sci.* **2015**, *72*, 137.
[66] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, A. K. Dunker, *Proteins* **2001**, *42*, 38.
[67] R. M. Williams, Z. Obradovi, V. Mathura, W. Braun, E. C. Garner, J. Young, S. Takayama, C. J. Brown, A. K. Dunker, *Pac. Symp. Biocomput.* **2001**, 89.
[68] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, A. K. Dunker, *Biophys. J.* **2007**, *92*, 1439.
[69] F. Meng, V. N. Uversky, L. Kurgan, *Cell. Mol. Life Sci.* **2017**, *74*, 3069.
[70] V. N. Uversky, A. K. Dunker, *Anal. Chem.* **2012**, *84*, 2096.
[71] V. N. Uversky, *Adv. Exp. Med. Biol.* **2015**, *870*, 215.
[72] C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biochem.* **2014**, *83*, 553.
[73] D. P. Minde, A. K. Dunker, K. S. Lilley, *Proteomics* **2017**, *17*.
[74] V. N. Uversky, *Biochim. Biophys. Acta* **2013**, *1834*, 932.
[75] V. N. Uversky, *J. Biol. Chem.* **2016**, *291*, 6681–6688.
[76] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic, A. K. Dunker, *J. Mol. Biol.* **2002**, *323*, 573.
[77] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, V. N. Uversky, *FEBS J.* **2005**, *272*, 5129.
[78] V. N. Uversky, C. J. Oldfield, A. K. Dunker, *J. Mol. Recognit.* **2005**, *18*, 343.
[79] S. Vucetic, H. Xie, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, *J. Proteome Res.* **2007**, *6*, 1899.
[80] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, *J. Proteome Res.* **2007**, *6*, 1917.
[81] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky, Z. Obradovic, *J. Proteome Res.* **2007**, *6*, 1882.
[82] A. Patil, H. Nakamura, *FEBS Lett.* **2006**, *580*, 2041.
[83] D. Ekman, S. Light, A. K. Bjorklund, A. Elofsson, *Genome Biol.* **2006**, *7*, R45.
[84] C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, L. M. Iakoucheva, *PLoS Comput. Biol.* **2006**, *2*, e100.
[85] Z. Dosztanyi, J. Chen, A. K. Dunker, I. Simon, P. Tompa, *J. Proteome Res.* **2006**, *5*, 2985.
[86] G. P. Singh, D. Dash, *Proteins* **2007**, *68*, 602.
[87] G. P. Singh, M. Ganapathi, D. Dash, *Proteins* **2007**, *66*, 761.
[88] G. E. Schulz, in: M. Balaban (Ed.), *Molecular Mechanism of Biological Recognition*, Elsevier/North-Holland Biomedical Press, New York **1979**, pp. 79.
[89] V. N. Uversky, *Chem Soc Rev.* **2011**, *40*, 1623.
[90] C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, A. K. Dunker, *Biochemistry* **2005**, *44*, 12454.
[91] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, *J. Mol. Biol.* **2006**, *362*, 1043.
[92] Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky, A. K. Dunker, *Biochemistry* **2007**, *46*, 13468.
[93] F. M. Disfani, W. L. Hsu, M. J. Mizianty, C. J. Oldfield, B. Xue, A. K. Dunker, V. N. Uversky, L. Kurgan, *Bioinformatics* **2012**, *28*, i75.
[94] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, P. E. Wright, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11504.
[95] K. Landsteiner, *The Specificity of Serological Reactions*, Courier Dover Publications, Mineola, NY **1936**.
[96] L. Pauling, *J. Am. Chem. Soc.* **1940**, *62*, 2643.
[97] F. Karush, *J. Am. Chem. Soc.* **1950**, *72*, 2705.
[98] W. E. Meador, A. R. Means, F. A. Quiocho, *Science* **1993**, *262*, 1718.
[99] V. N. Uversky, *J. Biomol. Struct. Dyn.* **2003**, *21*, 211.
[100] M. Fuxreiter, P. Tompa, *Adv. Exp. Med. Biol.* **2012**, *725*, 1.
[101] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2.
[102] S. E. Permyakov, I. S. Millett, S. Doniach, E. A. Permyakov, V. N. Uversky, *Proteins* **2003**, *53*, 855.
[103] A. Sigalov, D. Aivazian, L. Stern, *Biochemistry* **2004**, *43*, 2049.
[104] A. B. Sigalov, A. V. Zhuravleva, V. Y. Orekhov, *Biochimie* **2007**, *89*, 419.
[105] J. Sambrook, *Nature* **1977**, *268*, 101.
[106] D. L. Black, *Annu. Rev. Biochem.* **2003**, *72*, 291.
[107] P. R. Romero, S. Zaidi, Y. Y. Fang, V. N. Uversky, P. Radivojac, C. J. Oldfield, M. S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic, A. K. Dunker, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8390.
[108] G. Ast, *Nat. Rev. Genet.* **2004**, *5*, 773.
[109] B. R. Graveley, *Trends Genet.* **2001**, *17*, 100.
[110] K. P. Minneman, *Mol. Interv.* **2001**, *1*, 108.
[111] T. H. Thai, J. F. Kearney, *J. Immunol.* **2004**, *173*, 4009.
[112] W. Scheper, R. Zwart, F. Baas, *Neurogenetics* **2004**, *5*, 223.
[113] V. Pejaver, W. L. Hsu, F. Xin, A. K. Dunker, V. N. Uversky, P. Radivojac, *Protein Sci.* **2014**, *23*, 1077.
[114] L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, A. K. Dunker, *Nucleic Acid. Res.* **2004**, *32*, 1037.
[115] P. Radivojac, V. Vacic, C. Haynes, R. R. Cocklin, A. Mohan, J. W. Heyen, M. G. Goebl, L. M. Iakoucheva, *Proteins* **2010**, *78*, 365.
[116] S. Li, L. M. Iakoucheva, S. D. Mooney, P. Radivojac, *Pac. Symp. Biocomput.* **2010**, 337.
[117] Y. Cheng, T. LeGall, C. J. Oldfield, A. K. Dunker, V. N. Uversky, *Biochemistry* **2006**, *45*, 10448.
[118] V. N. Uversky, C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biophys.* **2008**, *37*, 215.
[119] V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic, A. K. Dunker, *BMC Genomics* **2009**, *10*, S7.
[120] V. N. Uversky, *Expert Rev. Proteomics* **2010**, *7*, 543.
[121] V. N. Uversky, V. Dave, L. M. Iakoucheva, P. Malaney, S. J. Metallo, R. R. Pathak, A. C. Joerger, *Chem. Rev.* **2014**, *114*, 6844.
[122] V. N. Uversky, *Front. Mol. Biosci.* **2014**, *1*, 6.
[123] V. N. Uversky, *Front. Biosci. (Landmark Ed)* **2014**, *19*, 181.
[124] P. J. Thul, L. Akesson, M. Wiking, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, Å.

Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K. S. Lilley, M. Uhlén, E. Lundberg, *Science* **2017**, *356*, pii:eaal3321.

[125] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, *Biochim. Biophys. Acta* **2010**, *1804*, 996.

[126] X. Li, P. Romero, M. Rani, A. K. Dunker, Z. Obradovic, *Genome Inform. Ser. Workshop Genome Inform* **1999**, *10*, 30.

[127] K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, Z. Obradovic, *J. Bioinform. Comput. Biol.* **2005**, *3*, 35.

[128] K. Rajagopalan, S. M. Mooney, N. Parekh, R. H. Getzenberg, P. Kulkarni, *J. Cell. Biochem.* **2011**, *112*, 3256.

[129] K. D. Pruitt, J. Harrow, R. A. Harte, C. Wallin, M. Diekhans, D. R. Maglott, S. Searle, C. M. Farrell, J. E. Loveland, B. J. Ruef, E. Hart, M.-M. Suner, M. J. Landrum, B. Aken, S. Ayling, R. Baertsch, J. Fernandez-Banet, J. L. Cherry, V. Curwen, M. DiCuccio, M. Kellis, J. Lee, M. F. Lin, M. Schuster, A. Shkeda, C. Amid, G. Brown, O. Dukhanina, A. Frankish, J. Hart, B. L. Maidak, J. Mudge, M. R. Murphy, T. Murphy, J. Rajan, B. Rajput, L. D. Riddick, C. Snow, C. Steward, D. Webb, J. A. Weber, L. Wilming, W. Wu, E. Birney, D. Haussler, T. Hubbard, J. Ostell, R. Durbin, D. Lipman, *Genome Res.* **2009**, *19*, 1316.

[130] R. A. Harte, C. M. Farrell, J. E. Loveland, M. M. Suner, L. Wilming, B. Aken, D. Barrell, A. Frankish, C. Wallin, S. Searle, M. Diekhans, J. Harrow, K. D. Pruitt, *Database (Oxford)* **2012**, *2012*, bas008.

[131] C. M. Farrell, N. A. O'Leary, R. A. Harte, J. E. Loveland, L. G. Wilming, C. Wallin, M. Diekhans, D. Barrell, S. M. J. Searle, B. Aken, S. M. Hiatt, A. Frankish, M. Suner, B. Rajput, C. A. Steward, G. R. Brown, R. Bennett, M. Murphy, W. Wu, M. P. Kay, J. Hart, J. Rajan, J. Weber, C. Snow, L. D. Riddick, T. Hunt, D. Webb, M. Thomas, P. Tamez, S. H. Rangwala, K. M. McGarvey, S. Pujar, A. Shkeda, J. M. Mudge, J. M. Gonzalez, J. G. R. Gilbert, S. J. Trevanion, R. Baertsch, J. L. Harrow, T. Hubbard, J. M. Ostell, , K. D. Pruitt, *Nucleic Acids Res.* **2014**, *42*, D865.

[132] B. Meszaros, I. Simon, Z. Dosztanyi, *PLoS Comput. Biol.* **2009**, *5*, e1000376.

[133] Z. Dosztanyi, B. Meszaros, I. Simon, *Bioinformatics* **2009**, *25*, 2745.

[134] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, *Bioinformatics* **2005**, *21*, 3433.

[135] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, *J. Mol. Biol.* **2005**, *347*, 827.

[136] B. Xue, C. J. Oldfield, A. K. Dunker, V. N. Uversky, *FEBS Lett.* **2009**, *583*, 1469.

[137] A. Mohan, W. J. Sullivan, Jr., P. Radivojac, A. K. Dunker, V. N. Uversky, *Mol. Biosyst.* **2008**, *4*, 328.

[138] F. Huang, C. J. Oldfield, J. Meng, W.-L. Hsu, B. Xue, V. N. Uversky, P. Romero, A. K. Dunker, *Pac. Symp. Biocomput.* **2012**, 128.

[139] D. M. Mitrea, J. A. Cika, C. S. Guy, D. Ban, P. R. Banerjee, C. B. Stanley, A. Nourse, A. A. Deniz, R. W. Kriwacki, *Elife* **2016**, *5*, e13571.

[140] H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, Andrew A. Bridges, Clifford P. Brangwynne, Amy S. Gladfelter, *Mol. Cell.* **2015**, *60*, 220.

[141] P. Anderson, N. Kedersha, *Trends Biochem. Sci.* **2008**, *33*, 141.

[142] N. Gilks, N. Kedersha, M. Ayodele, L. Shen, G. Stoecklin, L. M. Dember, P. Anderson, *Mol. Biol. Cell* **2004**, *15*, 5383.

[143] J. T. Wang, J. Smith, B. C. Chen, H. Schmidt, D. Rasoloson, A. Paix, B. G. Lambrus, D. Calidas, E. Betzig, G. Seydoux, *Elife* **2014**, *3*, e04591.

[144] S. A. Fromm, J. Kamenz, E. R. Noldeke, A. Neu, G. Zocher, R. Sprangers, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 7354.

[145] I. Na, F. Meng, L. Kurgan, V. N. Uversky, *Mol. Biosyst.* **2016**, *12*, 2798.

[146] V. Csizmok, A. V. Follis, R. W. Kriwacki, J. D. Forman-Kay, *Chem. Rev.* **2016**, *116*, 6424.

[147] P. A. Albertsson, *Partition of Cell Particles and Macromolecules*, Wiley, New York **1986**.

[148] W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389.

[149] E. M. Ahmed, *J. Adv. Res.* **2015**, *6*, 105.

[150] M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, H. S. Chan, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9650.

[151] T. Mittag, L. E. Kay, J. D. Forman-Kay, *J. Mol. Recognit.* **2010**, *23*, 105.

[152] R. Rohatgi, P. Nollau, H. Y. Ho, M. W. Kirschner, B. J. Mayer, *J. Biol. Chem.* **2001**, *276*, 26448.

[153] R. Nostramo, S. N. Varia, B. Zhang, M. M. Emerson, P. K. Herman, *Mol. Cell. Biol.* **2016**, *36*, 173.

[154] R. Nostramo, P. K. Herman, *Curr. Genet.* **2016**, *62*, 503.

[155] I. M. Blasutig, L. A. New, A. Thanabalasuriar, T. K. Dayarathna, M. Goudreault, S. E. Quaggin, S. S.-C. Li, S. Gruenheid, N. Jones, T. Pawson, *Mol. Cell. Biol.* **2008**, *28*, 2035.

[156] S. E. Radford, *Trends Biochem. Sci.* **2000**, *25*, 611.

[157] T. R. Jahn, S. E. Radford, *FEBS J.* **2005**, *272*, 5962.

[158] F. Huang, C. J. Oldfield, B. Xue, W. L. Hsu, J. Meng, X. Liu, L. Shen, P. Romero, V. N. Uversky, A. Dunker, *BMC Bioinformatics* **2014**, *15*, S4.

