FEBS Letters 589 (2015) 15–22

FEBS Letters

journal homepage: www.FEBSLetters.org

Hypothesis

Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates

Vladimir N. Uversky [a,b,c,d,*], Irina M. Kuznetsova [d,e], Konstantin K. Turoverov [d,e], Boris Zaslavsky [f]

[a] Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
[b] Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russian Federation
[c] Biology Department, Faculty of Science, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia
[d] Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russian Federation
[e] St. Petersburg State Polytechnical University, St. Petersburg, Russian Federation
[f] AnalizaDx Inc., 3615 Superior Ave., Suite 4407B, Cleveland, OH 44114, USA

ARTICLE INFO

Article history:
Received 17 August 2014
Revised 10 October 2014
Accepted 19 November 2014
Available online 29 November 2014

Edited by A. Valencia

Keywords:
Intrinsically disordered proteins
Liquid–liquid phase transition
Aqueous two-phase system
Coacervate
Partitioning
Membrane-less organelles

ABSTRACT

Here, we hypothesize that intrinsically disordered proteins (IDPs) serve as important drivers of the intracellular liquid–liquid phase separations that generate various membrane-less organelles. This hypothesis is supported by the overwhelming abundance of IDPs in these organelles. Assembly and disassembly of these organelles are controlled by changes in the concentrations of IDPs, their post-translational modifications, binding of specific partners, and changes in the pH and/or temperature of the solution. Each resulting phase provides a distinct solvent environment for other solutes leading to their unequal distribution within phases. The specificity and efficiency of such partitioning is determined by the nature of the IDP(s) and defines "targeted" enrichment of specific molecules in the resulting membrane-less organelles that determines their specific activities.

1. Introduction

1.1. Membrane-less organelles and liquid–liquid phase transitions

It is well known that the space inside the cell is crowded and inhomogeneous. Recent studies clearly indicate that the cytoplasm and nucleoplasm of any cell contain various membrane-less organelles, the dynamic assemblies typically containing both RNA and protein, and known as ribonucleoprotein (RNP) granules/bodies, or RNP droplets [1]. These membrane-less organelles form via colocalization of molecules at high concentrations within a small cellular micro-domain. Examples of such organelles include PML bodies or nuclear dots, or PODs [2], perinucleolar compartment (PNC) [3], the Sam68 nuclear body (SNB) [3], paraspeckles [4], nuclear speckles or interchromatin granule clusters [5], nucleoli [6], processing bodies [7], germline P granules [8,9], Cajal bodies (CBs; [10]), centrosomes [11], and stress granules [12]. Being devoid of membrane, these organelles or bodies are highly dynamic, and their components exist in direct contact with the surrounding nucleoplasm or cytoplasm [13,14]. Many of these structures were shown to be just slightly denser than the rest of the nucleoplasm or cytoplasm [15,16]. All this suggests that although these membrane-less organelles may be considered as a different "state" of cytoplasm or nucleoplasm, their major biophysical properties are rather similar to those of the rest of the intracellular fluid [1]. Therefore, these cellular bodies being only slightly denser than the bulk intracellular fluid and being characterized by high level of internal dynamics, can be considered as liquid-droplet phases of the nucleoplasm/cytoplasm [8,12,17–20].

Another important feature that various cellular membrane-less organelles have in common is the mechanism of their formation, which is believed to be related to the intracellular phase transitions [1]. These phase transitions in aqueous media originate from the different effects of macromolecules on the structure and solvent properties of water and are related to the high concentrations of macromolecular solutes. At low concentrations of macromolecules, the solution exists as a single phase, whereas at high concentrations, phase separation occurs [21].

* Corresponding author at: VNU, Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL 33612, USA.
E-mail address: vuversky@health.usf.edu (V.N. Uversky).

http://dx.doi.org/10.1016/j.febslet.2014.11.028
0014-5793/

Table 1
Disorder content in proteins found in various membrane-less cytoplasmic and nucleoplasmic organelles. For each organelle, proteins are arranged according to the decrease in the extent of their disorder evaluated as percentage of the residues predicted to be disordered (i.e., possessing disorder scores above 0.5) by PONDR® VSL2, which is among the more accurate disorder predictors.

Protein name	UniProt ID	Number of residues	PONDR® VSL2 (% disordered residues)	MobiDB (% disordered residues)[a]	Molecular functions (GO terms)[b]
PML bodies [2]					
Speckled 100 kDa protein (Sp100)	P23497	879	77.8	61.0	Protein binding, DNA binding, transcription corepressor, transcription coactivator, protein homodimerization, chromo shadow domain binding, identical protein binding, kinase binding, protein domain specific binding, transcription factor binding
Promyelocytic leukemia protein (PML)	P29590	882	53.1	27.0	Protein binding, DNA binding, transcription coactivator, zinc ion binding, SUMO binding, ubiquitin protein ligase binding, cobalt ion binding, protein homodimerization, protein heterodimerization
Perinucleolar compartment (PNC) [90]					
Nucleolin	P19338	710	86.2	64.9	Protein binding, RNA bonding, telomeric DNA binding, poly(A) RNA binding, identical protein binding, protein C-terminus binding, nucleotide binding
KH-type splicing regulatory protein (KSRP)	Q92945	711	76.2	67.8	DNA binding, poly(A) RNA binding
Ribonucleoprotein PTB-binding 1 (Raver1)	Q8IY67	606	74.2	45.4	Poly(A) RNA binding, nucleotide binding
CUG binding protein-2 (CUG-BP2)	O95319	508	60.4	15.6	Poly(A) RNA binding, RNA binding, nucleotide binding
Ribonucleoprotein PTB-binding 2 (Raver2)	Q9HCJ3	691	54.0	25.0	Poly(A) RNA binding, nucleotide binding
CUG binding protein-1 (CUG-BP1)	Q92879	486	50.0	21.6	BRE binding, poly(A) RNA binding, nucleotide binding, protein binding, translation repressor, nucleic acid binding, mRNA binding, RNA binding
Polypyrimidine tract-binding protein (PTB)	P26599	531	40.7	18.6	Poly(A) RNA binding, protein binding, nucleotide binding, RNA binding, pre-mRNA binding, poly-pyrimidine tract binding
Polypyrimidine tract-binding protein 3 (Rod1)	O95758	552	39.3	19.0	Poly(A) RNA binding, RNA binding, nucleotide binding,
Sam68 nuclear body (SNB) [3]					
Sam68 like mammalian-1 (SML1)	Q5VWX1	349	82.5	63.6	Poly(A) RNA binding, poly(U) RNA binding, protein binding,
Src associated in mitosis 68 kDa protein (Sam68)	Q07666	443	79.7	66.1	Poly(A) RNA binding, poly(U) RNA binding, protein binding, poly(A) binding, SH3/SH2 adaptor activity, DNA binding, RNA binding,
Sam68 like mammalian-2 (SML2)	O75525	346	68.2	54.3	Poly(A) RNA binding, RNA binding, protein binding
Paraspeckles [4,91]					
Polypyrimidine tract-binding protein-associated-splicing factor or splicing factor, proline- and glutamine-rich (PSF or SFPQ)	P23246	707	79.8	74.5	Poly(A) RNA binding, core promoter binding, transcription regulation sequence-specific DNA binding, protein binding, nucleotide binding
Paraspeckle protein 1 (PSPC1)	Q8WXF1	523	75.7	52.2	Poly(A) RNA binding, core promoter binding, protein binding, nucleotide binding
Non-POU domain-containing octamer-binding protein or 54 kDa nuclear RNA- and DNA-binding protein (NONO or P54NRB)	Q15233	471	76.9	56.1	Poly(A) RNA binding, core promoter binding, protein binding, nucleotide binding, identical protein binding
Nuclear speckles or interchromatin granule clusters [5,92]					
Transformer-2 protein homolog α (TRA2A)	Q13595	282	80.5	72.3	Nucleotide binding, poly(A) RNA binding
Transformer-2 protein homolog β (TRA2B)	P62995	288	79.5	72.6	Nucleotide binding, poly(A) RNA binding, mRNA binding, protein binding
Arginine/serine-rich domains-containing splicing factor, suppressor-of-white-apricot (SFSWAP)	Q12872	951	79.2		RNA binding
Nuclear inhibitor of protein phosphatase 1 (NIPP-1)	Q12972	351	69.5	64.7	DNA binding, RNA binding, protein binding, endonuclease activity, ribonuclease E activity, protein phosphatase type 1 regulator activity, protein serine/threonine phosphatase inhibitor activity
Threonine-proline repeats-containing splicing factor 3B subunit 1 (SF3B1)	O75533	1304	39.1	29.8	Chromatin binding, protein binding, poly(A) RNA binding
Nucleoli [6]					
Ribosomal proteins, many of which are known to be highly disordered [93]					rRNA binding, structural constituent of ribosome
Processing bodies or P-bodies					
Proline-rich nuclear receptor coactivator 2 (PNRC2)	Q9NPJ4	139	97.1	81.3	Protein binding
Trinucleotide repeat-containing gene 6A protein (TNRC6A)	Q8NDV7	1,962	96.5	79.9	Protein binding, poly(A) RNA binding, nucleotide binding
Eukaryotic translation initiation factor 4E transporter (EIF4ENIF1)	Q9NRA8	985	93.3	57.5	Protein binding, poly(A) RNA binding, protein transported activity
CCR4-NOT transcription complex subunit 3 (CNOT3)	O75175	753	80.5	58.3	Protein binding
LIM domain-containing protein 1 (LIMD1)	Q9UGP4	676	69.8	50.3	Protein binding, zinc ion binding, transcription corepressor activity

Table 1 (*continued*)

Protein name	UniProt ID	Number of residues	PONDR® VSL2 (% disordered residues)	MobiDB (% disordered residues)[a]	Molecular functions (GO terms)[b]
Protein PAT1 homolog 1	Q86TB9	770	69.7	48.2	Protein binding, RNA binding, poly(A) RNA binding, poly(U) RNA binding, poly(G) binding
Nanos homolog 3 (NANOS3)	P60323	173	64.2	52.6	RNA binding, zinc ion binding
PAB-dependent poly(A)-specific ribonuclease subunit PAN3	Q58A45	886	41.6	29.1	Protein binding, ATP binding, metal ion binding, protein kinase activity
Enhancer of mRNA-decapping protein 3 (EDC3)	Q96F86	508	38.6	27.0	Protein binding, RNA binding, identical protein binding
CCR4-NOT transcription complex subunit 1 (CNOT1)	A5YKK6	2376	27.8	8.5	Protein binding, poly(A) RNA binding, estrogen receptor binding, retinoic acid receptor binding
Germline P granules [94]					
Pharynx and intestine in excess protein 1 (PIE1)	Q94131	335	90.1	37.9	DNA binding, protein binding, metal ion binding
Zinc finger protein MEX5	Q9XUB2	468	88.7	50.9	DNA binding, mRNA 3′UTR binding, metal ion binding, poly-pyrimidine tract binding, protein kinase binding, protein domain specific binding
Pseudocleavage protein NOP1	Q09314	759	84.7	55.2	Unknown[c]
Zinc finger protein MEX6	Q09436	467	75.8	52.3	DNA binding, mRNA 3′UTR binding, metal ion binding, protein kinase binding, protein domain specific binding
Ectopic P granules protein 2 (EPG2)	Q95XR4	690	75.5	15.9	Unknown[c]
ATP-dependent RNA helicase GLH-4	O76743	1156	68.1	53.8	ATP binding, JUN kinase binding, RNA helicase activity, ATP-dependent helicase activity, RNA binding, zinc ion binding
Germline survival defective-1 (GLS-1)	Q8I4M5	1052	63.8	51.3	Unknown[c]
mRNA-decapping enzyme 2 (DCP2)	O62255	786	63.2	48.9	RNA binding, manganese ion binding, m7G(5')pppN diphosphatase activity
ATP-dependent RNA helicase GLH-2	Q966L9	974	61.9	51.8	ATP binding, zinc ion binding, protein self-association, RNA binding, JUN kinase binding, RNA helicase activity, protein binding, ATP-dependent helicase activity
Dual specificity tyrosine-phosphorylation-regulated kinase MBK2	Q9XTF3	817	61.6	49.0	ATO binding, protein binding, protein kinase activity, protein tyrosine kinase activity,protein serine/threonine kinase activity, protein serine/threonine/tyrosine kinase activity
P granule abnormality protein 1 (PGL-1)	Q9TZQ3	730	50.7	29.9	Protein binding, RNA binding, protein self-association
Defective in germ line development protein 3 (GLD3)	Q95ZK7	969	50.8	41.4	Protein binding, RNA binding, protein domain specific binding
ATP-dependent RNA helicase GLH-1	P34689	767	49.0	39.5	ATP binding, protein binding, protein self-association, RNA helicase activity, DEAD/H-box RNA helicase, RNA binding, zinc ion binding, JUN kinase binding, ATP-dependent helicase activity
Fem-3 mRNA-binding factor 2 (FBF2,)	Q09312	632	44.6	19.6	mRNA 3′UTR binding
Cajal bodies (CBs; [10])					
Coilin	P38432	576	70.1	56.6	Protein binding, disulfide oxidoreductase activity, identical protein binding, protein C-terminus binding
Survival motor neuron protein (SMN)	Q16637	294	69.7	57.8	Protein binding, identical protein binding, RNA binsing
Spliceosomal SM proteins are known to be highly disordered from previous bioinformatics studies [95,96]					Protein binding, RNA binding
Centrosome [11,97]					
Pericentrin (PCNT)	O95613	3,336	93.0	35.4	Protein binding
Centrosomal protein of 152 kDa (CEP152)	O94986	1710	88.9	24.3	Protein binding, protein kinase binding
Centromere protein J (CENPJ)	Q9HC77	1338	83.5	51.2	Protein binding, protein kinase binding, protein domain specific binding, tubulin binding
A-kinase anchor protein 9 (AKAP9)	Q99996	3911	81.5	17.0	Ion channel binding, protein complex scaffold, protein binding, receptor binding
CDK5 regulatory subunit-associated protein 2 (CK5P2)	Q96SN8	1893	74.6	31.0	Calmodulin binding, protein binding, transcription regulatory region DNA binding, tubulin binding, microtubule binding, protein kinase binding
Centrosomal protein of 192 kDa (CEP192)	Q8TEP8	1941	48.1	25.7	Phosphatase binding, protein binding
Stress granules [12]					
Ataxin 2 (ATX2)	Q99700	1313	93.2	79.1	Protein binding, RNA binding, poly(A) RNA binding, protein C-terminus binding, epidermal growth factor receptor binding
RNA-binding protein fused in sarcoma (FUS)	P35637	526	90.7	86.1	DNA binding, RNA binding, nucleotide binding, protein binding, zinc ion binding, identical protein binding, poly(A) RNA binding
Eukaryotic translation initiation factor 4 gamma 1 (eIF4G1)	Q04637	1599	71.3	58.1	Poly(A) RNA binding, translation factor activity, nucleic acid binding, translation initiation factor activity, protein binding
Ras GTPase-activating protein-binding protein 2 (G3BP2)	Q9UN86	482	64.9	58.1	Nucleotide binding, poly(A) RNA binding, receptor signaling complex scaffold activity
Ras GTPase-activating protein-binding protein 1 (G3BP1)	Q13283	466	64.0	56.0	ATP binding, ATP-dependent RNA helicase activity, ATP-dependent DNA helicase activity, endonuclease activity, poly(A) RNA binding, DNA binding, mRNA binding, protein binding

(*continued on next page*)

Table 1 (*continued*)

Protein name	UniProt ID	Number of residues	PONDR® VSL2 (% disordered residues)	MobiDB (% disordered residues)[a]	Molecular functions (GO terms)[b]
TAR DNA-binding protein 43 (TDP-43)	Q13148	414	57.2	37.0	Double-stranded DNA binding, mRNA 3′-UTR binding, poly(A) RNA binding, RNA binding, identical protein binding, sequence-specific DNA binding transcription factor activity, nucleotide binding, protein binding
Eukaryotic translation initiation factor 4 gamma 2 (eIF4G2)	P78344	907	51.6	31.0	Protein binding, poly(A) RNA binding, translation factor activity, nucleic acid binding, translation initiation factor activity
Nucleolysin TIA-1 isoform p40 (TIA-1)	P31483	386	38.6	22.5	AU-rich element binding, poly(A) binding, poly(A) RNA binding, protein binding, nucleotide binding

[a] Consensus disorder content of a given protein evaluated by MobiDB (http://mobidb.bio.unipd.it/) [67]. This consensus MobiDB disorder score is based on the outputs of ten disorder predictors, such as ESpritz in its three flavors [68], IUPred in its two flavors [69], DisEMBL in two of its flavors [70], GlobPlot [71], VSL2b [72,73], and JRONN [74].
[b] Functional information is provided as characteristic Gene Ontology (GO) terms in the "Molecular function" category. This information is taken from the corresponding entry at UniProt (http://www.uniprot.org/uniprot/P23497).
[c] Protein functions are annotated as unknown if the information on "molecular function" GO terms was unavailable at the time of analysis.

1.2. Aqueous two-phase systems (ATPSes) and coacervates

Aqueous two-phase systems (ATPSes) are formed in aqueous mixtures of different water-soluble polymers, or in a solution of a single polymer and certain salt [22]. In such systems, two or more distinct aqueous phases arise with a well-defined interface. When two specific polymers, such as dextran and Ficoll, are mixed in water above certain concentrations, the mixture separates into two immiscible aqueous layers. There is a clear interfacial boundary separating two distinct aqueous-based phases, each preferentially rich in one of the polymers, with the aqueous solvent in both phases suitable for biological products [23–25]. These systems are unique in that each of the phases typically contains well over 80% water on a molal basis, and yet they are immiscible and differ in their solvent properties [23,26–31]. Phase separation is known to occur also in aqueous solutions of a single polymer in response to temperature change or salt concentration increase. Phase separation is also observed for strongly interacting polymers such as oppositely charged polyelectrolytes (e.g., mixtures of oppositely charged proteins or mixtures of positively charged proteins and negatively charged nucleic acids), which are known to form complex coacervates that represent another form of liquid–liquid phase separation constituting a dilute phase and a concentrated coacervate phase enriched in both polyelectrolytes [21,32,33]. In the process of the coacervate formation, the efficiency of interactions between the polyelectrolytes is controlled via their charge screening by dissolved salts and therefore is strongly dependent on the solution ionic strength [24].

1.3. Membrane-less organelles as ATPSes and coacervates

As mentioned above phase separation in solution of macromolecules depends on their concentration. Typical phase separation thresholds in aqueous mixtures of polymers, proteins, and polysaccharides are within a range of several weight percent of each macromolecule [21,34]. The cell cytoplasm is well known to resemble a thick soup with a very crowded environment where the concentration of macromolecules, including proteins, nucleic acids, and carbohydrates, can be as high as 400 g/L [35,36]. These intracellular solutes occupy as much as 20–30% of the total cellular volume [35,37–39], creating a crowded medium where, in general, no individual macromolecular species are present at very high concentration [39,40]. It is likely, however, that the cytoplasm of living cells may contain a range of coexisting aqueous phases that defines the known compartmentalization of living cells [41–43]. Since cytoplasm and nucleoplasm contain the large number of different biomacromolecules, the intracellular phase separation can generate multiple phases [21,41–43]. This hypothesis proposed in the 1990s [23,40] is supported by the discovery of various membrane-less organelles briefly discussed above.

2. Intrinsically disordered proteins as major drivers of physiological phase separations

2.1. Some related peculiarities of intrinsically disordered proteins

Although for a very long time it was believed that the specific functionality of a given protein is predetermined by its unique 3-D structure, evidence is rapidly accumulating now that many protein regions and even entire proteins lack stable and/or secondary structure in solution yet possess crucial biological functions [44–52]. These intrinsically disordered proteins (IDPs) and IDP regions (IDPRs) are very abundant in nature [53–57], possess the highly heterogeneous structures [58], and differ from structured globular proteins and domains at multiple levels, such as amino acid composition, sequence complexity, hydrophobicity, charge, and flexibility. For example, IDPs/IDPRs are significantly depleted in a number of order-promoting residues, such as Ile, Leu, Val, Trp, Tyr, Phe, Cys, and Asn, being substantially enriched in the disorder-promoting amino acids, such as Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys [46,59–61]. Since many IDPs and IDPRs possess extremely high net charges, might have mosaic structure with alternating regions of opposite charges, might have highly repetitive sequences, and since many of them are present at high enough concentrations, they can clearly serve as potential players in phase separation. Among admittedly limited number of proteins known to phase separate in aqueous mixtures with other proteins or polysaccharides are casein, fibrinogen, glycinin, prolamine, gliadin, legumin [62], and β-crystalline [63] – all proteins being members of the family of IDPs or hybrid proteins containing ordered domains and long IDPRs. Elastin-like polypeptides (ELPs) known to phase separate in aqueous solutions in response to the temperature increase [64] (see below) also belong to the IDP family.

2.2. Abundance of IDPs in cellular membrane-less organelles

In agreement with this hypothesis, many proteins responsible for the formation of the cytoplasmic or nucleoplasmic membrane-less organelles are in fact intrinsically disordered. This is supported by Table 1 that lists the disorder propensities of the several proteins found in various membrane-less organelles. Proteins in Table 1 are arranged according to the decrease in the extent of their disorder evaluated as percentage of the residues predicted to be disordered (i.e., possessing disorder scores above 0.5) by

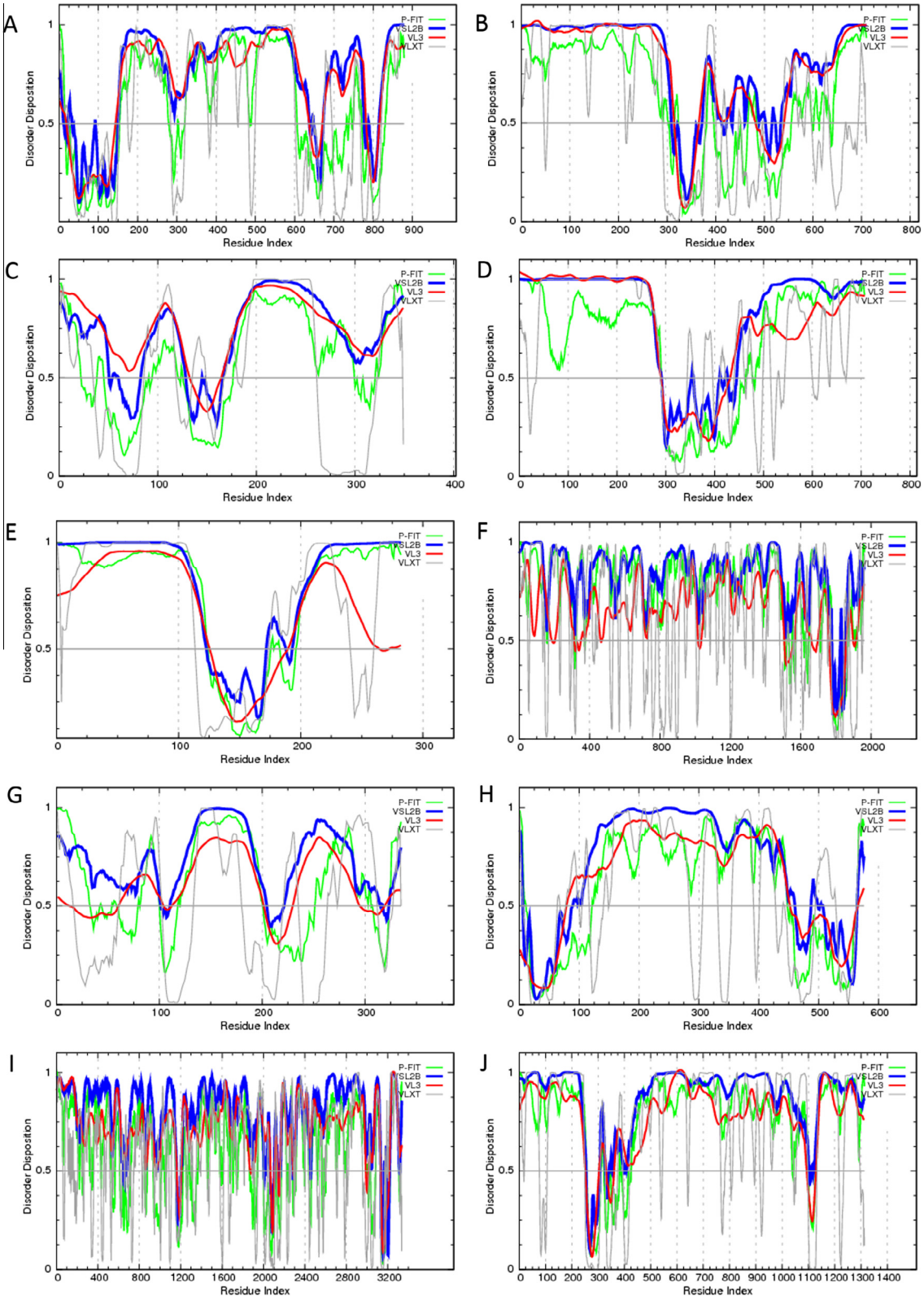

Fig. 1. Illustrative examples of highly disordered proteins found in various membrane-less organelles. (A) Speckled 100 kDa protein (P23497) from the PML bodies; (B) nucleolin (P19338) from the perinucleolar compartment; (C) Sam68 like mammalian-1 (Q5VWX1) from the Sam68 nuclear bodies; (D) polypyrimidine tract-binding protein-associated-splicing factor or splicing factor, proline- and glutamine-rich (P23246) from the paraspeckles; (E) Transformer-2 protein homolog α (Q13595) from the nuclear speckles or interchromatin granule clusters; (F) trinucleotide repeat-containing gene 6A protein (Q8NDV7) from the processing bodies or P-bodies; (G) pharynx and intestine in excess protein 1 (Q94131) from the germline P granules; (H) coilin (P38432) from the Cajal bodies; (I) pericentrin (O95613) from the centrosome; (J) ataxin 2 (Q99700) from the stress granules. Intrinsic disorder propensities are evaluated by PONDR® FIT (green lines), PONDR® VLXT (gray lines), PONDR® VSL2B (blue lines), and PONDR® VL3 (red lines). Scores above 0.5 correspond to disordered residues/regions.

PONDR® VSL2, which is among the more accurate disorder predictors. In fact, based on the comprehensive assessment of in silico predictors of intrinsic disorder [65,66], PONDR® VSL2 was shown to perform reasonably well. Table 1 also contains consensus disorder contents evaluated by MobiDB (http://mobidb.bio.unipd.it/) [67]. These consensus MobiDB disorder scores are based on the outputs of ten disorder predictors, such as ESpritz in its three flavors [68], IUPred in its two flavors [69], DisEMBL in two of its flavors [70], GlobPlot [71], VSL2b [72,73], and JRONN [74]. Table 1 clearly shows that proteins found in various cellular membrane-less organelles are extremely disordered.

Fig. 1 represents the peculiarities of distribution of intrinsic disorder propensities within the sequences of some of the constituents of these organelles. Here, for each organelle, an illustrative example is chosen as its most disordered protein. Disorder was evaluated by a family of PONDR predictors. Here, scores above 0.5 correspond to disordered residues/regions. PONDR® VSL2B is one of the most accurate stand-alone disorder predictors [75], PONDR® VL3 possesses high accuracy in finding long IDPRs [76], PONDR® VLXT is not the most accurate predictor but has high sensitivity to local sequence peculiarities which are often associated with disorder-based interaction sites [59], whereas PONDR-FIT represents a metapredictor which, being moderately more accurate than each of the component predictors, is one of the most accurate disorder predictors [77]. Fig. 1 provides further support to the notion that these proteins are excessively disordered.

It is known that functional repertoire of IDPs complements functions of ordered proteins [47,50,51,78–81]. Furthermore, comprehensive computational analysis revealed that many functional keywords found in Swiss-Prot strongly correlate with intrinsic disorder [82–84]. Among the molecular functions possessing strongest correlation with protein disorder were ribonucleoprotein, ribosomal protein, chromatin regulator, repressor, activator, and developmental protein [82]. Among the biological processes whose proteins show strongest preference for intrinsic disorder were differentiation, transcription, transcription regulation, DNA condensation, cell cycle, mRNA processing, mRNA splicing, and cell division [82]. Table 1 lists functional keywords ascribed to proteins found in various cellular membrane-less organelles and shows that vast majority of these proteins possess functions relying on intrinsic disorder.

The further proof of potential validity of the hypothesis that IDPs can serve as important players in biologically relevant liquid–liquid phase transitions comes from the analysis of the unique behavior of a series of synthetic repetitive peptides, elastin-like polypeptides (ELPs), originally derived from the mammalian elastin proteins. ELPs contain the repeats of the VPGXG pentapeptide with the "guest residue" X being any amino acid with the exception of proline. The peculiar feature of ELPs is their ability to undergo a completely reversible inverse phase transition resulting in the formation of ATPS, where an ELP converts from the structurally disordered, highly solvated conformation below the inverse transition temperature (T_t) to a new phase comprised of desolvated and aggregated polypeptides when the temperature is raised above T_t [85–88]. In addition to the strong dependence on the nature of the host residue X, the T_t of the transition depends on a polypeptide concentration and a polypeptide length, and can be triggered by temperature and the addition of chaotropic salts. Experimental analysis revealed that the ELP-based ATPSes can be generated in vivo in *Escherichia coli* and tobacco cells [89]. Furthermore, the formation of ELP-based droplets was shown to protect the content of the ELP-enriched phase from the proteolytic degradation by the preferential protease exclusion from the phase where the protein targeted for proteolysis was present [89]. The authors concluded that in the analyzed ELP-based ATPS, the overall rate of proteolysis was dramatically reduced due to the fact that cleavage reaction can only happen at the interface of the two phases [89].

2.3. Some implications of the hypothesis

Based on the above considerations combined with the ability of IDPs and hybrid proteins possessing ordered and disordered domains to be involved in a wide spectrum of weak interactions of different physico-chemical nature and with the abundant presence of IDPs in various membrane-less organelles found in cytoplasm and nucleoplasm, we hypothesize that IDPs serve as perfect candidates for the formation of these cellular ATPSes and coacervates. This hypothesis has several important outcomes related to the mechanisms of formation and disassembly of such organelles and even to the potential functional roles of such cellular ATPSes.

As far as the formation and disassembly of IDP-based cellular ATPSes are concerned, the appearance of the membrane-less organelles represents the result of a liquid–liquid phase separation. The emergence of a new liquid phase may be triggered by changes in concentration of critical ATPS-forming constituents, changes in the concentrations of specific small molecules, and changes in the pH and/or temperature of the solution. The liquid–liquid phase separations generating micrometre-sized liquid droplets in aqueous media of cyto- or nucleoplasm may be further regulated by the various posttranslational modifications of the related proteins [18], or by the specific binding of the phase forming proteins to some definite partners.

An important feature of all known ATPSes is their ability to modulate partitioning of various solutes. In fact, each phase of ATPS provides a distinct solvent environment for proteins, nucleic acids, RNPs, or other solutes. Differences in solute–solvent interactions in the two phases commonly lead to unequal solute distribution. As a result, a new liquid phase may be specifically enriched or depleted in particular solutes. Aforementioned exclusion of proteases from the ELP-containing phase represents an illustration of this phenomenon. Cellular ATPSes, such as membrane-less organelles in the *Xenopus* oocyte nucleus or germinal vesicle (GV) were shown to have a low-density structure that provides access to macromolecules from the nucleoplasm [15]. Importantly, some membrane-less organelles are known to act as liquid-phase micro-reactors where the cytoplasmic reactions are accelerated due to the increased concentrations of related RNA and protein components [1]. The aforementioned increased concentrations of the reactive components inside these organelles may result from the preferential partitioning of the related RNAs, proteins and other compounds to these IDP-based liquid phases. In other words, there is a good chance that the functionality of a given organelle is pre-determined by the nature of the IDP(s) responsible for the formation of a corresponding ATPS.

3. Concluding remarks

(a) Highly disordered proteins are very common in various cytoplasmic and neuroplasmic membrane-less organelles.
(b) These organelles are formed as a result of IDP-based liquid–liquid phase transitions.
(c) These transitions are controlled by IDP concentrations, post-translational modifications, and peculiarities of local environment.
(d) The resulting organelles are biological ATPSes or coacervates.
(e) These biological ATPSes define the inequality of solute distributions between phases, where some solutes are preferentially partitioned into the IDP-containing phase.

(f) Phase-specific enrichment in distinct solutes is driven by their preferred diffusion to a phase formed by specific IDPs.
(g) Such concentration enrichment may define the functional peculiarity of a given membrane-less organelle.

Acknowledgements

This work was supported in part by a grant from Russian Science Foundation RSCF No. 14-24-00131.

References

[1] Brangwynne, C.P. (2013) Phase transitions and size scaling of membrane-less organelles. J. Cell Biol. 203, 875–881.
[2] Maul, G.G., Negorev, D., Bell, P. and Ishov, A.M. (2000) Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. J. Struct. Biol. 129, 278–287.
[3] Huang, S. (2000) Review: perinucleolar structures. J. Struct. Biol. 129, 233–240.
[4] Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M. and Lamond, A.I. (2002) Paraspeckles: a novel nuclear domain. Curr. Biol. 12, 13–25.
[5] Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. Nat. Rev. Mol. Cell Biol. 4, 605–612.
[6] Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B.T., Patton, J.G., Singer, R.H. and Zipori, D. (2005) Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Mol. Biol. Cell 16, 2395–2413.
[7] Decker, C.J., Teixeira, D. and Parker, R. (2007) Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. J. Cell Biol. 179, 437–449.
[8] Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F. and Hyman, A.A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 324, 1729–1732.
[9] Chuma, S., Hosokawa, M., Tanaka, T. and Nakatsuji, N. (2009) Ultrastructural characterization of spermatogenesis and its evolutionary conservation in the germline: germinal granules in mammals. Mol. Cell. Endocrinol. 306, 17–23.
[10] Strzelecka, M., Trowitzsch, S., Weber, G., Luhrmann, R., Oates, A.C. and Neugebauer, K.M. (2010) Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. Nat. Struct. Mol. Biol. 17, 403–409.
[11] Decker, M., Jaensch, S., Pozniakovsky, A., Zinke, A., O'Connell, K.F., Zachariae, W., Myers, E. and Hyman, A.A. (2011) Limiting amounts of centrosome material set centrosome size in *C. elegans* embryos. Curr. Biol. 21, 1259–1267.
[12] Wippich, F., Bodenmiller, B., Trajkovska, M.G., Wanka, S., Aebersold, R. and Pelkmans, L. (2013) Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. Cell 152, 791–805.
[13] Phair, R.D. and Misteli, T. (2000) High mobility of proteins in the mammalian cell nucleus. Nature 404, 604–609.
[14] Pederson, T. (2001) Protein mobility within the nucleus–what are the right moves? Cell 104, 635–638.
[15] Handwerger, K.E., Cordero, J.A. and Gall, J.G. (2005) Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure. Mol. Biol. Cell 16, 202–211.
[16] Updike, D.L., Hachey, S.J., Kreher, J. and Strome, S. (2011) P granules extend the nuclear pore complex environment in the *C. elegans* germ line. J. Cell Biol. 192, 939–948.
[17] Brangwynne, C.P., Mitchison, T.J. and Hyman, A.A. (2011) Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. U.S.A. 108, 4334–4339.
[18] Li, P. et al. (2012) Phase transitions in the assembly of multivalent signalling proteins. Nature 483, 336–340.
[19] Aggarwal, S. et al. (2013) Myelin membrane assembly is driven by a phase transition of myelin basic proteins into a cohesive protein meshwork. PLoS Biol. 11, e1001577.
[20] Feric, M. and Brangwynne, C.P. (2013) A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. Nat. Cell Biol. 15, 1253–1259.
[21] Keating, C.D. (2012) Aqueous phase separation as a possible route to compartmentalization of biological molecules. Acc. Chem. Res. 45, 2114–2124.
[22] Ananthapadmanabhan, K.P. and Goddard, E.D. (1987) Aqueous biphase formation in polyethylene oxide-inorganic salt systems. Langmuir 3, 25–31.
[23] Zaslavsky, B. (1994) Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications, Marcel Dekker, New York.
[24] Albertsson, P.A. (1986) Partition of Cell Particles and Macromolecules, Wiley, New York.
[25] Walter, H., Brooks, D.E. and Fisher, D. (1985) Partitioning in Aqueous Two-Phase Systems: Theory, Methods, Use, and Applications to Biotechnology, Academic Press, Orlando, FL.
[26] Madeira, P.P., Reis, C.A., Rodrigues, A.E., Mikheeva, L.M. and Zaslavsky, B.Y. (2010) Solvent properties governing solute partitioning in polymer/polymer aqueous two-phase systems: nonionic compounds. J. Phys. Chem. B 114, 457–462.
[27] Ferreira, L.A., Parpot, P., Teixeira, J.A., Mikheeva, L.M. and Zaslavsky, B.Y. (2012) Effect of NaCl additive on properties of aqueous PEG-sodium sulfate two-phase system. J. Chromatogr. A 1220, 14–20.
[28] Madeira, P.P., Bessa, A., Alvares-Ribeiro, L., Aires-Barros, M.R., Reis, C.A., Rodrigues, A.E. and Zaslavsky, B.Y. (2012) Salt effects on solvent features of coexisting phases in aqueous polymer/polymer two-phase systems. J. Chromatogr. A 1229, 38–47.
[29] Madeira, P.P., Reis, C.A., Rodrigues, A.E., Mikheeva, L.M., Chait, A. and Zaslavsky, B.Y. (2011) Solvent properties governing protein partitioning in polymer/polymer aqueous two-phase systems. J. Chromatogr. A 1218, 1379–1384.
[30] Moody, M.L., Willauer, H.D., Griffin, S.T., Huddleston, J.G. and Rogers, R.D. (2005) Solvent property characterization of poly(ethylene glycol)/dextran aqueous biphasic systems using the free energy of transfer of a methylene group and a linear solvation energy relationship. Ind. Eng. Chem. Res. 44, 3749–3760.
[31] Willauer, H.D., Huddleston, J.G. and Rogers, R.D. (2002) Solvent properties of aqueous biphasic systems composed of polyethylene glycol and salt characterized by the free energy of transfer of a methylene group between the phases and by a linear solvation energy relationship. Ind. Eng. Chem. Res. 41, 2591–2601.
[32] Oparin, A.I. (1938) The Origin of Life (English Translation), Macmillan, New York.
[33] van der Gucht, J., Spruijt, E., Lemmers, M. and Cohen Stuart, M.A. (2011) Polyelectrolyte complexes: bulk phases and colloidal systems. J. Colloid Interface Sci. 361, 407–422.
[34] Tolstoguzov, V. (2000) Phase behaviour of macromolecular components in biological and food systems. Nahrung 44, 299–308.
[35] Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J. Mol. Biol. 222, 599–620.
[36] Zimmerman, S.B. and Minton, A.P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu. Rev. Biophys. Biomol. Struct. 22, 27–65.
[37] Fulton, A.B. (1982) How crowded is the cytoplasm? Cell 30, 345–347.
[38] Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. 26, 597–604.
[39] Minton, A.P. (1997) Influence of excluded volume upon macromolecular structure and associations in 'crowded' media. Curr. Opin. Biotechnol. 8, 65–69.
[40] Minton, A.P. (2000) Implications of macromolecular crowding for protein assembly. Curr. Opin. Struct. Biol. 10, 34–39.
[41] Walter, H. and Brooks, D.E. (1995) Phase separation in cytoplasm, due to macromolecular crowding, is the basis for microcompartmentation. FEBS Lett. 361, 135–139.
[42] Walter, H. (2000) Consequences of phase separation in cytoplasm. Int. Rev. Cytol. 192, 331–343.
[43] Brooks, D.E. (2000) Can cytoplasm exist without undergoing phase separation? Int. Rev. Cytol. 192, 321–330.
[44] Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321–331.
[45] Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415–427.
[46] Dunker, A.K. et al. (2001) Intrinsically disordered protein. J. Mol. Graph. Model. 19, 26–59.
[47] Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradovic, Z. (2002) Intrinsic disorder and protein function. Biochemistry 41, 6573–6582.
[48] Tompa, P. (2002) Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527–533.
[49] Uversky, V.N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? Cell. Mol. Life Sci. 60, 1852–1871.
[50] Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.
[51] Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579, 3346–3354.
[52] Turoverov, K.K., Kuznetsova, I.M. and Uversky, V.N. (2010) The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. Prog. Biophys. Mol. Biol. 102, 73–84.
[53] Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C. and Brown, C.J. (2000) Intrinsic protein disorder in complete genomes. Genome Inform. Ser. Workshop Genome Inform. 11, 161–171.
[54] Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Garner, E., Guilliot, S. and Dunker, A.K. (1998) Thousands of proteins likely to have long disordered regions. Pac. Symp. Biocomput., 437–448.
[55] Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. and Jones, D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. 337, 635–645.
[56] Xue, B., Dunker, A.K. and Uversky, V.N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J. Biomol. Struct. Dyn. 30, 137–149.
[57] Peng, Z. (2014) Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell. Mol. Life Sci.
[58] Uversky, V.N. (2013) Unusual biophysics of intrinsically disordered proteins. Biochim. Biophys. Acta 1834, 932–951.

[59] Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. and Dunker, A.K. (2001) Sequence complexity of disordered protein. Proteins 42, 38–48.
[60] Williams, R.M. et al. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac. Symp. Biocomput., 89–100.
[61] Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovic, Z., Uversky, V.N. and Dunker, A.K. (2007) Intrinsic disorder and functional proteomics. Biophys. J. 92, 1439–1456.
[62] Polyakov, V.I., Grinberg, V.Y. and Tolstoguzov, V.B. (1997) Thermodynamic incompatibility of proteins. Food Hydrocolloids 11, 171–180.
[63] Clark, J.I. and Clark, J.M. (2000) Lens cytoplasmic phase separation. Int. Rev. Cytol. 192, 171–187.
[64] Nettles, D.L., Chilkoti, A. and Setton, L.A. (2010) Applications of elastin-like polypeptides in tissue engineering. Adv. Drug Deliv. Rev. 62, 1479–1485.
[65] Peng, Z.L. and Kurgan, L. (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. Curr. Protein Pept. Sci. 13, 6–18.
[66] Fan, X. and Kurgan, L. (2014) Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. J. Biomol. Struct. Dyn. 32, 448–464.
[67] Di Domenico, T., Walsh, I., Martin, A.J. and Tosatto, S.C. (2012) MobiDB: a comprehensive database of intrinsic protein disorder annotations. Bioinformatics 28, 2080–2081.
[68] Walsh, I., Martin, A.J.M., Di Domenico, T. and Tosatto, S.C.E. (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28, 503–509.
[69] Dosztanyi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434.
[70] Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J. and Russell, R.B. (2003) Protein disorder prediction: implications for structural proteomics. Structure 11, 1453–1459.
[71] Linding, R., Russell, R.B., Neduva, V. and Gibson, T.J. (2003) GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31, 3701–3708.
[72] Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. and Dunker, A.K. (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins-Struct. Funct. Bioinf. 61, 176–182.
[73] Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K. and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics, 7.
[74] Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21, 3369–3376.
[75] Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. and Dunker, A.K. (2005) Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61 (Suppl. 7), 176–182.
[76] Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. and Dunker, A.K. (2003) Predicting intrinsic disorder from amino acid sequence. Proteins 53 (Suppl. 6), 566–572.
[77] Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. and Uversky, V.N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim. Biophys. Acta 1804, 996–1010.
[78] Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11, 739–756.
[79] Uversky, V.N. and Dunker, A.K. (2010) Understanding protein non-folding. Biochim. Biophys. Acta 1804, 1231–1264.
[80] Uversky, V.N. (2013) A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci. 22, 693–724.
[81] Oldfield, C.J. and Dunker, A.K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. Annu. Rev. Biochem. 83, 553–584.
[82] Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N. and Obradovic, Z. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J. Proteome Res. 6, 1882–1898.
[83] Vucetic, S., Xie, H., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z. and Uversky, V.N. (2007) Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. J. Proteome Res. 6, 1899–1916.
[84] Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z. and Uversky, V.N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J. Proteome Res. 6, 1917–1932.
[85] Urry, D.W. (1988) Entropic elastic processes in protein mechanisms. I. Elastic structure due to an inverse temperature transition and elasticity due to internal chain dynamics. J. Protein Chem. 7, 1–34.
[86] Urry, D.W. (1992) Free energy transduction in polypeptides and proteins based on inverse temperature transitions. Prog. Biophys. Mol. Biol. 57, 23–57.
[87] Urry, D.W. (1997) Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. J. Phys. Chem. B 101, 11007–11028.
[88] Meyer, D.E. and Chilkoti, A. (1999) Purification of recombinant proteins by fusion with thermally-responsive polypeptides. Nat. Biotechnol. 17, 1112–1115.
[89] Ge, X., Conley, A.J., Brandle, J.E., Truant, R. and Filipe, C.D. (2009) In vivo formation of protein based aqueous microcompartments. J. Am. Chem. Soc. 131, 9094–9099.
[90] Pollock, C. and Huang, S. (2010) The perinucleolar compartment. Cold Spring Harb. Perspect. Biol. 2, a000679.
[91] Bond, C.S. and Fox, A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. J. Cell Biol. 186, 637–644.
[92] Spector, D.L. and Lamond, A.I. (2011) Nuclear speckles. Cold Spring Harb. Perspect. Biol., 3.
[93] Peng, Z., Oldfield, C.J., Xue, B., Mizianty, M.J., Dunker, A.K., Kurgan, L. and Uversky, V.N. (2014) A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. Cell. Mol. Life Sci. 71, 1477–1504.
[94] Marcello, M.R. and Singson, A. (2011) Germline determination: don't mind the P granules. Curr. Biol. 21, R155–R157.
[95] Korneta, I. and Bujnicki, J.M. (2012) Intrinsic disorder in the human spliceosomal proteome. PLoS Comput. Biol. 8, e1002641.
[96] Coelho Ribeiro Mde, L. et al. (2013) Malleable ribonucleoprotein machine: protein intrinsic disorder in the *Saccharomyces cerevisiae* spliceosome. PeerJ 1, e2.
[97] Woodruff, J.B., Wueseke, O. and Hyman, A.A. (2014) Pericentriolar material structure and dynamics. Philos. Trans. R. Soc. Lond. B Biol. Sci., 369.