

Review

Protein Condensates and Protein Aggregates: *In Vitro*, in the Cell, and *In Silico*

Katja Venko^{1,*}, Eva Žerovnik^{2,3,*}¹Theory Department, National Institute of Chemistry, 1000 Ljubljana, Slovenia²Department of Biochemistry and Molecular and Structural Biology, Jožef Stefan Institute, 1000 Ljubljana, Slovenia³Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia*Correspondence: katja.venko@ki.si (Katja Venko); eva.zerovnik@ijs.si (Eva Žerovnik)

Academic Editor: Samuela Cataldi

Submitted: 23 May 2023 Revised: 25 July 2023 Accepted: 3 August 2023 Published: 28 August 2023

Abstract

Similar to other polypeptides and electrolytes, proteins undergo phase transitions, obeying physicochemical laws. They can undergo liquid-to-gel and liquid-to-liquid phase transitions. Intrinsically disordered proteins are particularly susceptible to phase separation. After a general introduction, the principles of *in vitro* studies of protein folding, aggregation, and condensation are described. Numerous recent and older studies have confirmed that the process of liquid-liquid phase separation (LLPS) leads to various condensed bodies in cells, which is one way cells manage stress. We review what is known about protein aggregation and condensation in the cell, notwithstanding the protective and pathological roles of protein aggregates. This includes membrane-less organelles and cytotoxicity of the prefibrillar oligomers of amyloid-forming proteins. We then describe and evaluate bioinformatic (*in silico*) methods for predicting protein aggregation-prone regions of proteins that form amyloids, prions, and condensates.

Keywords: neurodegeneration; protein aggregation; protein condensation; LLPS; membrane-less organelles; amyloid; intrinsically disordered proteins (IDPs); prion; prediction by *in silico* methods

1. Introduction

There are at least three states of proteins at the end-point of folding, unfolding, and misfolding equilibria: the native, the unfolded, and the amyloid state. A fourth state is sometimes quoted as a metastable molten globule intermediate. However, looking at the kinetics of unfolding by several probes, even more variants of the molten-globule intermediate are detectable [1].

Protein folding, oligomerization, misfolding, and aggregation are all determined by the primary structure, the protein sequence. However, the environment and protein concentration also play an important role and influence the final folding, the kinetics of folding, and the aggregation pathway.

In general, proteins are soluble, but under some conditions, they can undergo phase separation (liquid-liquid and liquid-gel transitions), similar to other polypeptides and electrolytes. Intrinsically disordered proteins (IDPs) are especially prone to phase separation. Together or with other biomolecules, such as RNA, they form biomolecular condensates that are important for proteostasis, compartmentalization, and regulation of the cell [2].

The aim of this review is to update and summarize the current research on the condensed and aggregated states of proteins. We consider protein aggregation a normal, albeit transient, physiological process based on physicochemical laws. The forces at play in the processes of protein folding, misfolding, and aggregation are described at the begin-

ning of this review. We describe different forms of protein aggregates and condensates as observed in cells. We then highlight the membrane-less organelles and the cytotoxicity of the prefibrillar oligomers of amyloid-forming proteins. Finally, we review and evaluate numerous *in silico* prediction programs for predicting the propensity of proteins to either condense or aggregate. Prions are discussed in a special subsection of amyloids.

2. *In Vitro* Studies of the Physicochemical Backgrounds of Amyloid Formation and Reversible Condensation of Proteins

Protein folding is a phenomenon of physical chemistry. The path that a protein takes to build a unique three-dimensional structure starting from its amino-acid sequence, i.e., the primary structure, is pre-defined by physical forces, such as hydrogen bonds, electrostatic, dipole-dipole, Van der Waals, and hydrophobic effect. Enthalpy-entropy compensation occurs when protein side chains fix and stabilize the structure. Proteins fold by following specific routes, not randomly. If a protein could attempt all possible conformations, it would take a vast, unimaginable amount of time (Levinthal's paradox [3]). Some proteins initially undergo a hydrophobic collapse, in which a hydrophobic core forms and others form some mobile elements of the secondary structure, following the framework model (leading to folding intermediates of the pre-molten globule and molten globule type and wet and dry molten



globules) [4–7]. The structure of the molten globule has long remained elusive, yet recent NMR studies are more decisive [8]. During protein folding, more specific interactions occur between the side chains. We studied an interesting example of human stefins A and B folding [9,10]. We concluded that a mixed mechanism spanning between a secondary structure governed framework model and hydrophobic collapse model with a non-native α -helix applies in this case [10].

A large number of proteins in the human proteome, nearly 40% [11], remain intrinsically disordered (IDPs) [12], also termed natively unfolded proteins [13,14]. Their folding is not dictated by hydrophobic collapse, but rather, it is directed by another molecular surface, serving as a template (template-like folding [11]). This type of protein forms a secondary structure along with binding [15]. Their interaction partners can be multiple and different IDPs can bind to a chosen partner [16].

Similar physical forces, as in folding and template-assisted folding, apply to protein-protein intermolecular interactions, leading to oligomerization and aggregation [17]. A general scheme of protein deposition (aggregation) and condensation pathways is shown in Fig. 1 (Ref. [18]) (as adapted from Vendruscolo, M. and Fuxreiter, M., 2022). Specifically, linear colloidal aggregation describes the initial events in the deposition pathway of some proteins, such as yeast Sup35, that lead to protofibrils appearing as a string of beads [19]. The mechanism of protofibril elongation is likely due to a dipole-dipole interaction between these beads [19], also referred to as “critical oligomers” [20]. However, the formation of amyloid fibrils is not a simple polymerization reaction but often involves nucleation and off-pathway aggregation [21,22]. This leads to a lag phase during which disordered intermediates refold into non-native secondary structures (β -strands), associated with the formation of oligomers [23]. Native or amyloid states can sometimes form from folding intermediates with extended (non-native) α -helices in the process of α - β structure transition [24,25]. Amyloid fibrils resulting from the reaction of nucleated polymerization are highly ordered and rigid protein states, held together by a network of backbone hydrogen bonds and by the stacking of aromatic rings [26–28] and can be of different morphologies [29].

The mechanisms of amyloid fibril formation have been extensively studied *in vitro* for both pathological and non-pathological proteins. One of the authors wrote an overview of the mechanisms of amyloid fibril formation in 2002 [30]. Others have also discussed the nucleated polymerization reaction with off-pathway intermediates [21], which has been shown for our model protein, human stefin B [22]. Since then, many other reviews have appeared. Among them is a comprehensive review by Chiti and Dobson [31], in which the authors describe, among other advances, the structures of amyloid fibrils and prefibrillar oligomers and explain the mechanisms of amyloid transfor-

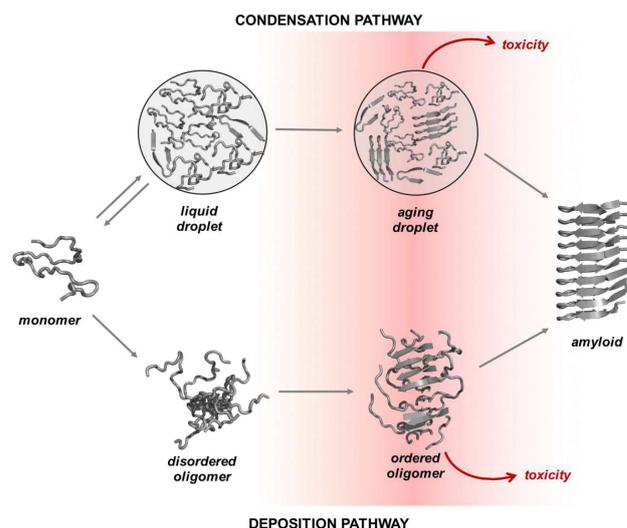


Fig. 1. Overview of the condensation and the deposition pathways for amyloid formation. Along the deposition pathway, proteins move from the native state to the amyloid state through increasingly ordered oligomer aggregates. Some of these oligomers are highly cytotoxic. Along the condensation pathway, proteins convert from the native state to the amyloid state through a dense liquid state (the droplet state). For many proteins under cellular conditions, the native and droplet states are metastable. The droplet state is functional for specific proteins, and it is stabilized by extrinsic factors, such as RNA and post-translational modifications (adapted from Vendruscolo, M. and Fuxreiter, M., 2022) [18]. This image is reproduced under the Creative Commons license: <https://creativecommons.org/licenses/by/4.0/>.

mation. They also discuss how cells combat the aberrations caused by protein aggregates. More reviews have recently been published [32–34].

The transition between the native and amyloid states of proteins can proceed via oligomeric intermediates as described above or via a condensation pathway involving liquid droplet intermediates generated through liquid-liquid phase separation [35]. Multivalent weak interactions between peptides are the driving force of phase separation. Proteins that undergo phase separation promote biomolecular condensate formation, which has a significant role in many biological processes. Further, these proteins can be divided into two categories according to their underlying driving force when forming condensates: self-assembling proteins (interacting with the same protein species and whose interactions are mediated mainly by intrinsically disordered regions) and partner-dependent proteins (interacting with different biomolecule species and whose interactions are mediated by multiple modular domains or motifs). Condensate proteome validation revealed that partner-dependent proteins are widespread in cells [36].

Similar to the amyloid transformation, the phase separation behavior of a protein is determined by general prop-

erties contained in the amino acid sequence and by environmental conditions, such as temperature and protein concentration, as well as post-translational modifications [37]. The IDPs are more prone to phase separation, even though condensation might be another generic property of proteins [38]. Cited from Vazquez *et al.* [38]: "... any sequence or conformation is susceptible to phase separation, provided that the appropriate concentration, temperature, and solvent condition are reached".

The sequences of several RNA-binding proteins comprise prion-like LCDs (low complexity domains), which are enriched in uncharged polar amino acids. Normally, these sequences are at least 60 residues long, are predicted to be intrinsically disordered, and enable the replication of a particular protein conformation from one copy to another (a template-like mechanism). LCDs are enriched in glutamine and asparagine amino acid residues. Based on sequence prediction models, the disease-linked RNA-binding proteins usually have the greatest tendency to aggregate.

Although IDPs, including prions, are thought to be more prone to undergo LLPS, structured, globular proteins may also do so under appropriate solution conditions [39]. External factors such as temperature, pH, ionic strength, shear stress, and protein concentration strongly affect the condensation of proteins. The molecular interactions that stabilize condensates at high ionic strength are mainly aromatic, hydrophobic, and nonionic interactions, whereas electrostatic interactions play a major role at lower ionic strengths [40]. Depending on the pH and ionic strength of the solution and at sufficiently high protein concentrations, the aggregates eventually form a gel-like network.

Of interest, proteins from extremophiles show an interesting shift towards intrinsic disorder [41] and, consequently, a preference for condensation over amyloid fibrillation.

3. Protein Aggregates and Condensates in the Cell

Protein misfolding in the cell occurs either in the cytoplasm or in the nucleus due to internal and external stressors, such as heat shock or oxidative stress. Misfolded proteins are prone to aggregate and are sensed by cellular defense mechanisms, such as unfolded protein response (UPR) in the endoplasmic reticulum or two degradation machineries in the cytosol. Pathological mutants, which are prone to aggregate, can exacerbate or even cause certain amyloidoses, among them neurodegenerative diseases.

Two types of protein aggregate deposits in eukaryotic (yeast) cells were previously reported, already in 2008: the IPOD and JUNQ [42,43]. The juxtannuclear inclusions harbor misfolded, still soluble proteins, which can exchange with the cytoplasmic proteins; therefore, this compartment is called "juxtannuclear protein quality control" (JUNQ). The perivacuolar peripheral inclusion contains aggregated and insoluble proteins, hence, the term "insoluble protein

deposit" (IPOD). Both compartments have certain features in common, such as the binding of the chaperone Hsp 104 and probably also Hsp 70. Only JUNQ is connected with proteasome subunits, whereas IPOD is close to autophagosomes; therefore, it is likely that the insoluble proteins get degraded by autophagy. By using electron microscopy, it has been demonstrated that JUNQ has an intranuclear localization adjacent to the nucleolus [44], and it was redefined as the INQ (intranuclear quality control compartment). The INQ serves as a deposit for both misfolded nuclear and cytosolic proteins [45]. All these aggregates are sequestered to their final location by microtubular transport, and the composition of inclusions is regulated by other regulatory proteins [42].

A third kind of inclusion exists in mammalian cells, similar to JUNQ – the aggresomes [46]. They contain fibrillar aggregates of amyloid-forming proteins, which get sequestered into the perinuclear space by the microtubule-organizing center (MTOC). The aggresomes are wrapped by a shield of intermediate filament protein vimentin [47]. All the described regulated protein aggregates are thought to exert cytoprotective functions, which are vital for cell integrity and survival [45]. The escaping soluble oligomers may bind to plasma and intracellular membranes and cause more damage by perforating them, resembling pore-forming toxins [48–50]. (See section 3.2)

Apart from protein aggregates, which contain mainly one type of protein molecule in an altered conformation, usually rich in β -structure and ending as amyloid fibrillar deposits, smaller protein puncta, i.e., condensed bodies, appear transiently upon cellular stress [51]. Albeit condensation of biomolecules has been known for some time, many new studies have appeared recently on the physicochemical process of **liquid-liquid phase separation (LLPS)**, by which condensates form from proteins and RNAs.

One of the key differences between protein aggregates and condensates is their reversibility. Protein condensates in distinction to more toxic forms of protein aggregates are at least initially reversible, as Shin *et al.* [52] have shown. When tagged with a light-sensitive tag, the proteins became condensed and later dissolved when the light was turned off. The gels were initially reversible, but over time and using a high-intensity light or high protein concentration, irreversible clumps formed [52], similar to those seen in neurodegenerative diseases.

Another key difference between protein condensates and aggregates is the specificity of the molecular interactions playing a major role. Protein aggregates of amyloid-type tend to be formed by specific interactions. The β -sheets are zipped together by a backbone of hydrogen bonds and π - π bonds between the side-chain phenylalanine rings, as well as by salt bridges between charge pairs (e.g., glutamic acid–lysine) [53]. In contrast, protein condensates are formed by more "general interactions", including electrostatic, hydrophobic, and π - π interactions, that allow dy-

dynamic assembly and disassembly. These weak interactions are less specific and involve a wide range of biomolecules, such as RNA and other proteins, which may contribute to their dynamic and diverse structures.

3.1 Membrane-Less Organelles

The intracellular biomolecular condensates, i.e., the membrane-less organelles, are important for cellular compartmentalization and regulation. They are considered protective [54] because they sequester aggregation-prone proteins and prevent amyloid formation, despite the local increase in protein concentration. They can serve as reservoirs of peptide hormones and other proteins that are released when needed.

In recent years, membrane-less organelles have been discovered a-new. Such bodies in the cytoplasm and the nucleus have been known for a long time, yet their function and mechanism of formation remained unknown. Under different forms of cellular stress, proteins and RNAs undergo reversible transitions from the liquid to the gel state, forming biomolecular condensates. Different condensates can be found from bacteria to eukaryotic and mammalian cells [55]. In eukaryotic cells, such condensed bodies exist both in the nucleus and in the cytoplasm. The two best-studied stress assemblies in the cytoplasm are the RNA-based processing bodies (Pbodies) and stress granules (SGs) that form in response to oxidative, endoplasmic reticulum (ER), osmotic and nutrient stress, UV light, etc. Stress granules (SGs) are membrane-less organelles formed in the cytoplasm by liquid-liquid phase separation (LLPS) of translationally-stalled mRNA and RNA-binding proteins, such as TDP-43. P-bodies are also composed of translationally-stalled mRNAs and proteins involved in translation repression and mRNA turnover (see, Fig. 2 (Ref. [56]) for a simplified view of the biomolecular condensates in a human cell). Apart from SGs and PBs, the RNA translation initiation complex eIF2 also forms condensed bodies in the cytoplasm [57]. Furthermore, nutrient stress (starvation) leads to the formation of a variety of cytoplasmic stress assemblies, some of which do not contain RNA, such as proteasome storage granules, metabolic enzyme bodies, and Sec bodies [57]. Sec bodies are formed by Sec16, a large scaffold protein important for secretion from ER to the Golgi. The good news is that all these entities are transient-reversible and, in most cases, pro-survival [57].

Reversible protein aggregation also occurs within the nuclei of stress-treated cells (for a schematic view, see Fig. 2). For example, mammalian cells protect thermosensitive nuclear proteins by their condensation into amyloid bodies (ABs). ABs assemble through the rapid accumulation of proteins and ribosomal RNA spacers and promote local nuclear translation during heat stress [56]. Nuclear stress bodies (nSBs) are also formed in nuclei from RNA and proteins upon heat shock. By sequestration of transcription factors, they inhibit RNA transcription [56].

As said, IDPs or proteins with intrinsically disordered regions (IDRs), such as prion and α -synuclein are most prone to phase separation. From approximately 91 cases in 2000 [58], their number increased to >1100 documented cases in 2015 [59], and there are many more IDPs and IDRs predicted. Their function is not unique and depends on the binding partners, which can be very adaptable. This makes them good candidates for regulatory and signaling functions [60].

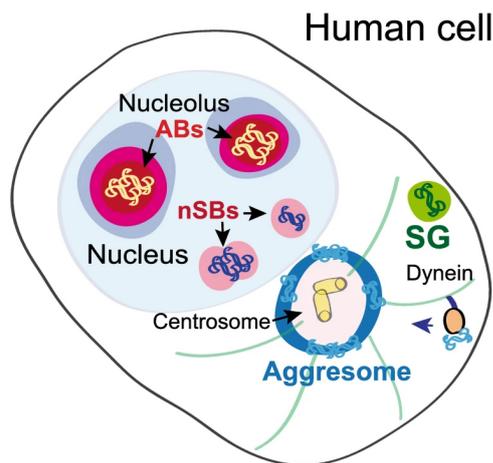
However, not only RNA-bound proteins or IDPs can form condensates upon stress. Folded globular proteins usually do not condense, yet their unfolded states can, forming the so-called unfolded protein deposits — UPODs [61]; proteins that are less stable and contain many aromatic amino acids, such as Tyr and Phe, can form UPODs [61]. Using lysozyme, a popular model protein [62], as an example, the aggregates were found to account for $\sim 10^{-4}$ of the total soluble protein. The reversible aggregates undergo a dynamic molecular exchange with the protein in solution [63]. Another example is SOD1 (superoxide dismutase). While soluble wild-type (WT) SOD1 forms a homodimer, which is stabilized by metal binding and an intra-subunit disulfide bond, a less stable mutant can unfold and is prone to form deposits [64].

3.2 Protein Oligomers Induce Cytotoxicity Upon Interaction with Membranes

The two-dimensional liquid environments provided by lipid bilayers (Fig. 3, Ref. [65]) can profoundly alter protein structure and dynamics by both specific and non-specific interactions. Kinetic and thermodynamic studies indicate that significant conformational changes can be induced in proteins encountering lipid surfaces, which can play a critical role in nucleating aggregate formation or stabilizing specific aggregation states.

Pore-forming proteins (PFP) are found in virtually all domains of life [49], and, by disrupting cell membranes, depending on the pore size, they cause ion disbalance, small molecules, or even protein efflux/influx, influencing cell signaling routes and fate. Such pore-forming proteins exist from bacteria to viruses and also shape host defense systems, including innate immunity. There is strong evidence that amyloid toxicity is also caused by prefibrillar oligomers forming amyloid pores in cellular membranes. It is believed that the prefibrillar, still soluble oligomeric intermediates would interact with cell membranes or even make the so-called “amyloid pores” [48] that exhibit structural and functional properties similar to those of pore-forming toxins.

Smaller oligomeric structures, in some cases, seem sufficient to perforate the lipid bilayer. In amyloid- β (1-42), pores with diameters of 1.7, 2.1, 2.4, and 2.9 nm were observed [66]. In our studies of the model non-pathological protein human stefin B, cytotoxicity was shown to correlate with the type and size of the aggregates [67–70]. In contrast to the insertion of amyloid- β (1-42) into lipid bilayers, toxi-



SG (Stress Granule)	Cytoplasmic
A-Bodies (Amyloid Bodies)	Nuclear / Nucleolar
nSBs (Nuclear Stress-Bodies)	Nuclear
Aggresome	Cytoplasmic (Pericentrosomal)

Fig. 2. Schematic presentation of various condensed bodies in a human cell adapted from [56]. ABs, amyloid bodies. This image is available via <http://creativecommons.org/licenses/by/4.0/>.

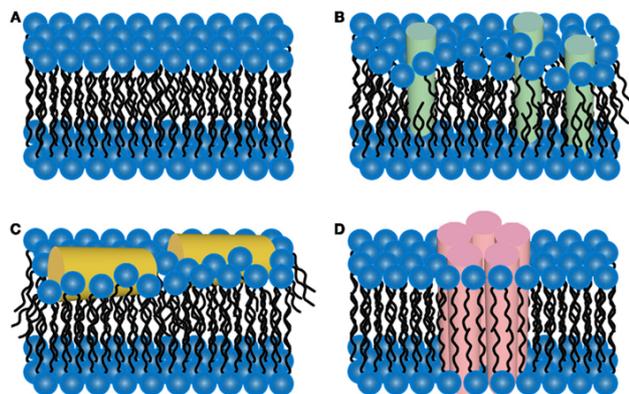


Fig. 3. Schematic representations of potential mechanisms of amyloid/lipid association. (A) A schematic representation of simplified, undisrupted lipid bilayer. This lipid bilayer structure can be perturbed by (B) membrane protein insertion or (C) association of amphiphilic α -helices with lipid surface, leading to membrane thinning and non-specific membrane leakage. (D) Many amyloid-forming proteins have been shown to form pore-like structures that can act as unregulated ion-selective channels. Reproduced from Burke *et al.* [65], *Frontiers in Neurology*. Copyright: © 2013 Burke, Yates and Legleiter. This is an open-access article distributed under the terms of the Creative Commons Attribution License.

city was higher for higher-order oligomers, i.e., 6–12 mers, >4 nm of size, in comparison to dimers or tetramers [68].

Di Scala *et al.* [50] described a common molecular mechanism of amyloid pore formation by $A\beta$ and α -synuclein (α -syn). They compared a panel of amyloid-forming fragments of the above-mentioned proteins and concluded that a two-step mechanism applies, in which gangliosides and cholesterol components of lipid membranes interact with specific structural motifs of $A\beta$ and α -syn, respectively.

The mechanism of amyloid pore formation has recently been followed by kinetic simulations [71] and single-channel measurements [72]. Kaye *et al.* [73] detected endogenous oligomeric and multimeric species in α -synucleopathies, whereas Bode *et al.* [66] showed the interaction of $A\beta$ with cellular membranes. In a *C. elegans* study, Julien *et al.* [74] showed that the membrane repair response was turned on when $A\beta$ was fed to animals, indirectly confirming the amyloid-pore hypothesis.

Both plasma and mitochondrial membranes can be affected by extracellular and intracellular prefibrillar oligomers [75]. Camilleri *et al.* [76] initially observed that mito-mimetic lipid vesicles were more permeable to oligomers of $A\beta$ (1-42), α -syn, and tau. Mito-mimetic membranes were enriched with 15% phospholipid cardiolipin (CL), which mimicked the mitochondrial inner membrane. Electrophysiological measurements displayed a twofold higher permeability in CL-enriched membranes. Recently, the same group investigated how the oligomers of $A\beta$ (1-42), α -syn, and tau interacted with isolated mitochondria and observed that all three amyloidogenic peptides, prepared as soluble oligomers, triggered a robust mitochondrial swelling, cytochrome c release and lowered the mitochondrial membrane potential [77,78]. Oligomers formed by the bacterial model amyloidogenic protein HypF-N behaved similarly [79].

4. *In Silico* Methods: Prediction Methods for Condensates and Aggregates

Based on protein sequence, environmental, and spatial factors, bioinformatic (*in silico*) methods can predict the propensity of proteins to transition to condensed or aggregated states; however, predicting the kinetics of aggregation remains a challenge.

Nowadays, *in silico* platforms provide not only single predictors but also so-called meta-predictors that gen-

erate highly accurate predictions based on algorithms that combine results from multiple computational models, including predictions of condensable, amyloidogenic, or/and intrinsically disordered regions [36]. Here, we describe and compare the state-of-the-art bioinformatic tools for *in silico* prediction of condensates and aggregates. As prediction methods are evolving rapidly, we have compiled a list of >40 commonly used tools accessible on public web servers (Support information, **Supplementary Table 1**).

Several computational methods are available to produce sequence-based predictions of the propensity of proteins to aggregate via the deposition pathway, but the physicochemical principles underlying condensation are known less well [18,80]. Although prediction methods exist to estimate the propensity of proteins to undergo liquid-liquid phase separation (LLPS), it is not clear how to predict amyloid aggregation within condensates [81]. Recently, Vendruscolo and Fuxreiter [18] have provided insights into the amino acid code for protein conversion between liquid-like and solid-like condensates. As different parameters have been proposed to determine the propensity of proteins to form condensates, various *in silico* strategies based on machine learning models, have been developed to understand the relationship between protein sequence and protein phase behavior.

4.1 Predicting Protein Condensation

In general, most *in silico* tools for predicting the propensity for condensation in two phases (liquid-liquid phase separation — LLPS) are based on amino acid sequences. The relatively small number of experimentally validated proteins prone to phase separation and the difficulty in detecting them remain a bottleneck in developing accurate predictors for condensates. Despite this limitation, new approaches are constantly being developed, and the integration of newly updated databases into their development is increasing the accuracy of the new predictors. **PScore** (<http://abragam.med.utoronto.ca/~JFKlab/Software/psp.htm>) [82], **catGRANULE** (http://service.tartaglialab.com/update_submission/277133/0b8f3740ac) [83] and **LARKS** (Low-complexity Aromatic-Rich Kinked Segments) [84] are first-generation LLPS propensity predictors. Compared to the first-generation predictors, the second-generation predictors **FuzDrop** (<https://fuzdrop.bio.unipd.it/predictor>) [85], **DeePhase** (<https://deephase.ch.cam.ac.uk/>) [86], **PSPer** (Phase Separating Protein, <https://www.bio2byte.be/b2btools/psp>) [87] and **PSPredictor** (<http://www.pkumdl.cn/PSPredictor>) [88] were developed based on larger training datasets (LLPSDB (<http://bio-comp.org.cn/llpsdb>) [89], PhaSepDB (<http://db.phasepro.pro/>) [90], PhaSePro (<https://phasepro.elte.hu/>) [91]), allowing for a broader range of LLPS protein screening. Each of the new predictors has specific properties; **DeePhase** is very

powerful in distinguishing LLPS-prone proteins from structured proteins and identifying them in the human proteome. The authors of this tool highlight that LLPS-prone proteins are more disordered, less hydrophobic, and of lower Shannon entropy [86]. **FuzDrop** (<https://fuzdrop.bio.unipd.it/predictor>) can identify droplet-promoting and aggregation-promoting regions in protein sequences that spontaneously phase separate [85]. **PSPer** prioritizes phase-separating proteins among proteins with similar RNA-binding domains, intrinsically disordered regions, and prions [87]. **PSPredictor** allows users to determine the most similar proteins in the LLPSDB under experimentally validated phase separation conditions [88]. **PSAP** (<https://github.com/vanheeringen-lab/psap>) is a random forest classifier trained on a set of 90 human proteins that condense with high confidence [81]. The LLPS predictors listed above were developed based on different underlying concepts, architectures, and training sets. This makes comparison difficult, as each method is suitable for different applications. Nevertheless, their combined use can significantly increase their utility, as reported by Pancsa *et al.* [92]. Their comparable analysis included five methods: PScore, PSPer, PLAAC, catGRANULE, and PSPredictor. By summarizing their findings, the PLAAC performs well in identifying prion-like LLPS proteins. PSPer and PScore show good synergy, as PSPer mainly detects PLDs and RNA-driven phase separation, whereas PScore detects LLPS driven by π - π stacking interactions. CatGRANULE and PSPredictor provide the truest positives and find hits missed by other methods. The novel meta-predictor **PhaSePred** (<http://predict.phasepro.pro/>) by Chen *et al.* [36] integrates several machine-learning models for predicting phase-separating proteins, including catGRANULE [83], CIDER [93], DeePhase [86], FuzDrop [85], LARKS [84], PLAAC [94], PScore [82], ZipperDB [95]. Not all tools included in **PhaSePred** are pure LLPS predictors, as PLAAC and ZipperDB predict prion-like domains and fibril-forming segments, whereas CIDER calculates various sequence parameters associated with disordered protein sequences. Unlike other currently available computational tools that preferentially predict only self-assembling proteins but perform poorly in screening partner-dependent proteins, this predictor can adequately predict self-assembling or partner-dependent protein categories [36].

Despite the abundance of existing computational tools, accurately predicting protein phase transitions remains a challenge, and the establishment of research and innovation hubs seems to be a possible future prospect. The PhasAGE project (<https://phasage.eu/>, PhasAGE – Excellence Hub on Phase Transitions in Aging and Age-Related Disorders) could be a good example of such an approach.

4.2 Predicting Aggregation Propensity

In recent decades, researchers have attempted to gain a better understanding of protein aggregation and to develop computational methods to predict aggregation propensity. It has been more difficult to predict the kinetics of aggregation. However, **AggreRATE-Pred** (<http://www.iitm.ac.in/bioinfo/aggrerate-pred/>) is the first tool to determine aggregation regions (APR prediction) and detect the change in aggregation kinetics [96]. Compared to several previous aggregation prediction models, AggreRATE-Pred considers both structural and sequence-based properties. It also predicts the change in aggregation rate upon point mutations. Compared with first-generation APR prediction methods such as **TANGO** (<http://tango.crg.es>) [97], **AGGRESKAN** (aggregation-prone segments in proteins, <http://bioinf.uab.es/aggreskan/>) [98] and **GAP** (Generalized Aggregation Proneness, <http://www.iitm.ac.in/bioinfo/GAP>) [99], the aggregation propensities determined by these methods do not correlate with the aggregation rate determined by AggreRATE-Pred [96]. The old methods only calculate the overall aggregation propensity of a polypeptide chain and do not provide information on the growth of aggregates over time. GAP deficiency is a small data set used in its development [99]; whereas the **AGGRESKAN** algorithm is simple and fast, the last implementation of the online software was performed in early 2023 [100]. In general, APRs are usually buried in the hydrophobic core of the native protein and enriched with residues that favor the formation of β -strands, contributing to increased hydrophobicity and low charge content [101]. These principles are also integrated into the **TANGO** statistical mechanics algorithm [97]. However, recent studies have shown that mutations outside APRs also affect aggregation kinetics [96].

SODA (Protein SOLubility based on Disorder and Aggregation, <http://old.protein.bio.unipd.it/soda/>) provides, in addition to aggregation propensity, information about the intrinsic disorder, hydrophobicity, and secondary structure preferences. In addition, a score to evaluate the difference in aggregation and solubility introduced by mutations can be evaluated [102]. In general, most tools for identifying APRs use amino acid residue composition and/or sequence patterns [96]. In this regard, **ANuPP** (Aggregation Nucleation Prediction in Peptides and Proteins, <https://web.iitm.ac.in/bioinfo2/ANuPP/homeseq1/>), a web meta-classifier for ARP identification, is a novelty [103]. It is unique since it is based on atom-level features and considers the diversity of aggregation mechanisms. The performance of ANuPP was evaluated on several datasets, and the results show that ANuPP is one of the best prediction methods for both the prediction of amyloidogenic hexapeptides and the identification of APRs compared with other currently available methods.

Predicting the aggregation propensity of folded proteins is a bottleneck due to the lack of known 3D structures with high resolution. Although algorithms for detect-

ing aggregation-nucleating sequences from the primary sequences of proteins work reasonably well, many of these sequences in the folded state become part of the inner core of the protein, which does not contribute to aggregation unless the protein unfolds extensively [104]. Therefore, the development of algorithms that can detect APRs at protein surfaces is of great interest and has been under constant development in recent years. These new tools, which combine structure- and sequence-based features into integrated predictors, bear improved accuracy. Such servers for aggregation propensity prediction and protein solubility engineering based on features associated with the 3D structure of proteins are **SAP** (Spatial Aggregation Propensity) [105], **Aggrescan3D** (A3D, <http://biocomp.chem.uw.edu.pl/A3D/>) [106], **Aggrescan3D 2.0** (A3D2, <http://biocomp.chem.uw.edu.pl/A3D2/>) [107], **SOLart** (<http://babylone.ulb.ac.be/SOLART/>) [108], **SolubiS** (<https://solubis.switchlab.org/>) [109], **CamSol Structurally Corrected** (<https://www-coh-software.ch.cam.ac.uk/>) [110], and **AggScore** [111].

With the advent of AlphaFold [112] and the establishment of AlphaFoldDB (<https://alphafold.ebi.ac.uk/>), the limitations due to the number of 3D protein structures identified are disappearing. Consequently, it is likely that in the next few years, we will foresee the development of many new tools for predicting the aggregation of protein 3D structures, which will enable new biomedical applications such as antibodies and beta-sheet-breaking peptides to treat diseases caused by protein aggregation [100]. In any case, the last decade has seen impressive innovations in ARP prediction. Several currently available algorithms enable an automated, sequence, and structure-based design strategy to improve the aggregation properties of proteins of scientific or industrial interest.

4.3 Predicting Amyloidogenicity

The transition of soluble proteins into insoluble amyloid fibrils is driven by specific self-propagating short-sequence segments that can be predicted from input sequences at the genomic level. In this regard, the propensity of different protein sequences to aggregate into amyloids mostly depends on the stability of the amyloid cross- β structure. Nowadays, the prediction of amyloidogenicity can be performed using various computational tools. Nevertheless, accurate prediction of amyloid-forming determinants remains a challenge [101].

AmyLoad (<http://comprec-lin.iar.pwr.edu.pl/amyloid/database/>) is a web server of amyloidogenic sequence fragments (over 1480 different entries, and continues to increase). It allows users to add their sequences to the database in FASTA format and to analyze the queried sequences with implemented amyloid predictors [113]. In addition, the updated and significantly expanded database **WALTZ-DB 2.0** (<http://waltzdb.switchlab.org/>) is now the largest freely accessible repository for determinants of amyloid fibril formation, determined experimentally based on amyloid-forming hexapeptide sequences [101].

First-generation amyloid predictors such as **FoldAmyloid** (<http://bioinfo.protres.ru/fold-amyloid/>) [114], **Waltz** [115], **SALSA** (http://amypdb.genouest.org/e107_plugins/amypdb_aggregation/db_prediction_salsa.php), are integrated into the AMYPdb database [116]. The aggregation prediction method **PASTA 2.0** (<http://protein.bio.unipd.it/pasta2/>) [117] is complemented and enriched by other information, such as intrinsic disorder and secondary structure predictions. In this regard, the amyloid-forming regions can be correctly identified with high specificity from a larger dataset of globular protein domains [117]. Further, more advanced amyloid identification methods based on machine learning approaches are available: **NetCSSP** (Neural networks for calculating Contact-dependent Secondary Structure Propensity), <http://cssp2.sookmyung.ac.kr/> [118], **FiSH Amyloid** (<http://comprec-lin.iar.pwr.edu.pl/>) [119], **AmyloGram** (<http://biongram.biotech.uni.wroc.pl/AmyloGram/>) [120], **APPNN** (Amyloid Propensity Prediction Neural Network), <https://cran.r-project.org/web/packages/appnn/index.html> [121], **BAP** (Budapest Amyloid Predictor <https://pitgroup.org/bap/>) [122] and **AmyLoad** (<http://comprec-lin.iar.pwr.edu.pl/amyload/database/>) [113]. However, meta-predictors based on a consensus approach, which combines the strength of different individual predictors into a single predictor, exceed the accuracy of these individual predictors. Such meta-predictors are **MetAmyl** and **AmylPred2**. **MetAmyl** (<http://metamyl.genouest.org>) produces a meta-prediction of sequence amyloidogenicity based on four individual predictors: Pafig, SALSA, Waltz and FoldAmyloid [123]. **AmylPred2** (http://thalis.biol.uoa.gr/AMYL_PRED2/) is an improved version of the earlier amyloid propensity prediction method (http://biophysics.biol.uoa.gr/AMYL_PRED/). It produces a consensus prediction based on 11 algorithms [124]. The method is useful for understanding the misfolding of disease proteins, and it also enables protein aggregation/solubility control in biotechnology.

Interestingly, in the 3D structures of most disease-related amyloid fibrils, the structures have been shown to contain a β -strand loop β -strand motif termed a β -arch. Accordingly, assuming that protein sequences capable of forming β -arches are amyloidogenic, a novel bioinformatics approach **ArchCandy** (<https://bioinfo.crbm.cnrs.fr/index.php?route=tools&tool=7>) was developed. Benchmark analysis demonstrated the high performance of the **ArchCandy** method [125].

4.4 Predicting Prions and Prion-Like Proteins

The first definition of prions was formulated by S.B. Prusiner as “small proteinaceous infectious particles that are resistant to inactivation by most procedures that modify nucleic acids” [126]. He posited two possible models for infectious replication of prions: either by a nucleic acid, which would be hidden and enwrapped by the pro-

tein part of the prion, or by a protein void of nucleic acid. This protein-only hypothesis was already proposed by Griffith, J. [127] and later substantiated by experimental evidence [128]. A broader definition of prions encompasses the fact that the mechanism of propagation is template-directed conformational change [129]. Prions were subsequently detected in other organisms, e.g., Sup35NM in yeast [130], and more intriguingly, several of the pathological amyloidogenic proteins, such as amyloid- β involved in Alzheimer’s disease and Tau in Parkinson’s disease, may spread in a prion-like fashion [131].

Since prions are a particular class of amyloids that can propagate their misfolded conformation and have unique compositional features, several bioinformatic tools capable of identifying novel pathological and functional polypeptides with prion-like properties have also been developed. Below, we discuss the features of several databases and algorithms that have been developed to study prions and prion-like proteins. **ZipperDB** (<https://services.mbi.ucla.edu/zipperdb/>) is a database that contains predictions of fibril-forming segments within proteins identified by the 3D profiling method [132]. This method is a unique approach that uses structural information to assess the likelihood of fibril formation of a given sequence [95]. Another interesting new database for predicting prion domains in complete proteomes is **PrionScan** (<http://webapps.bifi.es/prionscan>) [133], which has been used to understand the functions of prion/prionogenic protein and how their interaction networks substantially affect gene regulation, to identify regions driving LLPS [92] or proposed as a predictor of prion-like proteins capable of LLPS [134].

First-generation tools for predicting prion-like domains (PrLD) include **pWaltz** (<http://bioinf.uab.es/pWALTZ/>) [135], **PrionW** (<http://bioinf.uab.cat/prionw/>) [136] and **PLAAC** (Prion-Like Amino Acid Composition, <http://plaac.wi.mit.edu/>) [94]. pWaltz was originally inspired by the Waltz amyloid prediction strategy but used a lower detection threshold to identify milder amyloids and used a larger sliding window for the minimum transmissible β -fold size [135]. In addition, it is an advanced tool with the implementation of the pWaltz algorithm, enabling it to work with complete protein sequences and identify the compositional context and structural features needed for prion conversion [94,136]. **SGnn** and **AMYCO** are new and state-of-the-art predictors. **SGnn** (<http://sgnn.ppmclab.com/>) enables the recruitment of PrLDs to heat-induced SGs (stress granules) of the complete proteomes [137]. **AMYCO** (<http://bioinf.uab.es/mycov04>) provides a rapid, automated, and graphical evaluation of the impact of mutations on the aggregation properties of prion-like proteins [138]. Its performance is better than that of the first-generation predictors. The AMYCO implementation has also been used to gain insights into prion evolution [137,138].

The formation of amyloid pores by prefibrillar oligomers shares several similarities with protein toxins and

antimicrobial peptides and can also be predicted; however, this was not included in this review as it has been collected previously [139]. For space and scope reasons, we also omitted the prediction of intrinsically disordered regions of proteins, as the new meta-predictors already include these predictions in their workflows.

Abbreviations

LLPS, liquid-liquid phase separation; IDPs, intrinsically disordered proteins; MTOC, micro-tubule-organizing center; LCDs, low complexity domains; 3D, three-dimensional.

Availability of Data and Materials

The data underlying this article are available in the article and in its online supplementary material (Supplementary Table 1).

Author Contributions

EŽ—Communication with the journal, Conceptualization, Writing—original draft, Writing—review & editing. KV—Methodology of in silico methods (she is responsible and corresponding author for that part of the manuscript), Writing—review & editing. Both authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Authors thank for financial support to the Slovenian Research Agency: ARRS Grant Numbers P1-0140 (led by Boris Turk) and P1-017 (led by Marjana Novič).

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest. Given the role as Guest Editor, EŽ had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to DC.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2808183>.

References

[1] Gupta MN, Uversky VN. Pre-Molten, Wet, and Dry Molten Globules en Route to the Functional State of Proteins. *International Journal of Molecular Sciences*. 2023; 24: 2424.

- [2] Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews. Molecular Cell Biology*. 2017; 18: 285–298.
- [3] Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America*. 1992; 89: 20–22.
- [4] Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters*. 1990; 262: 20–24.
- [5] Judy E, Kishore N. A look back at the molten globule state of proteins: thermodynamic aspects. *Biophysical Reviews*. 2019; 11: 365–375.
- [6] Acharya N, Jha SK. Dry Molten Globule-Like Intermediates in Protein Folding, Function, and Disease. *The Journal of Physical Chemistry. B*. 2022; 126: 8614–8622.
- [7] Naiyer A, Hassan MI, Islam A, Sundd M, Ahmad F. Structural characterization of MG and pre-MG states of proteins by MD simulations, NMR, and other techniques. *Journal of Biomolecular Structure & Dynamics*. 2015; 33: 2267–2284.
- [8] Galano-Frutos JJ, Torreblanca R, García-Cebollada H, Sancho J. A look at the face of the molten globule: Structural model of the *Helicobacter pylori* apoflavodoxin ensemble at acidic pH. *Protein Science: a Publication of the Protein Society*. 2022; 31: e4445.
- [9] Kenig M, Jenko-Kokalj S, Tusek-Znidaric M, Pompe-Novak M, Guncar G, Turk D, *et al.* Folding and amyloid-fibril formation for a series of human stefins' chimeras: any correlation? *Proteins*. 2006; 62: 918–927.
- [10] Jelinska C, Davis PJ, Kenig M, Zerovnik E, Kokalj SJ, Gunčar G, *et al.* Modulation of contact order effects in the two-state folding of stefins A and B. *Biophysical Journal*. 2011; 100: 2268–2274.
- [11] Toto A, Malagrino F, Visconti L, Troilo F, Pagano L, Brunori M, *et al.* Templated folding of intrinsically disordered proteins. *The Journal of Biological Chemistry*. 2020; 295: 6586–6593.
- [12] Toto A, Sormanni P, Paisonni C, Uversky VN. Editorial: Intrinsically Disordered Proteins and Regions: The Challenge to the Structure-Function Relationship. *Frontiers in Molecular Biosciences*. 2022; 9: 962643.
- [13] Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Science: a Publication of the Protein Society*. 2002; 11: 739–756.
- [14] Fink AL. Natively unfolded proteins. *Current Opinion in Structural Biology*. 2005; 15: 35–41.
- [15] Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*. 2007; 447: 1021–1025.
- [16] Uversky VN. Intrinsic disorder-based protein interactions and their modulators. *Current Pharmaceutical Design*. 2013; 19: 4191–4213.
- [17] Adamcik J, Mezzenga R. Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angewandte Chemie (International Ed. in English)*. 2018; 57: 8370–8382.
- [18] Vendruscolo M, Fuxreiter M. Sequence Determinants of the Aggregation of Proteins Within Condensates Generated by Liquid-liquid Phase Separation. *Journal of Molecular Biology*. 2022; 434: 167201.
- [19] Xu S, Bevis B, Arnsdorf MF. The assembly of amyloidogenic yeast sup35 as assessed by scanning (atomic) force microscopy: an analogy to linear colloidal aggregation? *Biophysical Journal*. 2001; 81: 446–454.
- [20] Modler AJ, Gast K, Lutsch G, Damaschun G. Assembly of amyloid protofibrils via critical oligomers—a novel pathway of amyloid formation. *Journal of Molecular Biology*. 2003; 325: 135–148.
- [21] Powers ET, Powers DL. Mechanisms of protein fibril formation:

- nucleated polymerization with competing off-pathway aggregation. *Biophysical Journal*. 2008; 94: 379–391.
- [22] Skerget K, Vilfan A, Pompe-Novak M, Turk V, Waltho JP, Turk D, *et al.* The mechanism of amyloid-fibril formation by stefin B: temperature and protein concentration dependence of the rates. *Proteins*. 2009; 74: 425–436.
- [23] Frieden C. Protein aggregation processes: In search of the mechanism. *Protein Science: a Publication of the Protein Society*. 2007; 16: 2334–2344.
- [24] Konuma T, Sakurai K, Yagi M, Goto Y, Fujisawa T, Takahashi S. Highly Collapsed Conformation of the Initial Folding Intermediates of β -Lactoglobulin with Non-Native α -Helix. *Journal of Molecular Biology*. 2015; 427: 3158–3165.
- [25] Ashraf GM, Greig NH, Khan TA, Hassan I, Tabrez S, Shakil S, *et al.* Protein misfolding and aggregation in Alzheimer's disease and type 2 diabetes mellitus. *CNS & Neurological Disorders Drug Targets*. 2014; 13: 1280–1293.
- [26] Cukalevski R, Boland B, Frohm B, Thulin E, Walsh D, Linse S. Role of aromatic side chains in amyloid β -protein aggregation. *ACS Chemical Neuroscience*. 2012; 3: 1008–1016.
- [27] Stanković IM, Niu S, Hall MB, Zarić SD. Role of aromatic amino acids in amyloid self-assembly. *International Journal of Biological Macromolecules*. 2020; 156: 949–959.
- [28] Stanković IM, Božinovski DM, Brothers EN, Belić MR, Hall MB, Zarić SD. Interactions of Aromatic Residues in Amyloids: A Survey of Protein Data Bank Crystallographic Data. *Crystal Growth & Design*. 2017; 17: 6353–6362.
- [29] Taylor AIP, Staniforth RA. General Principles Underpinning Amyloid Structure. *Frontiers in Neuroscience*. 2022; 16: 878869.
- [30] Zerovnik E. Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. *European Journal of Biochemistry*. 2002; 269: 3362–3371.
- [31] Chiti F, Dobson CM. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry*. 2017; 86: 27–68.
- [32] Linse S. Mechanism of amyloid protein aggregation and the role of inhibitors. *Pure and Applied Chemistry*. 2019; 91: 211–229.
- [33] Almeida ZL, Brito RMM. Structure and Aggregation Mechanisms in Amyloids. *Molecules*. 2020; 25: 1195.
- [34] Uversky VN, Finkelstein AV. Life in Phases: Intra- and Inter-Molecular Phase Transitions in Protein Solutions. *Biomolecules*. 2019; 9: 842.
- [35] Vendruscolo M, Fuxreiter M. Protein condensation diseases: therapeutic opportunities. *Nature Communications*. 2022; 13: 5550.
- [36] Chen Z, Hou C, Wang L, Yu C, Chen T, Shen B, *et al.* Screening membraneless organelle participants with machine-learning models that integrate multimodal features. *Proceedings of the National Academy of Sciences of the United States of America*. 2022; 119: e2115369119.
- [37] Darling AL, Uversky VN. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Frontiers in Genetics*. 2018; 9: 158.
- [38] Vazquez DS, Toledo PL, Gianotti AR, Ermácora MR. Protein conformation and biomolecular condensates. *Current Research in Structural Biology*. 2022; 4: 285–307.
- [39] Vlachy V, Blanch HW, Prausnitz JM. Liquid-liquid phase separations in aqueous solutions of globular proteins. *AIChE Journal*. 1993; 39: 215–223.
- [40] Krainer G, Welsh TJ, Joseph JA, Espinosa JR, Wittmann S, de Csilléry E, *et al.* Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nature Communications*. 2021; 12: 1085.
- [41] Vicedo E, Schlessinger A, Rost B. Environmental Pressure May Change the Composition Protein Disorder in Prokaryotes. *PLoS ONE*. 2015; 10: e0133990.
- [42] Bagola K, Sommer T. Protein quality control: on IPODs and other JUNQ. *Current Biology*. 2008; 18: R1019–R1021.
- [43] Kaganovich D, Kopito R, Frydman J. Misfolded proteins partition between two distinct quality control compartments. *Nature*. 2008; 454: 1088–1095.
- [44] Miller SBM, Ho CT, Winkler J, Khokhrina M, Neuner A, Mohamed MYH, *et al.* Compartment-specific aggregases direct distinct nuclear and cytoplasmic aggregate deposition. *The EMBO Journal*. 2015; 34: 778–797.
- [45] Miller SBM, Mogk A, Bukau B. Spatially organized aggregation of misfolded proteins as cellular stress defense strategy. *Journal of Molecular Biology*. 2015; 427: 1564–1574.
- [46] Kopito RR, Sitia R. Aggresomes and Russell bodies. Symptoms of cellular indigestion? *EMBO Reports*. 2000; 1: 225–231.
- [47] Johnston JA, Ward CL, Kopito RR. Aggresomes: a cellular response to misfolded proteins. *The Journal of Cell Biology*. 1998; 143: 1883–1898.
- [48] Lashuel HA, Hartley D, Petre BM, Walz T, Lansbury PT Jr. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature*. 2002; 418: 291.
- [49] Anderlüh G, Zerovnik E. Pore formation by human stefin B in its native and oligomeric states and the consequent amyloid induced toxicity. *Frontiers in Molecular Neuroscience*. 2012; 5: 85.
- [50] Di Scala C, Yahi N, Boutemour S, Flores A, Rodriguez L, Chahinian H, *et al.* Common molecular mechanism of amyloid pore formation by Alzheimer's β -amyloid peptide and α -synuclein. *Scientific Reports*. 2016; 6: 28781.
- [51] Lee DSW, Choi CH, Sanders DW, Beckers L, Riback JA, Brangwynne CP, *et al.* Size distributions of intracellular condensates reflect competition between coalescence and nucleation. *Nature Physics*. 2023; 19: 586–596.
- [52] Shin Y, Berry J, Pannucci N, Haataja MP, Toettcher JE, Brangwynne CP. Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell*. 2017; 168: 159–171.e14.
- [53] Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102: 315–320.
- [54] Küffner AM, Linsenmeier M, Grigolato F, Prodan M, Zuccarini R, Capasso Palmiero U, *et al.* Sequestration within biomolecular condensates inhibits A β -42 amyloid formation. *Chemical Science*. 2021; 12: 4373–4382.
- [55] Fefilova AS, Fonin AV, Vishnyakov IE, Kuznetsova IM, Tur-overov KK. Stress-Induced Membraneless Organelles in Eukaryotes and Prokaryotes: Bird's-Eye View. *International Journal of Molecular Sciences*. 2022; 23: 5010.
- [56] Gallardo P, Salas-Pino S, Daga RR. Reversible protein aggregation as cytoprotective mechanism against heat stress. *Current Genetics*. 2021; 67: 849–855.
- [57] van Leeuwen W, Rabouille C. Cellular stress leads to the formation of membraneless stress assemblies in eukaryotic cells. *Traffic (Copenhagen, Denmark)*. 2019; 20: 623–638.
- [58] Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*. 2000; 41: 415–427.
- [59] DeForte S, Uversky VN. Intrinsically disordered proteins in PubMed: what can the tip of the iceberg tell us about what lies below? *RSC Advances*. 2016; 6: 11513–11521.
- [60] Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*. 2019; 7.
- [61] Ruff KM, Choi YH, Cox D, Ormsby AR, Myung Y, Ascher DB, *et al.* Sequence grammar underlying the unfolding and phase separation of globular proteins. *Molecular Cell*. 2022; 82: 3193–3208.e8.

- [62] Shiryayev A, Pagan DL, Gunton JD, (eds.) Introduction. Protein Condensation: Kinetic Pathways to Crystallization and Disease (pp. 1–8). Cambridge University Press: Cambridge. 2007.
- [63] Nikfarjam S, Jouravleva EV, Anisimov MA, Woehl TJ. Effects of Protein Unfolding on Aggregation and Gelation in Lysozyme Solutions. *Biomolecules*. 2020; 10: 1262.
- [64] Nordlund A, Leinartaite L, Saraboji K, Aisenbrey C, Gröbner G, Zetterström P, *et al.* Functional features cause misfolding of the ALS-provoking enzyme SOD1. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 9667–9672.
- [65] Burke KA, Yates EA, Legleiter J. Biophysical insights into how surfaces, including lipid membranes, modulate protein aggregation related to neurodegeneration. *Frontiers in Neurology*. 2013; 4: 17.
- [66] Bode DC, Baker MD, Viles JH. Ion Channel Formation by Amyloid- β 42 Oligomers but Not Amyloid- β 40 in Cellular Membranes. *The Journal of Biological Chemistry*. 2017; 292: 1404–1413.
- [67] Ceru S, Zerovnik E. Similar toxicity of the oligomeric molten globule state and the prefibrillar oligomers. *FEBS Letters*. 2008; 582: 203–209.
- [68] Ceru S, Kokalj SJ, Rabzelj S, Skarbot M, Gutierrez-Aguirre I, Kopitar-Jerala N, *et al.* Size and morphology of toxic oligomers of amyloidogenic proteins: a case study of human stefin B. *Amyloid: the International Journal of Experimental and Clinical Investigation: the Official Journal of the International Society of Amyloidosis*. 2008; 15: 147–159.
- [69] Rabzelj S, Viero G, Gutiérrez-Aguirre I, Turk V, Dalla Serra M, Anderluh G, *et al.* Interaction with model membranes and pore formation by human stefin B: studying the native and prefibrillar states. *The FEBS Journal*. 2008; 275: 2455–2466.
- [70] Anderluh G, Gutiérrez-Aguirre I, Rabzelj S, Ceru S, Kopitar-Jerala N, Macek P, *et al.* Interaction of human stefin B in the prefibrillar oligomeric form with membranes. Correlation with cellular toxicity. *The FEBS Journal*. 2005; 272: 3042–3051.
- [71] Shah SI, Demuro A, Ullah G. Modeling the kinetics of amyloid beta pores and long-term evolution of their Ca²⁺ toxicity. *bioRxiv*. 2022. (preprint)
- [72] Ferreira C, Couceiro J, Tenreiro S, Quintas A. A biophysical perspective on the unexplored mechanisms driving Parkinson's disease by amphetamine-like stimulants. *Neural Regeneration Research*. 2021; 16: 2213–2214.
- [73] Kaye R, Dettmer U, Lesné SE. Soluble endogenous oligomeric α -synuclein species in neurodegenerative diseases: Expression, spreading, and cross-talk. *Journal of Parkinson's Disease*. 2020; 10: 791–818.
- [74] Julien C, Tomberlin C, Roberts CM, Akram A, Stein GH, Silverman MA, *et al.* *In vivo* induction of membrane damage by β -amyloid peptide oligomers. *Acta Neuropathologica Communications*. 2018; 6: 131.
- [75] Vassallo N. Amyloid pores in mitochondrial membranes. *Neural Regeneration Research*. 2021; 16: 2225–2226.
- [76] Camilleri A, Zarb C, Caruana M, Ostermeier U, Ghio S, Högen T, *et al.* Mitochondrial membrane permeabilisation by amyloid aggregates and protection by polyphenols. *Biochimica et Biophysica Acta*. 2013; 1828: 2532–2543.
- [77] Ghio S, Camilleri A, Caruana M, Ruf VC, Schmidt F, Leonov A, *et al.* Cardiolipin Promotes Pore-Forming Activity of Alpha-Synuclein Oligomers in Mitochondrial Membranes. *ACS Chemical Neuroscience*. 2019; 10: 3815–3829.
- [78] Camilleri A, Ghio S, Caruana M, Weckbecker D, Schmidt F, Kamp F, *et al.* Tau-induced mitochondrial membrane perturbation is dependent upon cardiolipin. *Biochimica et Biophysica Acta. Biomembranes*. 2020; 1862: 183064.
- [79] Farrugia MY, Caruana M, Ghio S, Camilleri A, Farrugia C, Cauchi RJ, *et al.* Toxic oligomers of the amyloidogenic HypF-N protein form pores in mitochondrial membranes. *Scientific Reports*. 2020; 10: 17733.
- [80] Vernon RM, Forman-Kay JD. First-generation predictors of biological protein phase separation. *Current Opinion in Structural Biology*. 2019; 58: 88–96.
- [81] van Mierlo G, Jansen JRG, Wang J, Poser I, van Heeringen SJ, Vermeulen M. Predicting protein condensate formation using machine learning. *Cell Reports*. 2021; 34: 108705.
- [82] Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife*. 2018; 7: e31486.
- [83] Bolognesi B, Lorenzo Gotor N, Dhar R, Cirillo D, Baldrighi M, Tartaglia GG, *et al.* A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Reports*. 2016; 16: 222–231.
- [84] Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, Cascio D, *et al.* Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science (New York, N.Y.)*. 2018; 359: 698–701.
- [85] Hardenberg M, Horvath A, Ambrus V, Fuxreiter M, Vendruscolo M. Widespread occurrence of the droplet state of proteins in the human proteome. *Proceedings of the National Academy of Sciences of the United States of America*. 2020; 117: 33254–33262.
- [86] Saar KL, Morgunov AS, Qi R, Arter WE, Krainer G, Lee AA, *et al.* Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proceedings of the National Academy of Sciences of the United States of America*. 2021; 118: e2019053118.
- [87] Orlando G, Raimondi D, Tabaro F, Codicè F, Moreau Y, Vranken WF. Computational identification of prion-like RNA-binding proteins that form liquid phase-separated condensates. *Bioinformatics (Oxford, England)*. 2019; 35: 4617–4623.
- [88] Chu X, Sun T, Li Q, Xu Y, Zhang Z, Lai L, *et al.* Prediction of liquid-liquid phase separating proteins using machine learning. *BMC Bioinformatics*. 2022; 23: 72.
- [89] Li Q, Peng X, Li Y, Tang W, Zhu J, Huang J, *et al.* LLPSDB: a database of proteins undergoing liquid-liquid phase separation *in vitro*. *Nucleic Acids Research*. 2020; 48: D320–D327.
- [90] You K, Huang Q, Yu C, Shen B, Sevilla C, Shi M, *et al.* PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Research*. 2020; 48: D354–D359.
- [91] Mészáros B, Erdős G, Szabó B, Schád É, Tantos Á, Abukhairan R, *et al.* PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Research*. 2020; 48: D360–D367.
- [92] Pancsa R, Vranken W, Mészáros B. Computational resources for identifying and describing proteins driving liquid-liquid phase separation. *Briefings in Bioinformatics*. 2021; 22: bbaa408.
- [93] Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical Journal*. 2017; 112: 16–21.
- [94] Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics (Oxford, England)*. 2014; 30: 2501–2502.
- [95] Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107: 3487–3492.
- [96] Rawat P, Prabakaran R, Kumar S, Gromiha MM. AggreRATE-Pred: a mathematical model for the prediction of change in aggregation rate upon point mutation. *Bioinformatics (Oxford, England)*. 2020; 36: 1439–1444.

- [97] Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology*. 2004; 22: 1302–1306.
- [98] Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics*. 2007; 8: 65.
- [99] Thangakani AM, Kumar S, Nagarajan R, Velmurugan D, Gromiha MM. GAP: towards almost 100 percent prediction for β -strand-mediated aggregating peptides with distinct morphologies. *Bioinformatics (Oxford, England)*. 2014; 30: 1983–1990.
- [100] Pintado-Grima C, Bárcenas O, Bartolomé-Nafria A, Fornt-Suñé M, Iglesias V, Garcia-Pardo J, *et al*. A Review of Fifteen Years Developing Computational Tools to Study Protein Aggregation. 2023; 3: 1–20.
- [101] Louros N, Orlando G, De Vleeschouwer M, Rousseau F, Schymkowitz J. Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities. *Nature Communications*. 2020; 11: 3314.
- [102] Paladín L, Piovesan D, Tosatto SCE. SODA: prediction of protein solubility from disorder and aggregation propensity. *Nucleic Acids Research*. 2017; 45: W236–W240.
- [103] Prabakaran R, Rawat P, Kumar S, Gromiha MM. Evaluation of *in silico* tools for the prediction of protein and peptide aggregation on diverse datasets. *Briefings in Bioinformatics*. 2021; 22: bbab240.
- [104] Graña-Montes R, Ventura S. Protein Aggregation and Its Prediction. In Scapin G, Patel D, Arnold E (eds.) *Multifaceted Roles of Crystallography in Modern Drug Discovery* (pp. 115–127). Springer Netherlands: Dordrecht. 2015.
- [105] Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL. Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 11937–11942.
- [106] Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Research*. 2015; 43: W306–W313.
- [107] Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. *Nucleic Acids Research*. 2019; 47: W300–W307.
- [108] Hou Q, Kwasigroch JM, Rooman M, Pucci F. SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics (Oxford, England)*. 2020; 36: 1445–1452.
- [109] Van Durme J, De Baets G, Van Der Kant R, Ramakers M, Ganevan A, Wilkinson H, *et al*. Solubis: a webserver to reduce protein aggregation through mutation. *Protein Engineering, Design & Selection: PEDS*. 2016; 29: 285–289.
- [110] Sormanni P, Amery L, Ekizoglou S, Vendruscolo M, Popovic B. Rapid and accurate *in silico* solubility screening of a monoclonal antibody library. *Scientific Reports*. 2017; 7: 8200.
- [111] Sankar K, Krystek SR Jr, Carl SM, Day T, Maier JKX. AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins*. 2018; 86: 1147–1156.
- [112] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596: 583–589.
- [113] Wozniak PP, Kotulska M. AmyLoad: website dedicated to amyloidogenic protein fragments. *Bioinformatics (Oxford, England)*. 2015; 31: 3395–3397.
- [114] Garbuzynskiy SO, Lobanov MY, Galzitskaya OV. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics (Oxford, England)*. 2010; 26: 326–332.
- [115] Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, *et al*. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Methods*. 2010; 7: 237–242.
- [116] Pawlicki S, Le Béhec A, Delamarche C. AMYPdb: a database dedicated to amyloid precursor proteins. *BMC Bioinformatics*. 2008; 9: 273.
- [117] Walsh I, Seno F, Tosatto SCE, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Research*. 2014; 42: W301–W307.
- [118] Kim C, Choi J, Lee SJ, Welsh WJ, Yoon S. NetCSSP: web application for predicting chameleon sequences and amyloid fibril formation. *Nucleic Acids Research*. 2009; 37: W469–W473.
- [119] Gasior P, Kotulska M. FISH Amyloid - a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids. *BMC Bioinformatics*. 2014; 15: 54.
- [120] Burdukiewicz M, Sobczyk P, Rödiger S, Duda-Madej A, Mackiewicz P, Kotulska M. Amyloidogenic motifs revealed by n-gram analysis. *Scientific Reports*. 2017; 7: 12961.
- [121] Família C, Dennison SR, Quintas A, Phoenix DA. Prediction of Peptide and Protein Propensity for Amyloid Formation. *PLoS ONE*. 2015; 10: e0134679.
- [122] Keresztes L, Szögi E, Varga B, Farkas V, Perczel A, Grolmusz V. The Budapest Amyloid Predictor and Its Applications. *Biomolecules*. 2021; 11: 500.
- [123] Emily M, Talvas A, Delamarche C. MetAmyl: a METa-predictor for AMYLoid proteins. *PLoS ONE*. 2013; 8: e79722.
- [124] Tsoilis AC, Papandreou NC, Iconomidou VA, Hamodrakas SJ. A consensus method for the prediction of ‘aggregation-prone’ peptides in globular proteins. *PLoS ONE*. 2013; 8: e54175.
- [125] Ahmed AB, Znassi N, Château MT, Kajava AV. A structure-based approach to predict predisposition to amyloidosis. *Alzheimer’s & Dementia: the Journal of the Alzheimer’s Association*. 2015; 11: 681–690.
- [126] Prusiner SB. Novel proteinaceous infectious particles cause scrapie. *Science (New York, N.Y.)*. 1982; 216: 136–144.
- [127] Griffith JS. Self-replication and scrapie. *Nature*. 1967; 215: 1043–1044.
- [128] Soto C, Castilla J. The controversial protein-only hypothesis of prion propagation. *Nature Medicine*. 2004; 10: S63–S37.
- [129] Fraser PE. Prions and prion-like proteins. *The Journal of Biological Chemistry*. 2014; 289: 19839–19840.
- [130] Krammer C, Kryndushkin D, Suhre MH, Kremmer E, Hofmann A, Pfeifer A, *et al*. The yeast Sup35NM domain propagates as a prion in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 462–467.
- [131] Scialò C, De Cecco E, Manganotti P, Legname G. Prion and Prion-Like Protein Strains: Deciphering the Molecular Basis of Heterogeneity in Neurodegeneration. *Viruses*. 2019; 11: 261.
- [132] Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, Eisenberg D. The 3D profile method for identifying fibril-forming segments of proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103: 4074–4078.
- [133] Espinosa Angarica V, Angulo A, Giner A, Losilla G, Ventura S, Sancho J. PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genomics*. 2014; 15: 102.
- [134] Maziuk B, Ballance HI, Wolozin B. Dysregulation of RNA Binding Protein Aggregation in Neurodegenerative Disorders. *Frontiers in Molecular Neuroscience*. 2017; 10: 89.
- [135] Sabate R, Rousseau F, Schymkowitz J, Ventura S. What makes a protein sequence a prion? *PLoS Computational Biology*. 2015;

11: e1004013.

- [136] Zambrano R, Conchillo-Sole O, Iglesias V, Illa R, Rousseau F, Schymkowitz J, *et al.* PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Research*. 2015; 43: W331–W337.
- [137] Iglesias V, Santos J, Santos-Suárez J, Pintado-Grima C, Ventura S. SGnn: A Web Server for the Prediction of Prion-Like Domains Recruitment to Stress Granules Upon Heat Stress. *Frontiers in Molecular Biosciences*. 2021; 8: 718301.
- [138] Iglesias V, Conchillo-Sole O, Batlle C, Ventura S. AMYCO: evaluation of mutational impact on prion-like proteins aggregation propensity. *BMC Bioinformatics*. 2019; 20: 24.
- [139] Venko K, Novič M, Stoka V, Žerovnik E. Prediction of Transmembrane Regions, Cholesterol, and Ganglioside Binding Sites in Amyloid-Forming Proteins Indicate Potential for Amyloid Pore Formation. *Frontiers in Molecular Neuroscience*. 2021; 14: 619496.