

Review

# A Review of Fifteen Years Developing Computational Tools to Study Protein Aggregation

Carlos Pintado-Grima <sup>1</sup>, Oriol Bárcenas <sup>1</sup>, Andrea Bartolomé-Nafria <sup>1</sup>, Marc Fornt-Suñé <sup>1</sup>,  
Valentín Iglesias <sup>1,2</sup>, Javier Garcia-Pardo <sup>1,\*</sup> and Salvador Ventura <sup>1,\*</sup>

<sup>1</sup> Departament de Bioquímica i Biologia Molecular, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

<sup>2</sup> Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, Universitat de Barcelona, Rosselló 149-153, 08036 Barcelona, Spain

\* Correspondence: javiergarciapardo@msn.com (J.G.-P.); salvador.ventura@uab.cat (S.V.); Tel.: +34-93-586-8936 (J.G.-P.); +34-93-586-8956 (S.V.)

**Abstract:** The presence of insoluble protein deposits in tissues and organs is a hallmark of many human pathologies. In addition, the formation of protein aggregates is considered one of the main bottlenecks to producing protein-based therapeutics. Thus, there is a high interest in rationalizing and predicting protein aggregation. For almost two decades, our laboratory has been working to provide solutions for these needs. We have traditionally combined the core tenets of both bioinformatics and wet lab biophysics to develop algorithms and databases to study protein aggregation and its functional implications. Here, we review the computational toolbox developed by our lab, including programs for identifying sequential or structural aggregation-prone regions at the individual protein and proteome levels, engineering protein solubility, finding and evaluating prion-like domains, studying disorder-to-order protein transitions, or categorizing non-conventional amyloid regions of polar nature, among others. In perspective, the succession of the tools we describe illustrates how our understanding of the protein aggregation phenomenon has evolved over the last fifteen years.

**Keywords:** protein aggregation; bioinformatics; biophysics; computational tools; amyloid; protein structure; protein folding



**Citation:** Pintado-Grima, C.; Bárcenas, O.; Bartolomé-Nafria, A.; Fornt-Suñé, M.; Iglesias, V.; Garcia-Pardo, J.; Ventura, S. A Review of Fifteen Years Developing Computational Tools to Study Protein Aggregation. *Biophysica* **2023**, *3*, 1–20. <https://doi.org/10.3390/biophysica3010001>

Academic Editor: Jaume Casademunt

Received: 30 December 2022

Revised: 11 January 2023

Accepted: 16 January 2023

Published: 18 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

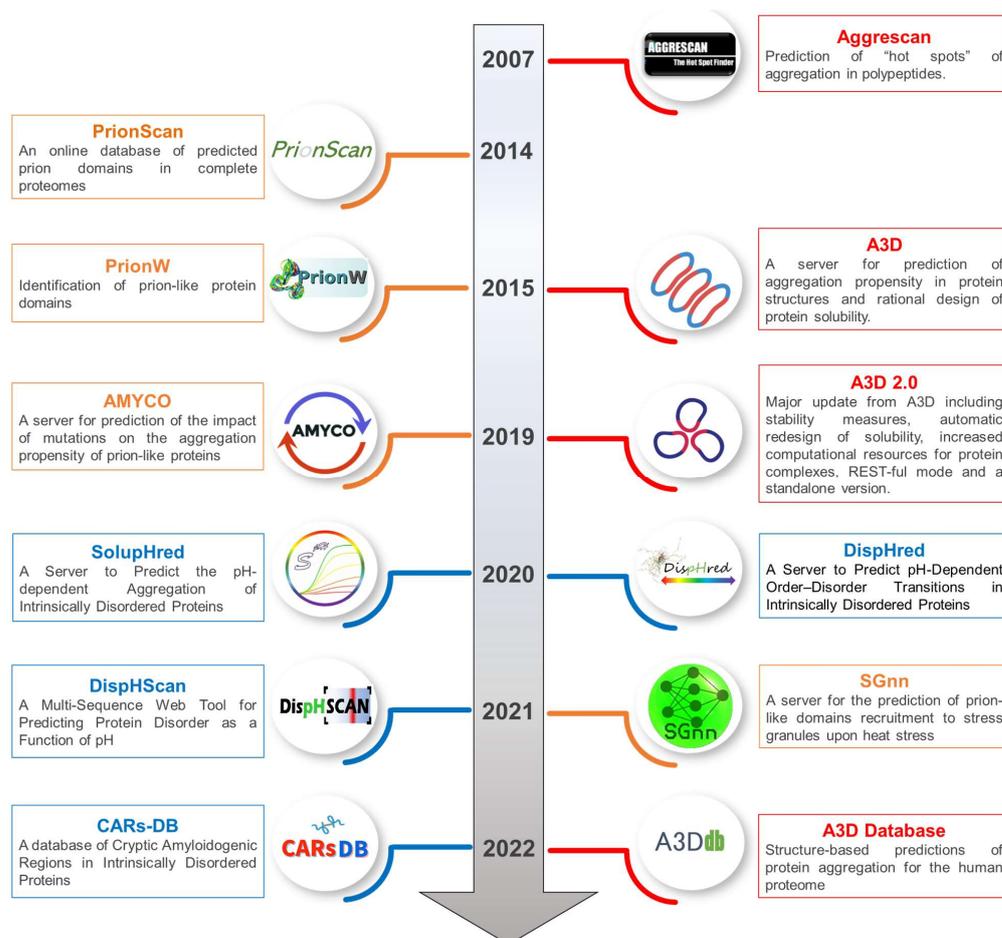
Proteins are prevalent macromolecules in living organisms and are essential to most biological functions. The establishment of functional native intra- and interchain interactions is a key feature of protein biology that controls protein folding, binding, and activity [1,2]. Proteins navigate several conformational states in the crowded environment of living cells in pursuit of a free-energy minimum, which can correspond to a monomeric state or a wide range of assemblies [3,4]. Intracellular assemblies come in a variety of forms, from multi-component, dynamic, and reversible biomolecular condensates to irreversible protein clumps. In this latter case, the original protomers undergo partial or global unfolding, and native contacts are replaced by non-native intermolecular interactions [5–7] resulting in the formation of non-structured amorphous aggregates or highly ordered amyloid fibrils characterized by a cross- $\beta$  conformation [8,9].

A wide range of pathologies, including neurodegenerative diseases such as Alzheimer's and Parkinson's and nonneuronal localized or systemic amyloidoses, are all closely linked to the formation of protein aggregates [10–13]. However, despite this association with disease, aggregation propensity is a general property of polypeptide chains [14]. This results from the physicochemical requirements to form native interactions overlapping with the molecular determinants driving aggregation [15,16]. Therefore, aggregation-prone regions are ubiquitous in proteins and preserved throughout proteomes over millions of years of evolution [17]. Under this constant and inevitable pressure of aggregation, proteins

have evolved to adjust their solubilities to the maximum necessary to function in their natural context [18].

The protein quality control system continuously monitors the balance between protein aggregation and solubility in vivo [19,20]. However, events such as genetic mutations [21,22], post-translational modifications [23], or the breakdown of proteostasis [24] place proteins out of their usual environment and favor the initiation of non-native contacts leading to aggregation. A similar situation is faced during the biotechnological production, purification, and storage of proteins [25–27], where they are exposed to solvent conditions divergent from those in the cell at concentrations that are many orders of magnitude higher than their biological levels [28]. Proteins have not evolved to be soluble in these conditions [29,30]; as a result, they precipitate in tissues in human disorders or during the development of therapeutic proteins.

Rationalization of the causes underlying protein aggregation prompted the development of tools that can predict protein solubility, diagnose the impact of mutations or chemical modifications in disease, and assist the engineering of optimized protein-based drugs. In 2007, we developed Aggrescan, one of the first protein aggregation prediction algorithms [31]. As our lab has a combined theoretical/experimental character, one of the features of this pioneering program was that it exploited validated biophysical data to generate its aggregation propensity scale [32]. This has been a constant of the different tools we have developed along the years, in a journey that has taken us from identifying the regions responsible for aggregation in individual protein sequences [33] to describing the aggregation properties of the totality of human protein structures [34] (Figure 1).



**Figure 1.** Overview of computational tools developed during the last 15 years by Ventura's Lab. Red squares indicate aggregation-related predictors and databases, orange refer to prion-like domain resources and blue is indicative of tools dedicated to study intrinsically disordered proteins.

The present review attempts to illustrate the contribution of this Spanish group to the prediction of protein aggregation, describing the characteristics of the different developed algorithms and databases and providing our view on the state of the art of this field.

## 2. Computational Tools to Study Protein Aggregation

The present section provides a brief overview of three different computational tools specifically developed to predict aggregation propensity from polypeptide sequences, identify aggregation-prone regions in globular proteins, and evaluate mutations' impact on aggregation.

### 2.1. Aggrescan: Prediction of "Hot Spots" of Aggregation in Polypeptides

In 2007, we developed Aggrescan [31], a web-based software that provides a tool for predicting aggregation-prone regions in polypeptide sequences. Aggrescan implements an aggregation propensity scale for natural amino acids derived from *in vivo* experiments performed with the  $\beta$ -amyloid peptide [32]. Precisely, we used a model consisting in the A $\beta$ 42 peptide fused to the green fluorescence protein (GFP), in which GFP fluorescence acts as a reporter of aggregation of the fusion protein. We mutated the middle position (Phe19) of the central hydrophobic cluster of A $\beta$ 42 for the other 19 possible natural amino acidic residues and analyzed the fluorescence levels of the different A $\beta$ 42-GFP fusions upon expression in *Escherichia coli*. As a result, variants with increased aggregation propensity decreased the fluorescence levels as they interfered with the proper folding of the fluorescent protein. The derived experimental data were employed to parametrize the algorithm behind Aggrescan [31–33,35].

Thus, Aggrescan assumes that protein aggregation is nucleated and driven by specific short sequence stretches that are exposed to the solvent, known as aggregation-prone regions (APRs) or "hot spots". The latest Aggrescan implementation is available online (<http://bioinf.uab.es/aggrescan/> (accessed on 12 January 2023)) and allows users to evaluate single or multiple protein sequences. In both cases, the polypeptide sequence/s must be provided in FASTA format as input. Then, the program determines the aggregation propensity values for each individual amino acid, based on the experimentally derived scale, generating an aggregation profile where APRs can be identified. It also calculates the average score of the sequence using a sliding window; this value provides an estimation of the overall aggregation propensity of the protein of interest.

Aggrescan is a simple and fast algorithm incorporated into different protein aggregation and stability prediction pipelines. In particular, in 2013, it was implemented in AMYLPRED2, a consensus predictor that integrates analysis performed by 11 top-tier tools to identify aggregation-prone regions in proteins [36]. Ultimately, Aggrescan has been used for many different experimental applications, including the characterization of individual biomedically relevant human proteins and their mutants [37–39] or the comparative analysis of the aggregation propensity of entire proteomes [40,41].

### 2.2. Aggrescan 3D: A Server for Prediction of Aggregation Propensity in Protein Structures and Rational Design of Protein Solubility

The establishment of intermolecular contacts driven by solvent-exposed APRs has shown to be a successful concept in predicting protein aggregation in the context of newly formed proteins or IDPs. However, for folded proteins, the detected APRs are usually located within hydrophobic cores, inaccessible regions, or highly stable secondary structures, whose exposure or  $\beta$ -sheet conversion is thermodynamically prevented [42,43]. Typically, globular proteins aggregate by the spatial clustering of often non-contiguous in sequence hydrophobic amino acids in the protein surface, forming structural APRs (STAPs) [44], by local or global structural destabilization [45] or by stochastic fluctuations that lead to the exposure of previously buried APRs [46]. Therefore, weighting a protein's spatial context becomes necessary to understand the forces that lead to its aggregation, a task that sequence-based prediction methods cannot undertake.

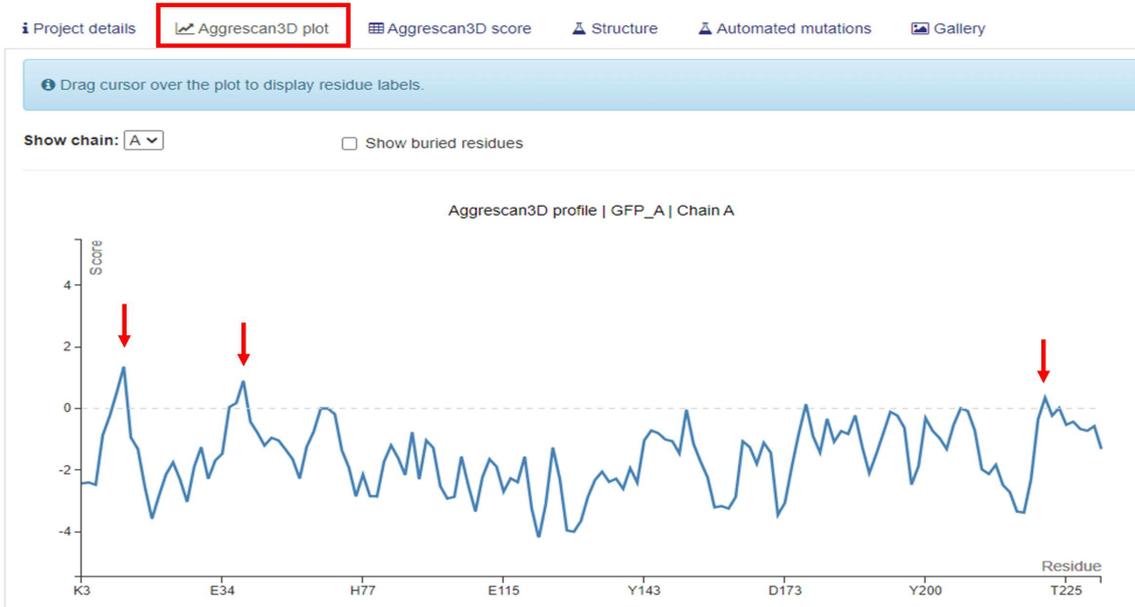
To overcome these limitations, in 2015, we developed the Aggrescan 3D (A3D) algorithm (<http://biocomp.chem.uw.edu.pl/A3D2/> (accessed on 12 January 2023)) [47]. A3D makes use of Aggrescan's aggregation propensity scale and projects it into a protein structure three-dimensional context. This novel algorithm modulates each residue's aggregation propensity by accounting for its surface exposure and summing the contributions from proximal residues' (i.e., at 5 Å or 10 Å radius) intrinsic aggregation propensity, distance, and exposed area, while disregarding non-exposed amino acids' contribution [47]. This makes accessible the study of protein aggregation using structural models for non-experts and significantly reduces the number of false positive hits compared to lineal prediction methods.

The first A3D implementation was equipped with FoldX energy force field to calculate the structural impact upon mutations [48] and CABS-Flex [49], a coarse-grained molecular dynamics simulator to estimate the proteins' most dominant structural fluctuations in the near-native ensemble. The integration of both approaches under the same pipeline allowed A3D to model and estimate mutation impact on stability and aggregation propensity. Using this strategy, it was possible to explain the mechanism of human  $\beta$ 2-microglobulin aggregation, which entails a severe complication for long-term hemodialysis patients. Aggregation-prone mutants in this protein tend to expose STAPs, which are protected in non-aggregating variants.

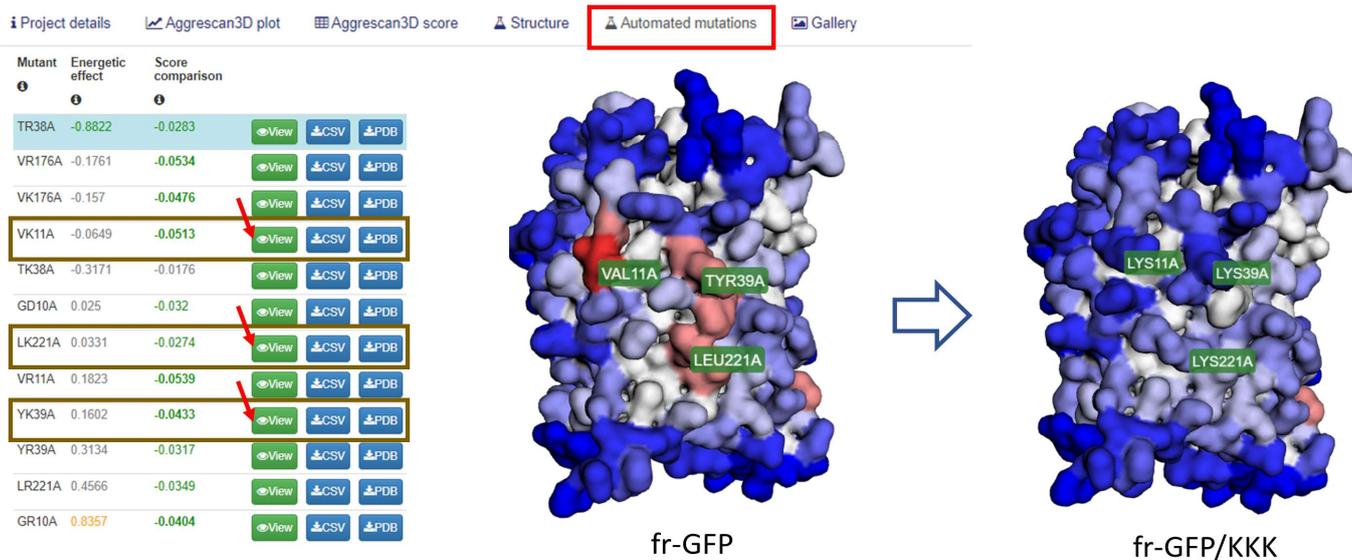
Due to its high computational costs, CABS-flex simulations were first restricted to small proteins. In 2019, the initial A3D release was subsequently updated to the 2.0 version [50,51], which allowed studies on large biomolecules such as antibodies, protein fibers, or multi-chain protein complexes [50]. In addition, the A3D 2.0 included other significant improvements, such as the automatic engineering of more soluble yet stable protein variants and a REST-ful service to incorporate the server into bioinformatic pipelines. This last algorithm version has been incorporated into a cost-effective routine tool specifically developed for designing and optimizing multimeric protein materials [52]. A standalone version of A3D 2.0 was recently released [53], which avoids erratic internet connections or deal with privacy concerns.

Notably, in July 2019, the Spanish Biophysical Society (SBE) designated the paper describing the method of A3D 2.0 as highlighted paper. Since its launch, A3D has aided the community in multiple experimental efforts, such as redesigning proteins for biotechnological approaches and engineering protein-based nanostructures [52,54–56]. For instance, A3D has allowed the *in silico* redesigning of one of the more soluble GFP variants [54]. The A3D-assisted redesign of this protein is shown in Figure 2. Other A3D applications included the study of the impact on the aggregation of pathogenic [37,57,58] and non-pathogenic protein variants [54,59–61], the analysis of the binding of antibacterial proteins to membranes [62], the understanding of chaperone client recognition [63], and the assistance with neglected tropical disease vaccine development [64] or model viral protein evolution throughout the SARS- CoV-2 pandemic [65].

**A**



**B**



**Figure 2.** A3D–assisted redesign of the green fluorescent protein (GFP). (A) Example of an A3D plot generated for the redesign of GFP. The red arrows point to aggregation-prone residues identified in a globular context. (B) Automatic mutations mode tab depicting three energetically favorable solubilizing mutations (indicated with brown squares). Note that A3D predicts three structural aggregation-prone regions (colored in red) that are solubilized (colored in blue) when applying a triple mutation to Lys amino residues. This example was generated using a previously solved structure of GFP (PDB code, 2B3Q, chain a).

### 2.3. A3D Database: Structure-Based Predictions of Protein Aggregation for the Human Proteome

In 2021, a comprehensive database containing highly accurate structure predictions for the human proteome was published [66]. These predictions were computed using AlphaFold (AF), a deep-learning neural network model developed by Jumper and coworkers [67]. As discussed above, A3D exploits the structural information from atomic models to identify surface-exposed aggregation-prone patches. We have exploited A3D to compute the aggregation propensity of the entire human proteome in the AF database. These data have been compiled in the A3D Database, which includes the precalculated A3D predictions for 23,391 human proteins [34]. This database is the first compiling aggregation in protein structures at this large scale and is freely available at (<http://biocomp.chem.uw.edu.pl/A3D2/hproteome> (accessed on 12 January 2023)).

The first release of this database included interesting features from the more recent implementation of A3D, such as the capacity to predict the effect of selected mutations on protein stability and aggregation propensity, as well as propose optimal solubility-enhancing mutations for every compiled human protein. Each entry of the A3D database includes a detailed description of the structure-based aggregation propensity for the protein of interest. The A3D database also incorporates user-friendly graphical tools for protein structure visualization and interpretation. Examples of potential applications include studying the impact of genetic mutations and engineering the solubility of pharmaceutically relevant human proteins, including antibodies, replacement enzymes, and growth factors.

## 3. Computational Tools to Study Prion-like Proteins

Prions are a particular class of amyloids that can propagate their misfolded conformation. These proteins have unique compositional features that have been exploited to develop dedicated bioinformatics tools capable of identifying novel pathological and functional polypeptides with prion-like properties. Herein, we discuss the features of four different algorithms developed by our group to study prions and prion-like proteins.

### 3.1. PrionScan: An Online Database of Predicted Prion Domains in Complete Proteomes

In 2014, we developed PrionScan (<http://webapps.bifi.es/prionscan> (accessed on 12 January 2023)) as an open-source database of organized and up-to-date predictions for putative prion-forming proteins for all the publicly available proteomes from all taxonomic subdivisions [68–71]. The PrionScan algorithm has been developed based on the assumption that prion propensity is determined by the composition of protein sequences [71,72]. Previously developed algorithms primarily focused on identifying amyloidogenic regions in pathogenic proteins based on local structural and primary sequence characteristics. However, most of these programs were not suited to analyze prion behavior since, globally, prion domains do not share the sequential characteristics common to disease-associated  $\beta$ -sheet amyloids [73].

PrionScan was designed to identify and score prion regions based on the compositional bias of prionogenic regions as deduced from an extensive set of experimentally validated prion and non-prion sequences from yeast. These data were exploited to build and train a probabilistic model that uses the statistical significance of individual amino acid propensities to detect Q/N-rich prion-like regions in all UniProtKB annotated proteomes [68,71,72,74]. In addition to storing information on putative prion proteins, PrionScan provides a function to predict prion regions in sequences not reported in public databases [68,71,75]. The data generated for a prediction comprises the sequence and localization of the highest-scoring putative prion domain and additional information about the protein, such as the Gene Ontology (GO) Terms and cross-references to other databases [68].

PrionScan has been used to understand prion/prionogenic proteins' functions and how their interaction networks have a substantial impact on gene regulation [76] or to identify regions driving liquid–liquid phase separation (LLPS) [77]. Recently, we have applied PrionScan to identify and characterize novel prion-like proteins in more than 800 bacteria proteomes, suggesting that prion-like presence is a common feature of different prokaryotic genomes [70].

### 3.2. *pWaltz and PrionW: Identification of Prion-like Protein Domains*

In 2015, we launched the pWaltz algorithm (<http://bioinf.uab.es/pWALTZ/> (accessed on 12 January 2023)) [78]. This predictor was inspired by the Waltz amyloid prediction strategy [79], but employed a lower detection threshold to identify milder amyloids and used a larger sliding window that fitted the size of the minimum transmissible  $\beta$ -fold described at that time [80,81]. As described, prion-like conversion was initially thought to be driven by compositional features alone [82,83]. However, in 2010 a seminal study by Toombs and co-workers using the yeast prion Sup35 suggested that certain stretches of the prion domain may play a driving role in its transition [84]. Surprisingly, their results indicated that these regions do not exhibit bias for residues overrepresented in yeast prions, such as asparagines (N) and glutamines (Q). On the contrary, hydrophobic residues were favored, while charged residues and prolines (P) harmed prion formation [84]. These biases were reminiscent of those used by pure amyloid predictors [79], which sparked the idea that prion conversion or propagation could rely on particular amyloid-like contributions. We further realized that previously reported prion or prion-like domains (PrLDs) had short stretches of mild amyloid propensity, and their mutation could explain observed differences in prion-like conversion. Based on these observations, we developed the pWaltz algorithm that could discriminate Q/N-rich domains with and without prion activity with higher accuracy than the compositional-only prediction methods available at the time [78].

Since its release, pWaltz has been applied, coupled to different PrLD boundary prediction algorithms, to detect soft amyloid cores in yeast and human prion-like proteins [85,86], to identify the first bacterial prion [87] and prion candidates in the malaria parasite [88], to evaluate mutation impact on prion-like protein aggregation [89], to understand the aggregation of human prion-like proteins [90], or to describe the mechanism of Med15 and TBP aggregation from initial coiled-coil conformations [60,91].

In 2015, we implemented PrionW (<http://bioinf.uab.cat/prionw/> (accessed on 12 January 2023)), a prion prediction algorithm that works with complete protein sequences, as it identifies the compositional context and the structural features needed for prion conversion [92]. PrionW first runs a disorder prediction over the input sequence, and those stretches deemed disordered are evaluated for a minimum Q/N enrichment. Then, the best candidate sequence is evaluated with the pWaltz algorithm, and the selected PrLD and soft amyloid core is presented. We employed PrionW to analyze the complete yeast proteome demonstrating that it recalls bona fide prion proteins with high accuracy. Over the past years, PrionW has helped scientists study telomeric-associated proteins' evolution in *Candida albicans* strains [93], to select yeast prion-like transcription factors that co-aggregate with Swi1 in prion state, explaining another layer of how the prion phenotype changes gene expression patterns [94]. PrionW has also been used to investigate the role of pathogenic SFPQ human protein in Alzheimer's and Creutzfeldt Jakob diseases [95], to understand the evolution of prions in fungal species [96], study the evolution of mammalian meiotic proteins [97], or proposed as a predictor of prion-like proteins capable of LLPS [77,98].

### 3.3. AMYCO: A Server for Prediction of the Impact of Mutations on the Aggregation Propensity of Prion-like Proteins

In 2017, the first extensive mutational study addressing the aggregation of a human prion-like protein *in vivo* was reported [99]. It studied the ribonucleoprotein hnRNPA2, whose aggregation is associated with the development of Amyotrophic Lateral Sclerosis (ALS) and multisystem proteinopathy [100]. This pioneering work provided a robust experimental framework to evaluate the determinants driving pathogenic PrLDs' aggregation. We used it to demonstrate that an equation that simultaneously considers the effects of mutations on PrLDs' composition and localized amyloid propensity best predicted the impact of amino acid substitutions on the intracellular aggregation of functional yeast prions and human disease-linked proteins [100–102]. The derived amino acid scoring system was implemented in 2019 into the publicly available AMYCO (combined AMYloid and COmposition-based prediction of prion-like propensity) algorithm [89].

AMYCO (<http://bioinf.uab.es/amycov04/> (accessed on 12 January 2023)) is a web server that allows the fast, automated, and graphical evaluation of the effect of mutations on the aggregation properties of prion-like proteins [89]. At that time, its performance was better than previous state-of-the-art predictors. Since its publication, AMYCO implementation has been used to gain insights into prion evolution, especially the appearance and conservation of prion-protective or -enhancing mutations in different mammals [103–106] and birds [107]. It has also been used to identify prion disease-related somatic mutation in the prion gene from cancer patients [108] or to rationalize the effect of point mutations in the hnRNPD gene on the onset of a rare type of muscular atrophy [109].

### 3.4. SGnn: A Server for the Prediction of Prion-like Domains Recruitment to Stress Granules upon Heat Stress

Stress granules (SGs) are dynamic and reversible biological condensates that form in response to different cellular stresses [110]. These intracellular structures are constituted mainly by mRNAs and proteins containing PrLDs similar to those found in yeast [110].

Stress Granules neural network (SGnn) (<http://sgnn.ppmclab.com/> (accessed on 12 January 2023)) is a web application developed by our group that predicts PrLDs' propensity to populate heat-induced SGs upon heat stress in complete proteomes [111]. To perform the predictions, the SGnn algorithm evaluates three relevant parameters that have been identified as important for SG localization: (i) PrLD aggregation propensity using the Aggrescan algorithm [31] and CamSol Intrinsic [112], (ii) the ability to establish electrostatic interactions (i.e., net charge per residue), and (iii) the free cysteine content. All these sequence- and composition-dependent features contribute to PrLDs' heat-induced assembly and can be read in their sequences. The predictive method implemented in SGnn to assess PrLDs' behavior is based on these assumptions and on Ross and coworkers' *in vivo* characterization of yeast PrLDs' recruitment to heat-induced SGs [110]. Exploiting this experimental data, a feed-forward neural network (FFNN) was trained, and its discriminatory potential was benchmarked against positive and negative PrLD sequences [111], this made it possible to define the three features providing the best prediction of the propensity of a PrLD to be recruited into SG upon heat stress, with accuracy and precision higher than contemporary algorithms using only compositional parameters, suggesting that specific interactions between defined residues play a role in the recruitment of proteins to these condensates. SGnn provides tabular results for all calculated parameters and a final decision (true/false) on the recruitment of PrLDs to SGs.

As a representative example, SGnn has been recently used by Harrison and coworkers to predict whether ortholog sequences from metazoans and plants can be recruited into SGs [113].

#### 4. Computational Tools to Study Intrinsically Disordered Proteins (IDPs)

Intrinsically Disordered Proteins (IDPs) have primary structures that combine low mean hydrophobicity and high net charge. The absence of a driving force for compaction and electrostatic repulsions causes the proteins populating this sequence space to present extended conformations in which amino acids are highly exposed to the solvent. Thus, solution conditions, including the pH, significantly impact the structure adopted by disordered protein regions. Here, we introduce a set of bioinformatics tools we developed to provide a framework to study IDPs' properties in a context-dependent manner.

##### 4.1. DispHred and DispHScan: Predicting Protein Disorder as a Function of pH

In 2020, we released DispHred (<https://ppmclab.pythonanywhere.com/DispHred> (accessed on 12 January 2023)) [114]. This tool was specifically developed to study the effect of pH on the order–disorder transitions of proteins possessing low secondary structure content. This server uses Henderson Hasselbalch's equation to calculate the protein's net charge and the pH-dependent hydrophobicity scale developed by Zamora et al. [115]. First, we validated the utility of this novel pH-dependent hydrophobicity scale, building up a dataset of experimentally validated disordered and single-chain folded proteins. Their associated net charge and hydrophobicity scores were computed and represented in charge-hydrophobicity plots, which were then used to assess the disorder-predicting potential of this representation. Receiver Operating Characteristic (ROC) analysis was performed on these plots, indicating high performance compared to the traditional Guy's and Kyle-Dolittle's hydrophobicity scales [116]. Afterward, the model was tested to predict pH-dependent order-to-disorder transitions. To do so, we used seven disordered proteins and peptides for which their pH-dependent conformations were validated experimentally. Using Support Vector Machines (SVMs), a linear boundary condition was defined. This classification system correctly discriminated folded and disordered proteins, avoiding overfitting and providing a margin of uncertainty near the boundary condition line.

DispHred is ready to use under its freely available web server implementation and allows the users access to the individual sequence order-to-disorder transition analysis. Among the variables of the analysis, the user can choose the sliding window size, the starting and ending pH interval, and the pH step used. The results page presents tabular and graphical data of the DispH score for the protein at every given pH. A score over 0 indicates that the protein is folded, while negative scores indicate that the protein is unfolded at the given pH.

DispHred has been used to predict pH-dependent order transitions in amphiphilic peptides to study their self-assembly [117] and disorder-to-order transitions in redox or alkali environments in viral IDPs [118]. It has also been used in biomedicine to study the ordered state of possible bioactive peptides regarding pH [119] and the effect of pH on the binding of drugs to the Human Serum Albumin's disordered regions [120].

DispHred represented the first bioinformatics tool specifically designed to predict protein disorder as a function of pH [114]. However, despite the novelty of the method, the DispHred algorithm was limited by the prediction of a single sequence at a time, thus precluding the analysis of pH-dependent disorder in large datasets such as proteomes. Moreover, specific information was not provided, such as identifying the pH of transition or the nature of identified conformational switches. For these reasons, in 2020, we also released DispHScan (<http://disphscan.ppmclab.com> (accessed on 12 January 2023)) [121]. The DispHScan pipeline looks for possible disorder transitions in a defined pH interval and determines the nature of such conformational changes (e.g., conditional folding, conditional unfolding, or multitransition). At the same time, if a transition occurs, the corresponding pH value(s) are specified. As discussed above, the most relevant novelty of DispHScan relies on its ability to run pH-dependent disorder predictions for multiple sequences. In this sense, the server performance was tested by running the proteomes of four model organisms, including human, *Saccharomyces cerevisiae*, *Escherichia coli*, and *Caenorhabditis elegans*, each at >25 different pHs. Beyond proteome analyses, the server has been used to predict pH-dependent disorder in low complexity sequences involved in liquid–liquid phase separation (LLPS), where a significant correlation between protein disorder and solubility was observed at neutral pH [122].

#### 4.2. SolupHred: A Server to Predict the pH-Dependent Aggregation of Intrinsically Disordered Proteins

Biophysicists have been long interested in predicting aggregation from protein sequences in defined conditions [79,122,123]. However, the protein microenvironment is highly dynamic, and the aggregation of polypeptides is influenced by external factors such as pH [124,125]. This influence is especially relevant for IDPs, whose lack of defined three-dimensional conformation makes them more susceptible to environmental fluctuations [126].

SolupHred (<https://ppmclab.pythonanywhere.com/SolupHred> (accessed on 12 January 2023)) represented the first aggregation predictor for IDPs to incorporate the effect of pH in its core [127]. In order to develop the predictive model, we engineered three different variants of the measles virus phosphoprotein (PNT) displaying different net charges and isoelectric points (pI). Interestingly, we discovered that not only the net charge but also the lipophilicity depended on the solution pH [128]. The SolupHred algorithm implements this evidence into an empirical equation based on the assumption that pH-dependent aggregation in IDPs is determined by both charge and lipophilicity. SolupHred successfully recapitulated the aggregation propensities of disease-linked proteins such as alpha-synuclein [129], islet amyloid polypeptide [130], abeta 40 [131], or tau [132] at different pH levels.

The SolupHred web server works on top of an individual or multiple sequences and predicts solubility either in a pH interval or at a specific pH. After submission, it provides a solubility profile in the selected pH range, indicating the 10% maximum and 10% minimum solubilities (Figure 3). SolupHred can be used as a fast, cost-effective method to optimize experimental conditions, purification, and storage of IDPs, as well as for conducting large-scale analyses of pH-dependent IDP aggregation. The server has been used to study the correlation between solubility and LLPS in low-complexity regions of proteins implied in neuronal diseases [122].

A



## SolupHred

The pH-dependent solubility predictor of IDPs



SolupHred is the first phenomenological predictor that considers protein environment pH when calculating aggregation propensity of Intrinsically Disordered Proteins (IDPs).

Submission	Help	References	Contact
------------	------	------------	---------

Introduce your sequence(s) in [FASTA format](#).  
All sequence(s) will be treated as disordered. Alternatively users can explore pH dependent disorder [here](#).

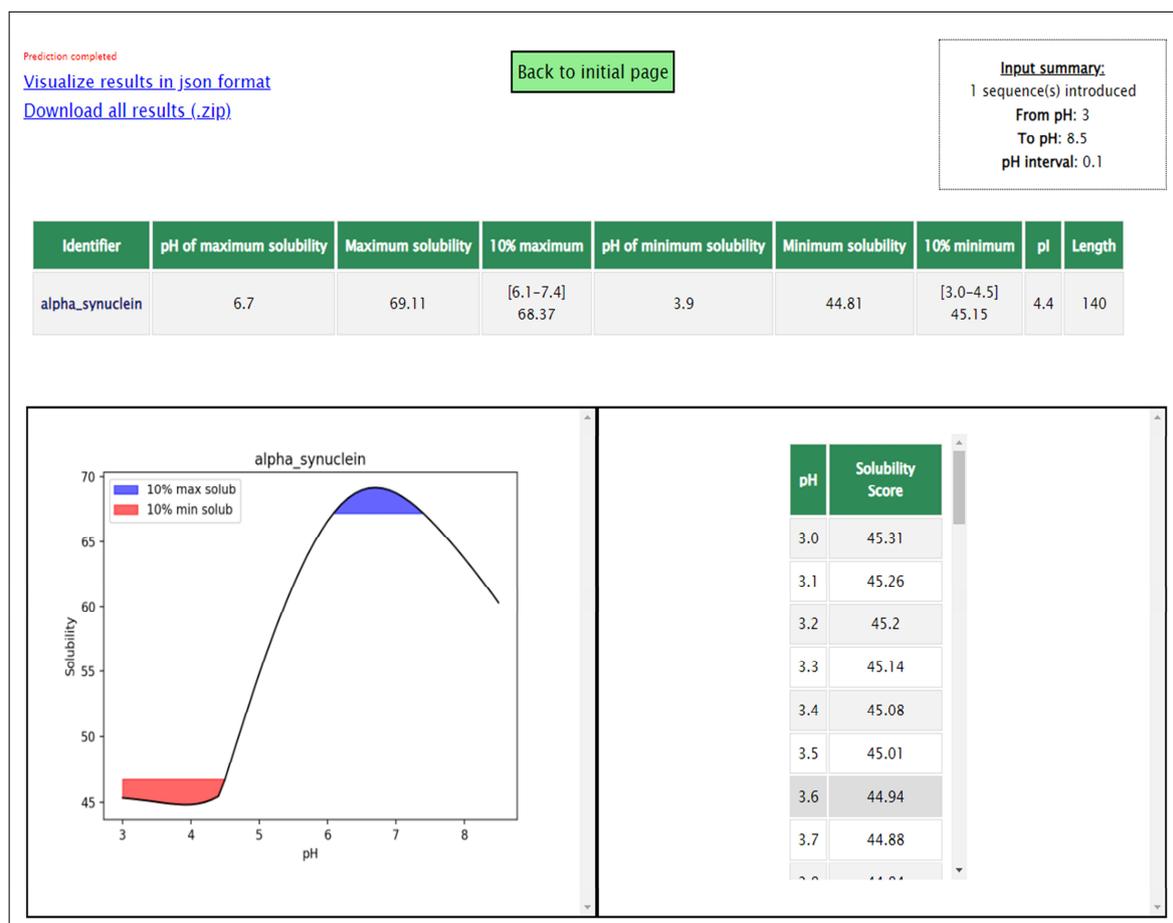
Choose File No file chosen [Example](#)

```
>alpha_synuclein
MDVFMKGLSKAKEGVVAEAAEKTQGVAAEAGKTEGVLVVGSKTEGVVHGVTVAEKTKEQVTNVGGAVVTGVTAVAQKTE
GAGSTAAATGFVKDOLGKNEEGAPQEGILEDMPVDPNEAYEMPSEEGYDYEPEA
```

From pH:  To pH:  pH interval:

Predict solubility at specific pH:

B



**Figure 3.** The SolupHred web server. (A) SolupHred's input interface requires the entry of one or many disordered regions in FASTA format and the selected pH range for the study. Alternatively, users can also predict solubility at a specific pH. (B) The results provide a table with the most relevant predictions such as 10% maximum and 10% minimum solubilities. Moreover, a graphical representation with the solubility profile in the pH range is also shown with the specific solubility scores. Different links can be downloaded for obtaining the results in the desired file format.

#### 4.3. CARs-DB: A Database of Cryptic Amyloidogenic Regions in Intrinsically Disordered Proteins

IDPs, lacking a defined secondary structure, were considered devoid of pro-aggregational regions. Classical amyloidogenic sequences are rich in nonpolar and aromatic amino acids, a compositional bias toward hydrophobicity not found in unstructured proteins. In 2021, we surveyed IDPs in search of non-canonical aggregation-prone segments [133]. This investigation led to the concept of Cryptically Amyloidogenic regions (CARs) of polar nature in Intrinsically Disordered Regions (IDRs). CARs play an essential role in mediating Protein–Protein Interactions (PPIs), but they are also connected to pathogenic processes, such as when non-native interactions occur. To provide a resource for researchers to assess the presence of amyloidogenic stretches in IDPs, CARs-DB was developed [134]. This database contains candidate CARs of all IDPs in the manually curated database, DisProt. To detect these segments, the Waltz algorithm, developed by Maurer-Stroh et al. [79], was employed, using a detection threshold lower than the one used to find conventional amyloid sequences. The lowest threshold of 73.5 was consistent with the one used to search for the amyloid cores of prion and prion-like proteins, for which experimental evidence of amyloid formation existed [85,86,135]. However, on this occasion, the identified sequence could be as short as seven residues since we have shown that highly polar peptides of this size form bona fide amyloids in vitro [136–138]. This evidence is consistent with the hypothesis that there exists an uncharted amyloid space away from hydrophobic sequences [139].

CARs-DB is a freely available database that users can access using the following link: <http://carsdb.ppmclab.com> (accessed on 12 January 2023). This precomputed database enables the detection of CARs in IDPs without the need to calculate their amyloidogenic propensity [134]. In the “Database” section, three thresholds are available: 85, 80, and 73.5. The lower the selected threshold, the more unconventional these amyloidogenic regions will be, presenting an increasingly polar nature. Among the information provided in the database, users can find the protein’s DisProt ID, UniProt Accession ID, its name, and the source organism. Both IDs provide links to each database (DisProt and UniProt, respectively). In addition, the start and end of the IDR (annotated in DisProt) are also specified. Finally, information regarding the CAR (start and end positions, length, sequence, and Waltz score) is included.

This recently published resource has been used to detect CARs in the Hendra and Nipah P proteins’ intrinsically disordered N-terminal domain (NTD) [140]. The P protein is a phosphoprotein that belongs to the viral RNA-dependent-RNA-polymerase complex (RpRd), which is necessary for the transcription and replication of these viruses. Several of the detected regions had already been experimentally validated [133], proving the potential of this database to analyze IDRs for the presence of polar amyloidogenic stretches.

## 5. Discussion

Due to its importance in biomedical research and the biotechnological sector, protein aggregation has changed over the past decades from a virtually unexplored study subject to a scorching research issue. We have seen ground-breaking scientific advancements, and now, we have a profound mechanistic view of how aggregation occurs. Computational tools, such as the ones developed by our lab and described here, have contributed significantly to this knowledge, helping to direct experimental efforts to elucidate the molecular pathways behind disorders related to protein aggregation. Additionally, they have accelerated the development of engineered protein variants with enhanced solubility and stability, reducing the time and money needed to produce therapeutic proteins.

All our programs, and the large majority of those developed by our colleagues, are freely available to the public in the format of a web server and/or as an executable file. In addition, all of them contain help files with detailed information for users, including a general description of the tool and relevant usage information. In silico approximations, such as the ones we detail here, are gradually included in many wet laboratories’ routines as a cost-effective way to design experimental pipelines. In this way, the primary articles

describing the algorithms discussed in this review have collectively received >1350 citations as of 22nd December and according to Google Scholar.

The different algorithms capture distinct aspects of protein aggregation and are intended for diverse applications. In a way, the timeline shown in Figure 1 reflects how the interests of the field have evolved over time and how the integration of experimental biophysical data and predictions has allowed us, as a community, to address challenges of increasing complexity. Initially thought to be a purely stochastic and thus unpredictable phenomenon, the realization that, as folding, aggregation was somehow imprinted in the sequence [141] opened an avenue for rationalization of aggregation reactions at the proteome scale. With sequence-based predictors available, very soon, new types of protein sequences attracted the community's interest, those belonging to prion and prion-like proteins. It was immediately evident that the sequence space of archetypical amyloids and prion-like proteins only partially overlapped, and a new generation of algorithms was generated. That effort was worthwhile because these algorithms, or those directly derived from them, are currently being used to study the propensity of proteins to form part of the fashionable membraneless organelles [142]. The need for different scales when dealing with different sequence sets already indicated that the amyloid sequence space was far broader than previously believed. It is now clear that highly soluble sequences with minimal aliphatic content and/or high net charge can form amyloids [139]. The idea that low solubility and aggregation propensity are interchangeable qualities was at the heart of most initial algorithms; new programs and databases are revisiting this idea to fish sequences in this new amyloid terrain.

Once intrinsic sequential factors were clarified for the different aggregation flavors, extrinsic factors had to be considered. They include viscosity, temperature, pH, ionic concentration, protein concentration, solvent identity, and interactions with other molecules. The absence of rigorous experimental data spanning all potential variable combinations for a group of sequentially unrelated proteins has been the fundamental obstacle to developing systems that can incorporate the protein microenvironment in their predictions. However, as we illustrate here, the first attempts to incorporate parameters such as the solution pH in the prediction pipeline are rendering their fruits, especially for IDPs, whose properties are especially sensitive to the solution conditions.

In addition to the intrinsic sequence, one should consider other factors when studying the aggregation of globular proteins, including stability, conformation, cooperativity, surface solubility, and dynamics. Structure-based algorithms were born to deal with all these parameters automatically. However, for a long time, the application of these tools was limited to a relatively reduced space of the protein universe: those for which a high-resolution structure exists or a model could be confidently constructed. However, with the avenue of programs such as AlphaFold [67], this limitation has been broken down, and databases containing accurate protein aggregation predictions for the complete set of globular proteins in a given proteome are already available online [34].

The time has come for artificial intelligence (AI) to enter the aggregation prediction arena [143]. The application of this technology requires the availability of a large number of biophysical studies that can feed it. Unfortunately, the acquisition of biochemical (stability, pH-dependence of conformational changes) and biophysical data (type of condensation or aggregation) is seen as a low-value objective. We should remember that AI successes such as AlphaFold would not have been possible without an extraordinarily well-curated database of protein structures [144]. Building a consortium that can generate a coherent set of information related to protein aggregation is now more than ever a necessity and a must, given the growing impact of protein aggregation-related diseases in our society.

## 6. Conclusions

We have seen impressive innovations in protein aggregation prediction in the last fifteen years. We are pleased to have played our part in this progress, together with an outstanding group of international researchers.

**Author Contributions:** Conceptualization, J.G.-P. and S.V.; writing—original draft preparation, C.P.-G., O.B., A.B.-N., M.F.-S., V.I., J.G.-P. and S.V.; supervision, J.G.-P. and S.V.; funding acquisition, S.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by European Union Horizon 2020 research and innovation programme under GA 952334 (PhasAGE), by the Spanish Ministry of Science and Innovation (PID2019-105017RB-I00) to S.V. and by ICREA, ICREA-Academia 2015 and 2020 to S.V. C.P.-G. was supported by the Secretariat of Universities and Research of the Catalan Government and the European Social Fund (2021 FI\_B 00087). M.F.-S. was supported by Ministry of Science and Innovation via a doctoral grant (FPU20/02897). V.I. was supported by the Spanish Ministry of Universities and the European Union-NextGenerationEU (ruling 02/07/2021, Universitat Autònoma de Barcelona). J.G.-P. was funded by the Spanish Ministry of Science and Innovation with a postdoctoral grant Juan de la Cierva Incorporación (IJC2019-041039-I).

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** We acknowledge the members of the Ventura's lab in the last two decades for their contributions to the studies we detail in this review.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dill, K.A.; MacCallum, J.L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042–1046. [[CrossRef](#)] [[PubMed](#)]
2. Daggett, V.; Fersht, A.R. Protein folding and binding: Moving into uncharted territory. *Curr. Opin. Struct. Biol.* **2009**, *19*, 1–2. [[CrossRef](#)] [[PubMed](#)]
3. Arolas, J.L.; Aviles, F.X.; Chang, J.-Y.; Ventura, S. Folding of small disulfide-rich proteins: Clarifying the puzzle. *Trends Biochem. Sci.* **2006**, *31*, 292–301. [[CrossRef](#)] [[PubMed](#)]
4. Mishra, P.; Jha, S.K. The native state conformational heterogeneity in the energy landscape of protein folding. *Biophys. Chem.* **2022**, *283*, 106761. [[CrossRef](#)]
5. Pallares, I.; Vendrell, J.; Aviles, F.X.; Ventura, S. Amyloid Fibril Formation by a Partially Structured Intermediate State of  $\alpha$ -Chymotrypsin. *J. Mol. Biol.* **2004**, *342*, 321–331. [[CrossRef](#)]
6. Ventura, S.; Zurdo, J.; Narayanan, S.; Parreño, M.; Mangués, R.; Reif, B.; Chiti, F.; Giannoni, E.; Dobson, C.M.; Aviles, F.X.; et al. Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7258–7263. [[CrossRef](#)]
7. Jahn, T.R.; Radford, S.E. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **2008**, *469*, 100–117. [[CrossRef](#)]
8. Sabate, R.; Ventura, S. Cross-beta-sheet supersecondary structure in amyloid folds: Techniques for detection and characterization. *Methods Mol. Biol.* **2012**, *932*, 237–257. [[CrossRef](#)]
9. Riek, R.; Eisenberg, D.S. The activities of amyloids from a structural perspective. *Nature* **2016**, *539*, 227–235. [[CrossRef](#)]
10. Selkoe, D.J. Folding proteins in fatal ways. *Nature* **2003**, *426*, 900–904. [[CrossRef](#)]
11. Chiti, F.; Dobson, C.M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27–68. [[CrossRef](#)] [[PubMed](#)]
12. Invernizzi, G.; Papaleo, E.; Sabate, R.; Ventura, S. Protein aggregation: Mechanisms and functional consequences. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 1541–1554. [[CrossRef](#)] [[PubMed](#)]
13. Dobson, C.M.; Knowles, T.P.J.; Vendruscolo, M. The Amyloid Phenomenon and Its Significance in Biology and Medicine. *Cold Spring Harb. Perspect. Biol.* **2020**, *12*, a033878. [[CrossRef](#)] [[PubMed](#)]
14. Dobson, C.M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **2004**, *15*, 3–16. [[CrossRef](#)] [[PubMed](#)]
15. Fraga, H.; Graña-Montes, R.; Illa, R.; Covalada, G.; Ventura, S. Association between Foldability and Aggregation Propensity in Small Disulfide-Rich Proteins. *Antioxid. Redox Signal.* **2014**, *21*, 368–383. [[CrossRef](#)]

16. Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.* **2004**, *342*, 345–353. [[CrossRef](#)]
17. Rousseau, F.; Serrano, L.; Schymkowitz, J.W. How Evolutionary Pressure against Protein Aggregation Shaped Chaperone Specificity. *J. Mol. Biol.* **2006**, *355*, 1037–1047. [[CrossRef](#)]
18. Tartaglia, G.G.; Pechmann, S.; Dobson, C.M.; Vendruscolo, M. Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **2007**, *32*, 204–206. [[CrossRef](#)]
19. Ventura, S.; Villaverde, A. Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **2006**, *24*, 179–185. [[CrossRef](#)]
20. Kim, Y.E.; Hipp, M.S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **2013**, *82*, 323–355. [[CrossRef](#)]
21. Chiti, F.; Calamai, M.; Taddei, N.; Stefani, M.; Ramponi, G.; Dobson, C.M. Studies of the aggregation of mutant proteins *in vitro* provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (Suppl. 4), 16419–16426. [[CrossRef](#)] [[PubMed](#)]
22. De Baets, G.; Van Doorn, L.; Rousseau, F.; Schymkowitz, J. Increased Aggregation Is More Frequently Associated to Human Disease-Associated Mutations Than to Neutral Polymorphisms. *PLoS Comput. Biol.* **2015**, *11*, e1004374. [[CrossRef](#)] [[PubMed](#)]
23. Marinelli, P.; Navarro, S.; Graña-Montes, R.; Bañó-Polo, M.; Fernández, M.R.; Papaleo, E.; Ventura, S. A single cysteine post-translational oxidation suffices to compromise globular proteins kinetic stability and promote amyloid formation. *Redox Biol.* **2018**, *14*, 566–575. [[CrossRef](#)] [[PubMed](#)]
24. Hamdan, N.; Kritsiligkou, P.; Grant, C.M. ER stress causes widespread protein aggregation and prion formation. *J. Cell Biol.* **2017**, *216*, 2295–2304. [[CrossRef](#)] [[PubMed](#)]
25. Cromwell, M.E.M.; Hilario, E.; Jacobson, F. Protein aggregation and bioprocessing. *AAPS J.* **2006**, *8*, E572–E579. [[CrossRef](#)]
26. Roberts, C.J. Protein aggregation and its impact on product quality. *Curr. Opin. Biotechnol.* **2014**, *30*, 211–217. [[CrossRef](#)]
27. Hamrang, Z.; Rattray, N.J.; Pluen, A. Proteins behaving badly: Emerging technologies in profiling biopharmaceutical aggregation. *Trends Biotechnol.* **2013**, *31*, 448–458. [[CrossRef](#)]
28. Castillo, V.; Graña-Montes, R.; Ventura, S. The aggregation properties of Escherichia coli proteins associated with their cellular abundance. *Biotechnol. J.* **2011**, *6*, 752–760. [[CrossRef](#)]
29. Ciryam, P.; Tartaglia, G.G.; Morimoto, R.I.; Dobson, C.M.; Vendruscolo, M. Widespread Aggregation and Neurodegenerative Diseases Are Associated with Supersaturated Proteins. *Cell Rep.* **2013**, *5*, 781–790. [[CrossRef](#)]
30. Tartaglia, G.G.; Pechmann, S.; Dobson, C.M.; Vendruscolo, M. A Relationship between mRNA Expression Levels and Protein Solubility in *E. coli*. *J. Mol. Biol.* **2009**, *388*, 381–389. [[CrossRef](#)]
31. Conchillo-Solé, O.; de Groot, N.S.; Avilés, F.X.; Vendrell, J.; Daura, X.; Ventura, S. AGGRESCAN: A server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinform.* **2007**, *8*, 65. [[CrossRef](#)]
32. de Groot, N.S.; Avilés, F.X.; Vendrell, J.; Ventura, S. Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer’s peptide. Side-chain properties correlate with aggregation propensities. *FEBS J.* **2006**, *273*, 658–668. [[CrossRef](#)] [[PubMed](#)]
33. de Groot, N.S.; Pallarés, I.; Avilés, F.X.; Vendrell, J.; Ventura, S. Prediction of “hot spots” of aggregation in disease-linked polypeptides. *BMC Struct. Biol.* **2005**, *5*, 18. [[CrossRef](#)] [[PubMed](#)]
34. Badaczewska-Dawid, A.E.; Garcia-Pardo, J.; Kuriata, A.; Pujols, J.; Ventura, S.; Kmiecik, S. A3D database: Structure-based predictions of protein aggregation for the human proteome. *Bioinformatics* **2022**, *38*, 3121–3123. [[CrossRef](#)] [[PubMed](#)]
35. de Groot, N.S.; Castillo, V.; Graña-Montes, R.; Ventura, S. AGGRESCAN: Method, Application, and Perspectives for Drug Design. *Methods Mol. Biol.* **2012**, *819*, 199–220. [[CrossRef](#)]
36. Tsolis, A.C.; Papandreou, N.C.; Iconomidou, V.A.; Hamodrakas, S.J. A Consensus Method for the Prediction of ‘Aggregation-Prone’ Peptides in Globular Proteins. *PLoS ONE* **2013**, *8*, e54175. [[CrossRef](#)]
37. Berdyński, M.; Miszta, P.; Safranow, K.; Andersen, P.M.; Morita, M.; Filipek, S.; Żekanowski, C.; Kuźma-Kozakiewicz, M. SOD1 mutations associated with amyotrophic lateral sclerosis analysis of variant severity. *Sci. Rep.* **2022**, *12*, 103. [[CrossRef](#)]
38. Martínez-Rubio, D.; Rodríguez-Prieto, A.; Sancho, P.; Navarro-Gonzalez, C.; Gorria-Redondo, N.; Miquel-Leal, J.; Marco-Marin, C.; Jenkins, A.; Soriano-Navarro, M.; Hernandez, A.; et al. Protein misfolding and clearance in the pathogenesis of a new infantile onset ataxia caused by mutations in PRDX3. *Hum. Mol. Genet.* **2022**, *31*, 3897–3913. [[CrossRef](#)]
39. Tavassoly, O.; Safavi, F.; Tavassoly, I. Seeding Brain Protein Aggregation by SARS-CoV-2 as a Possible Long-Term Complication of COVID-19 Infection. *ACS Chem. Neurosci.* **2020**, *11*, 3704–3706. [[CrossRef](#)]
40. De Groot, N.S.; Ventura, S. Protein Aggregation Profile of the Bacterial Cytosol. *PLoS ONE* **2010**, *5*, e9383. [[CrossRef](#)]
41. Graña-Montes, R.; de Oliveira, R.S.; Ventura, S. Protein aggregation profile of the human kinome. *Front. Physiol.* **2012**, *3*, 438. [[CrossRef](#)] [[PubMed](#)]
42. Castillo, V.; Chiti, F.; Ventura, S. The N-terminal Helix Controls the Transition between the Soluble and Amyloid States of an FF Domain. *PLoS ONE* **2013**, *8*, e58297. [[CrossRef](#)] [[PubMed](#)]
43. Santos, J.; Iglesias, V.; Ventura, S. Computational prediction and redesign of aberrant protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **2020**, *169*, 43–83. [[CrossRef](#)] [[PubMed](#)]

44. Castillo, V.; Ventura, S. Amyloidogenic Regions and Interaction Surfaces Overlap in Globular Proteins Related to Conformational Diseases. *PLoS Comput. Biol.* **2009**, *5*, e1000476. [[CrossRef](#)]
45. Castillo, V.; Espargaró, A.; Gordo, V.; Vendrell, J.; Ventura, S. Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria. *Proteomics* **2010**, *10*, 4172–4185. [[CrossRef](#)]
46. Graña-Montes, R.; de Groot, N.S.; Castillo, V.; Sancho, J.; Velazquez-Campoy, A.; Ventura, S.; Fraga, H.; Illa, R.; Covaleda, G. Contribution of Disulfide Bonds to Stability, Folding, and Amyloid Fibril Formation: The PI3-SH3 Domain Case. *Antioxid. Redox Signal.* **2012**, *16*, 1–15. [[CrossRef](#)]
47. 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 Res.* **2015**, *43*, W306–W313. [[CrossRef](#)]
48. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: An online force field. *Nucleic Acids Res.* **2005**, *33*, W382–W388. [[CrossRef](#)]
49. Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Res.* **2013**, *41*, W427–W431. [[CrossRef](#)]
50. Kuriata, A.; Iglesias, V.; Pujols, J.; Kurcinski, M.; Kmiecik, S.; Ventura, S. Aggrescan3D (A3D) 2.0: Prediction and engineering of protein solubility. *Nucleic Acids Res.* **2019**, *47*, W300–W307. [[CrossRef](#)]
51. Pujols, J.; Iglesias, V.; Santos, J.; Kuriata, A.; Kmiecik, S.; Ventura, S. A3D 2.0 Update for the Prediction and Optimization of Protein Solubility. *Methods Mol. Biol.* **2022**, *2406*, 65–84. [[CrossRef](#)]
52. Parladé, E.; Voltà-Durán, E.; Cano-Garrido, O.; Sánchez, J.M.; Unzueta, U.; López-Laguna, H.; Serna, N.; Cano, M.; Rodríguez-Mariscal, M.; Vazquez, E.; et al. An In Silico Methodology That Facilitates Decision Making in the Engineering of Nanoscale Protein Materials. *Int. J. Mol. Sci.* **2022**, *23*, 4958. [[CrossRef](#)]
53. Kuriata, A.; Iglesias, V.; Kurcinski, M.; Ventura, S.; Kmiecik, S. Aggrescan3D standalone package for structure-based prediction of protein aggregation properties. *Bioinformatics* **2019**, *35*, 3834–3835. [[CrossRef](#)]
54. Gil-Garcia, M.; Baño-Polo, M.; Varejao, N.; Jamroz, M.; Kuriata, A.; Caballero, M.D.; Lascorz, J.; Morel, B.; Navarro, S.; Reverter, D.; et al. Combining Structural Aggregation Propensity and Stability Predictions to Redesign Protein Solubility. *Mol. Pharm.* **2018**, *15*, 3846–3859. [[CrossRef](#)]
55. Ebo, J.S.; Saunders, J.C.; Devine, P.W.A.; Gordon, A.M.; Warwick, A.S.; Schiffrin, B.; Chin, S.E.; England, E.; Button, J.D.; Lloyd, C.; et al. An in vivo platform to select and evolve aggregation-resistant proteins. *Nat. Commun.* **2020**, *11*, 1816. [[CrossRef](#)]
56. Xia, X.; Kumru, O.S.; Blaber, S.L.; Middaugh, C.R.; Li, L.; Ornitz, D.M.; Sutherland, M.A.; Tenorio, C.A.; Blaber, M. Engineering a Cysteine-Free Form of Human Fibroblast Growth Factor-1 for “Second Generation” Therapeutic Application. *J. Pharm. Sci.* **2016**, *105*, 1444–1453. [[CrossRef](#)]
57. Bhandare, V.V.; Ramaswamy, A. The proteinopathy of D169G and K263E mutants at the RNA Recognition Motif (RRM) domain of tar DNA-binding protein (tdp43) causing neurological disorders: A computational study. *J. Biomol. Struct. Dyn.* **2018**, *36*, 1075–1093. [[CrossRef](#)]
58. Žerovnik, E. Putative alternative functions of human stefin B (cystatin B): Binding to amyloid-beta, membranes, and copper. *J. Mol. Recognit.* **2016**, *30*, e2562. [[CrossRef](#)]
59. Katina, N.S.; Balobanov, V.A.; Ilyina, N.B.; Vasiliev, V.D.; Marchenkov, V.V.; Glukhov, A.S.; Nikulin, A.D.; Bychkova, V.E. sw ApoMb Amyloid Aggregation under Nondenaturing Conditions: The Role of Native Structure Stability. *Biophys. J.* **2017**, *113*, 991–1001. [[CrossRef](#)]
60. Behbahanipour, M.; García-Pardo, J.; Ventura, S. Decoding the role of coiled-coil motifs in human prion-like proteins. *Prion* **2021**, *15*, 143–154. [[CrossRef](#)]
61. Gil-Garcia, M.; Navarro, S.; Ventura, S. Coiled-coil inspired functional inclusion bodies. *Microb. Cell Factories* **2020**, *19*, 117. [[CrossRef](#)]
62. Pulido, D.; Arranz-Trullén, J.; Prats-Ejarque, G.; Velázquez, D.; Torrent, M.; Moussaoui, M.; Boix, E. Insights into the Antimicrobial Mechanism of Action of Human RNase6: Structural Determinants for Bacterial Cell Agglutination and Membrane Permeation. *Int. J. Mol. Sci.* **2016**, *17*, 552. [[CrossRef](#)]
63. Pulido, P.; Llamas, E.; Llorente, B.; Ventura, S.; Wright, L.P.; Rodríguez-Concepción, M. Specific Hsp100 Chaperones Determine the Fate of the First Enzyme of the Plastidial Isoprenoid Pathway for Either Refolding or Degradation by the Stromal Clp Protease in Arabidopsis. *PLoS Genet.* **2016**, *12*, e1005824. [[CrossRef](#)]
64. Ishwarlall, T.Z.; Adeleke, V.T.; Maharaj, L.; Okpeku, M.; Adeniyi, A.A.; Adeleke, M.A. Identification of potential candidate vaccines against Mycobacterium ulcerans based on the major facilitator superfamily transporter protein. *Front. Immunol.* **2022**, *13*, 1023558. [[CrossRef](#)]
65. Flores-León, M.; Lázaro, D.F.; Shvachiy, L.; Krisko, A.; Outeiro, T.F. In silico analysis of the aggregation propensity of the SARS-CoV-2 proteome: Insight into possible cellular pathologies. *Biochim. Biophys. Acta Proteins Proteom.* **2021**, *1869*, 140693. [[CrossRef](#)]

66. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)]
67. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
68. Angarica, V.E.; Angulo, A.; Giner, A.; Losilla, G.; Ventura, S.; Sancho, J. PrionScan: An online database of predicted prion domains in complete proteomes. *BMC Genom.* **2014**, *15*, 102. [[CrossRef](#)]
69. Navarro, S.; Ventura, S. Computational methods to predict protein aggregation. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102343. [[CrossRef](#)]
70. Iglesias, V.; de Groot, N.S.; Ventura, S. Computational analysis of candidate prion-like proteins in bacteria and their role. *Front. Microbiol.* **2015**, *6*, 1123. [[CrossRef](#)]
71. Batlle, C.; Iglesias, V.; Navarro, S.; Ventura, S. Prion-like proteins and their computational identification in proteomes. *Expert Rev. Proteom.* **2017**, *14*, 335–350. [[CrossRef](#)]
72. Alberti, S.; Halfmann, R.; King, O.; Kapila, A.; Lindquist, S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* **2009**, *137*, 146–158. [[CrossRef](#)]
73. Pawar, A.P.; DuBay, K.F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C.M. Prediction of “Aggregation-prone” and “Aggregation-susceptible” Regions in Proteins Associated with Neurodegenerative Diseases. *J. Mol. Biol.* **2005**, *350*, 379–392. [[CrossRef](#)]
74. Angarica, V.E.; Ventura, S.; Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genom.* **2013**, *14*, 316. [[CrossRef](#)]
75. O’Carroll, A.; Coyle, J.; Gambin, Y. Prions and Prion-like assemblies in neurodegeneration and immunity: The emergence of universal mechanisms across health and disease. *Semin. Cell Dev. Biol.* **2020**, *99*, 115–130. [[CrossRef](#)]
76. Harbi, D.; Harrison, P.M. Interaction Networks of Prion, Prionogenic and Prion-Like Proteins in Budding Yeast, and Their Role in Gene Regulation. *PLoS ONE* **2014**, *9*, e100615. [[CrossRef](#)]
77. Pancsa, R.; Vranken, W.; Mészáros, B. Computational resources for identifying and describing proteins driving liquid–liquid phase separation. *Brief. Bioinform.* **2021**, *22*, bbaa408. [[CrossRef](#)]
78. Sabate, R.; Rousseau, F.; Schymkowitz, J.; Ventura, S. What Makes a Protein Sequence a Prion? *PLoS Comput. Biol.* **2015**, *11*, e1004013. [[CrossRef](#)]
79. Maurer-Stroh, S.; Debulpaep, M.; Kuemmerer, N.; Lopez De La Paz, M.; Martins, I.C.; Reumers, J.; Morris, K.L.; Copland, A.; Serpell, L.; Serrano, L.; et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods* **2010**, *7*, 237–242. [[CrossRef](#)]
80. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A.B.; Riek, R.; Meier, B.H. Amyloid Fibrils of the HET-s(218–289) Prion Form a  $\beta$  Solenoid with a Triangular Hydrophobic Core. *Science* **2008**, *319*, 1523–1526. [[CrossRef](#)]
81. Wasmer, C.; Zimmer, A.; Sabaté, R.; Soragni, A.; Saupe, S.J.; Ritter, C.; Meier, B.H. Structural Similarity between the Prion Domain of HET-s and a Homologue Can Explain Amyloid Cross-Seeding in Spite of Limited Sequence Identity. *J. Mol. Biol.* **2010**, *402*, 311–325. [[CrossRef](#)]
82. Ross, E.D.; Edskes, H.K.; Terry, M.J.; Wickner, R.B. Primary sequence independence for prion formation. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12825–12830. [[CrossRef](#)]
83. Toombs, J.A.; Petri, M.; Paul, K.R.; Kan, G.Y.; Ben-Hur, A.; Ross, E.D. De novo design of synthetic prion domains. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6519–6524. [[CrossRef](#)]
84. Toombs, J.A.; McCarty, B.R.; Ross, E.D. Compositional Determinants of Prion Formation in Yeast. *Mol. Cell. Biol.* **2010**, *30*, 319–332. [[CrossRef](#)]
85. Batlle, C.; de Groot, N.S.; Iglesias, V.; Navarro, S.; Ventura, S. Characterization of Soft Amyloid Cores in Human Prion-Like Proteins. *Sci. Rep.* **2017**, *7*, 12134. [[CrossRef](#)]
86. Sant’Anna, R.; Fernández, M.R.; Batlle, C.; Navarro, S.; de Groot, N.S.; Serpell, L.; Ventura, S. Characterization of Amyloid Cores in Prion Domains. *Sci. Rep.* **2016**, *6*, srep34274. [[CrossRef](#)]
87. Pallarès, I.; Iglesias, V.; Ventura, S. The Rho Termination Factor of *Clostridium botulinum* Contains a Prion-Like Domain with a Highly Amyloidogenic Core. *Front. Microbiol.* **2016**, *6*, 1516. [[CrossRef](#)]
88. Pallarès, I.; de Groot, N.S.; Iglesias, V.; Sant’Anna, R.; Biosca, A.; Fernández-Busquets, X.; Ventura, S. Discovering Putative Prion-Like Proteins in *Plasmodium falciparum*: A Computational and Experimental Analysis. *Front. Microbiol.* **2018**, *9*, 1737. [[CrossRef](#)]
89. Iglesias, V.; Conchillo-Sole, O.; Batlle, C.; Ventura, S. AMYCO: Evaluation of mutational impact on prion-like proteins aggregation propensity. *BMC Bioinform.* **2019**, *20*, 24. [[CrossRef](#)]
90. Navarro, S.; Marinelli, P.; Diaz-Caballero, M.; Ventura, S. The prion-like RNA-processing protein HNRPDL forms inherently toxic amyloid-like inclusion bodies in bacteria. *Microb. Cell Factories* **2015**, *14*, 102. [[CrossRef](#)]

91. Batlle, C.; Calvo, I.; Iglesias, V.; Lynch, C.J.; Gil-Garcia, M.; Serrano, M.; Ventura, S. MED15 prion-like domain forms a coiled-coil responsible for its amyloid conversion and propagation. *Commun. Biol.* **2021**, *4*, 414. [[CrossRef](#)]
92. Zambrano, R.; Conchillo-Sole, O.; Iglesias, V.; Illa, R.; Rousseau, F.; Schymkowitz, J.; Sabate, R.; Daura, X.; Ventura, S. PrionW: A server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. *Nucleic Acids Res.* **2015**, *43*, W331–W337. [[CrossRef](#)]
93. Dunn, M.J.; Shazib, S.U.A.; Simonton, E.; Slot, J.C.; Anderson, M.Z. Architectural groups of a subtelomeric gene family evolve along distinct paths in *Candida albicans*. *G3 (Bethesda)* **2022**, *12*, jkac283. [[CrossRef](#)]
94. Du, Z.; Regan, J.; Bartom, E.; Wu, W.-S.; Zhang, L.; Goncharoff, D.K.; Li, L. Elucidating the regulatory mechanism of Swi1 prion in global transcription and stress responses. *Sci. Rep.* **2020**, *10*, 21838. [[CrossRef](#)]
95. Younas, N.; Zafar, S.; Shafiq, M.; Noor, A.; Siegert, A.; Arora, A.S.; Galkin, A.; Zafar, A.; Schmitz, M.; Stadelmann, C.; et al. SFPQ and Tau: Critical factors contributing to rapid progression of Alzheimer’s disease. *Acta Neuropathol.* **2020**, *140*, 317–339. [[CrossRef](#)]
96. An, L.; Fitzpatrick, D.; Harrison, P.M. Emergence and evolution of yeast prion and prion-like proteins. *BMC Evol. Biol.* **2016**, *16*, 24. [[CrossRef](#)]
97. Papanikos, F.; Daniel, K.; Goercharn-Ramlal, A.; Fei, J.-F.; Kurth, T.; Wojtasz, L.; Dereli, I.; Fu, J.; Penninger, J.; Habermann, B.; et al. The enigmatic meiotic dense body and its newly discovered component, SCML1, are dispensable for fertility and gametogenesis in mice. *Chromosoma* **2017**, *126*, 399–415. [[CrossRef](#)]
98. Maziuk, B.; Ballance, H.I.; Wolozin, B. Dysregulation of RNA Binding Protein Aggregation in Neurodegenerative Disorders. *Front. Mol. Neurosci.* **2017**, *10*, 89. [[CrossRef](#)]
99. Paul, K.R.; Molliex, A.; Cascarina, S.; Boncella, A.E.; Taylor, J.P.; Ross, E.D. Effects of Mutations on the Aggregation Propensity of the Human Prion-Like Protein hnRNPA2B1. *Mol. Cell. Biol.* **2017**, *37*, e00652-16. [[CrossRef](#)]
100. Kim, H.J.; Kim, N.C.; Wang, Y.-D.; Scarborough, E.A.; Moore, J.; Diaz, Z.; MacLea, K.S.; Freibaum, B.; Li, S.; Molliex, A.; et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **2013**, *495*, 467–473. [[CrossRef](#)]
101. Paul, K.R.; Hendrich, C.G.; Waechter, A.; Harman, M.R.; Ross, E.D. Generating new prions by targeted mutation or segment duplication. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8584–8589. [[CrossRef](#)] [[PubMed](#)]
102. Vieira, N.M.; Naslavsky, M.; Licinio, L.; Kok, F.; Schlesinger, D.; Vainzof, M.; Sanchez, N.; Kitajima, J.P.; Gal, L.; Cavaçana, N.; et al. A defect in the RNA-processing protein HNRPD L causes limb-girdle muscular dystrophy 1G (LGMD1G). *Hum. Mol. Genet.* **2014**, *23*, 4103–4110. [[CrossRef](#)]
103. Kim, Y.; Kim, Y.-C.; Jeong, B.-H. Novel Single Nucleotide Polymorphisms (SNPs) and Genetic Features of the Prion Protein Gene (PRNP) in Quail (*Coturnix japonica*). *Front. Vet. Sci.* **2022**, *9*, 870735. [[CrossRef](#)] [[PubMed](#)]
104. Kim, D.-J.; Kim, Y.-C.; Kim, A.-D.; Jeong, B.-H. Novel Polymorphisms and Genetic Characteristics of the Prion Protein Gene (PRNP) in Dogs—A Resistant Animal of Prion Disease. *Int. J. Mol. Sci.* **2020**, *21*, 4160. [[CrossRef](#)] [[PubMed](#)]
105. Kim, Y.-C.; Won, S.-Y.; Do, K.; Jeong, B.-H. Identification of the novel polymorphisms and potential genetic features of the prion protein gene (PRNP) in horses, a prion disease-resistant animal. *Sci. Rep.* **2020**, *10*, 8926. [[CrossRef](#)] [[PubMed](#)]
106. Kim, S.-K.; Kim, Y.-C.; Won, S.-Y.; Jeong, B.-H. Potential scrapie-associated polymorphisms of the prion protein gene (PRNP) in Korean native black goats. *Sci. Rep.* **2019**, *9*, 15293. [[CrossRef](#)]
107. Kim, K.H.; Kim, Y.-C.; Jeong, B.-H. Novel Polymorphisms and Genetic Characteristics of the Prion Protein Gene in Pheasants. *Front. Vet. Sci.* **2022**, *9*, 935476. [[CrossRef](#)]
108. Kim, Y.-C.; Won, S.-Y.; Jeong, B.-H. Identification of Prion Disease-Related Somatic Mutations in the Prion Protein Gene (PRNP) in Cancer Patients. *Cells* **2020**, *9*, 1480. [[CrossRef](#)]
109. Batlle, C.; Yang, P.; Coughlin, M.; Messing, J.; Pesarrodona, M.; Szulc, E.; Salvatella, X.; Kim, H.J.; Taylor, J.P.; Ventura, S. hnRNPD L Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. *Cell Rep.* **2020**, *30*, 1117–1128.e5. [[CrossRef](#)]
110. Boncella, A.E.; Shattuck, J.E.; Cascarina, S.M.; Paul, K.R.; Baer, M.H.; Fomicheva, A.; Lamb, A.K.; Ross, E.D. Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5826–5835. [[CrossRef](#)]
111. 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. *Front. Mol. Biosci.* **2021**, *8*, 718301. [[CrossRef](#)] [[PubMed](#)]
112. Sormanni, P.; Aprile, F.A.; Vendruscolo, M. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* **2015**, *427*, 478–490. [[CrossRef](#)]
113. Luo, J.; Harrison, P.M. Evolution of sequence traits of prion-like proteins linked to amyotrophic lateral sclerosis (ALS). *PeerJ* **2022**, *10*, e14417. [[CrossRef](#)] [[PubMed](#)]

114. Santos, J.; Iglesias, V.; Pintado, C.; Santos-Suárez, J.; Ventura, S. DisPHred: A Server to Predict pH-Dependent Order–Disorder Transitions in Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 5814. [[CrossRef](#)]
115. Zamora, W.J.; Campanera, J.M.; Luque, F.J. Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations. *J. Phys. Chem. Lett.* **2019**, *10*, 883–889. [[CrossRef](#)]
116. Huang, F.; Oldfield, C.J.; Xue, B.; Hsu, W.-L.; Meng, J.; Liu, X.; Shen, L.; Romero, P.; Uversky, V.N.; Dunker, A.K. Improving protein order-disorder classification using charge-hydrophobicity plots. *BMC Bioinform.* **2014**, *15* (Suppl. 17), S4. [[CrossRef](#)]
117. Jacoby, G.; Asher, M.S.; Ehm, T.; Ionita, I.A.; Shinar, H.; Azoulay-Ginsburg, S.; Zemach, I.; Koren, G.; Danino, D.; Kozlov, M.M.; et al. Order from Disorder with Intrinsically Disordered Peptide Amphiphiles. *J. Am. Chem. Soc.* **2021**, *143*, 11879–11888. [[CrossRef](#)] [[PubMed](#)]
118. Pezzotti, G.; Ohgitani, E.; Fujita, Y.; Imamura, H.; Shin-Ya, M.; Adachi, T.; Yamamoto, T.; Kanamura, N.; Marin, E.; Zhu, W.; et al. Raman Fingerprints of the SARS-CoV-2 Delta Variant and Mechanisms of Its Instantaneous Inactivation by Silicon Nitride Bioceramics. *ACS Infect. Dis.* **2022**, *8*, 1563–1581. [[CrossRef](#)] [[PubMed](#)]
119. De Cena, G.L.; Scavassa, B.V.; Conceição, K. In Silico Prediction of Anti-Infective and Cell-Penetrating Peptides from *Thalassophryne nattereri* Natterin Toxins. *Pharmaceuticals* **2022**, *15*, 1141. [[CrossRef](#)]
120. Gomari, M.M.; Rostami, N.; Faradonbeh, D.R.; Asemaneh, H.R.; Esmailnia, G.; Arab, S.; Farsimadan, M.; Hosseini, A.; Dokholyan, N.V. Evaluation of pH change effects on the HSA folding and its drug binding characteristics, a computational biology investigation. *Proteins* **2022**, *90*, 1908–1925. [[CrossRef](#)]
121. Pintado-Grima, C.; Iglesias, V.; Santos, J.; Uversky, V.N.; Ventura, S. DisPHScan: A Multi-Sequence Web Tool for Predicting Protein Disorder as a Function of pH. *Biomolecules* **2021**, *11*, 1596. [[CrossRef](#)] [[PubMed](#)]
122. Pintado-Grima, C.; Bárcenas, O.; Ventura, S. In-Silico Analysis of pH-Dependent Liquid-Liquid Phase Separation in Intrinsically Disordered Proteins. *Biomolecules* **2022**, *12*, 974. [[CrossRef](#)] [[PubMed](#)]
123. Fernandez-Escamilla, A.-M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **2004**, *22*, 1302–1306. [[CrossRef](#)] [[PubMed](#)]
124. Pfeifferkorn, C.M.; McGlinchey, R.P.; Lee, J.C. Effects of pH on aggregation kinetics of the repeat domain of a functional amyloid, Pmel17. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21447–21452. [[CrossRef](#)]
125. Li, R.; Wu, Z.; Wang, Y.; Ding, L.; Wang, Y. Role of pH-induced structural change in protein aggregation in foam fractionation of bovine serum albumin. *Biotechnol. Rep.* **2016**, *9*, 46–52. [[CrossRef](#)]
126. Uversky, V.N. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *Protein J.* **2009**, *28*, 305–325. [[CrossRef](#)] [[PubMed](#)]
127. Pintado, C.; Santos, J.; Iglesias, V.; Ventura, S. SolupHred: A server to predict the pH-dependent aggregation of intrinsically disordered proteins. *Bioinformatics* **2020**, *37*, 1602–1603. [[CrossRef](#)]
128. Santos, J.; Iglesias, V.; Santos-Suárez, J.; Mangiagalli, M.; Brocca, S.; Pallarès, I.; Ventura, S. pH-Dependent Aggregation in Intrinsically Disordered Proteins Is Determined by Charge and Lipophilicity. *Cells* **2020**, *9*, 145. [[CrossRef](#)]
129. Uversky, V.N.; Li, J.; Fink, A.L. Evidence for a Partially Folded Intermediate in  $\alpha$ -Synuclein Fibril Formation. *J. Biol. Chem.* **2001**, *276*, 10737–10744. [[CrossRef](#)]
130. Jha, S.; Snell, J.M.; Sheftic, S.R.; Patil, S.M.; Daniels, S.B.; Kolling, F.W.; Alexandrescu, A.T. pH Dependence of Amylin Fibrillization. *Biochemistry* **2014**, *53*, 300–310. [[CrossRef](#)]
131. Hortschansky, P.; Schroeckh, V.; Christopheit, T.; Zandomeneghi, G.; Fändrich, M. The aggregation kinetics of Alzheimer’s  $\beta$ -amyloid peptide is controlled by stochastic nucleation. *Protein Sci.* **2005**, *14*, 1753–1759. [[CrossRef](#)] [[PubMed](#)]
132. Jeganathan, S.; von Bergen, M.; Mandelkow, E.-M.; Mandelkow, E. The Natively Unfolded Character of Tau and Its Aggregation to Alzheimer-like Paired Helical Filaments. *Biochemistry* **2008**, *47*, 10526–10539. [[CrossRef](#)] [[PubMed](#)]
133. Santos, J.; Pallarès, I.; Iglesias, V.; Ventura, S. Cryptic amyloidogenic regions in intrinsically disordered proteins: Function and disease association. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4192–4206. [[CrossRef](#)]
134. Pintado-Grima, C.; Bárcenas, O.; Manglano-Artuñedo, Z.; Vilaça, R.; Macedo-Ribeiro, S.; Pallarès, I.; Santos, J.; Ventura, S. CARs-DB: A Database of Cryptic Amyloidogenic Regions in Intrinsically Disordered Proteins. *Front. Mol. Biosci.* **2022**, *9*, 882160. [[CrossRef](#)] [[PubMed](#)]
135. Hughes, M.P.; Sawaya, M.R.; Boyer, D.R.; Goldschmidt, L.; Rodriguez, J.A.; Cascio, D.; Chong, L.; Gonen, T.; Eisenberg, D.S. Atomic structures of low-complexity protein segments reveal kinked  $\beta$  sheets that assemble networks. *Science* **2018**, *359*, 698–701. [[CrossRef](#)]
136. Diaz-Caballero, M.; Navarro, S.; Fuentes, I.; Teixidor, F.; Ventura, S. Minimalist Prion-Inspired Polar Self-Assembling Peptides. *ACS Nano* **2018**, *12*, 5394–5407. [[CrossRef](#)]
137. Diaz-Caballero, M.; Navarro, S.; Ventura, S. Functionalized Prion-Inspired Amyloids for Biosensor Applications. *Biomacromolecules* **2021**, *22*, 2822–2833. [[CrossRef](#)]

138. Peccati, F.; Díaz-Caballero, M.; Navarro, S.; Rodríguez-Santiago, L.; Ventura, S.; Sodupe, M. Atomistic fibrillar architectures of polar prion-inspired heptapeptides. *Chem. Sci.* **2020**, *11*, 13143–13151. [[CrossRef](#)]
139. 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. *Nat. Commun.* **2020**, *11*, 3314. [[CrossRef](#)]
140. Gondelaud, F.; Pesce, G.; Nilsson, J.F.; Bignon, C.; Ptchelkine, D.; Gerlier, D.; Mathieu, C.; Longhi, S. Functional benefit of structural disorder for the replication of measles, Nipah and Hendra viruses. *Essays Biochem.* **2022**, *66*, 915–934. [[CrossRef](#)]
141. Ventura, S. Sequence determinants of protein aggregation: Tools to increase protein solubility. *Microb. Cell Factories* **2005**, *4*, 11. [[CrossRef](#)] [[PubMed](#)]
142. Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* **2022**. [[CrossRef](#)] [[PubMed](#)]
143. Pinheiro, F.; Santos, J.; Ventura, S. AlphaFold and the amyloid landscape. *J. Mol. Biol.* **2021**, *433*, 167059. [[CrossRef](#)] [[PubMed](#)]
144. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.A.; Craig, P.; Crichtlow, G.V.; Dalenberg, K.; Duarte, J.M.; et al. RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **2022**, *51*, D488–D508. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.