Biochemistry and Biophysics Reports 29 (2022) 101202

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

The aqueous humor proteome is intrinsically disordered

Mak Djulbegovic [a,*], Vladimir N. Uversky [b,c]

[a] Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, 900 NW 17th St, Miami, FL, 33136, USA
[b] Department of Molecular Medicine, Morsani College of Medicine, University of South, Florida, Bruce B. Downs Blvd., MDC07, Tampa, FL, 33612, USA
[c] Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Institutskiy Pereulok, 9, Dolgoprudny, 141700, Moscow Region, Russia

1. Introduction

The aqueous humor is a colorless fluid that fills the anterior chamber of the eye. It functions to maintain intraocular pressure and nourish the lens and the cornea [1]. The composition of the aqueous humor is complex. One recent Fourier transform LTQ-Orbitrap Velos mass spectrometry study identified 763 proteins in the aqueous humor [1]. A knowledge gap remains as the intrinsic disorder propensity of the aqueous humor proteome has not yet been characterized.

Intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs) are proteins that local or global levels of non-distinct structural elements and are the most dynamic, functional regions of the protein [2–4]. IDPs and IDPRs have come into the literature as important biological entities and must be considered when thinking about proteins in any cell or compartment of the body. Characterizing these entities provide the community to understand the molecular dynamics that underlie the function of cells and other physiological systems.

We aim to characterize the previously identified 763 proteins of the aqueous humor and assess them for the presence of IDPs/IDPRs. If IDPs/IDPRs are abundant in the aqueous humor, then it can give insights into its molecular functions and interactions. The presence of these proteins may provide insight into the molecular interactions occurring within the milieu of the aqueous humor.

2. Methods

Protein sequences were collected from the Universal Protein Resource (UniProt) [5]. Of 763 proteins identified in the proteomics paper of the aqueous humor, 749 were used in subsequent analysis and 14 were not used as they were not on UniProt. The 749 protein sequences were first used the Composition Profiler (available at: http://www.cprofiler.org/) [6] to generate an amino acid composition profile of all the proteins contained within the aqueous humor. Our set of amino acid sequences was the query set and the 'Protein Data Bank Select 25' was the background set. We also generated a composition profile for experimentally validated disordered proteins from the DisProt database and distribution of amino acids in nature from the SwissProt database for comparison.

The next step in our analysis was directed at quantifying the intrinsic disorder. To quantify intrinsic disorder on per residue basis, we used the Predictor of Natural Disordered Protein Regions (PONDR®; available at: http://original.disprot.org/metapredictor.php), specifically PONDR® VSL2 for the initial part of the analysis [7]. To characterize intrinsic disorder on a whole protein basis, we used two binary predictors of disorder, charge-hydropathy (CH) plot and cumulative distribution function (CDF) (available at: http://www.pondr.com/). These binary predictors are used to characterize a protein as completely ordered or completely disordered and were combined to create a CH-CDF plot [8,9].

We then turned our analysis to further characterizing the most disordered proteins in the protein set. We determined the 10 most disordered proteins by PONDR®-VSL2 score. We used other predictors of intrinsic disorder to further characterize the most disordered protein and further validate our findings that these select proteins are highly disordered. Score and percent for three more PONDR® predictor were collected and these predictors were PONDR®-VL3 score, PONDR® VLXT score, and PONDR® FIT score [7,10,11]. In addition, the IUPred2A platform (available at: https://iupred2a.elte.hu/) was used to generate predictions for short and long disordered regions [11]. A mean disorder profile (MDP) score and percent were calculated by averaging the prediction value of the PONDR® and IUPred2A platforms. In these analyses, proteins were classified based on their percent of predicted disordered residues (PPDR). Here, two arbitrary cutoffs for the levels of intrinsic disorder are used to classify proteins as highly ordered (PPDR$<$10%), moderately disordered (10%$\leq$PPDR$<$30%) and highly disordered (PPDR$\geq$30%) [12]. Since the average disorder score (ADS) of a given protein is not directly related to its PPDR value (e.g. theoretically, a protein with the PPDR of 100% might have the ADS ranging from 0.5 to 1.0; whereas a protein with the PPDR of 0% might have any ADS $<$0.5), we also evaluated disorder status of proteins using this criterion. Here,

* Corresponding author.
E-mail address: mbd83@med.miami.edu (M. Djulbegovic).

https://doi.org/10.1016/j.bbrep.2022.101202
Received 7 December 2021; Received in revised form 3 January 2022; Accepted 4 January 2022

Abbreviations

IDP	Intrinsically disordered proteins
IDPR	Intrinsically disordered protein regions
UniProt	Universal Protein Resource
PONDR®	Predictor of Natural Disordered Protein Regions
CH	Charge-Hydropathy
CDF	Cumulative Distribution Function
MDP	Mean Disorder Profile
PPDR	Percent of Predicted Disordered Residues
ADS	Average Disorder Score
MoRF	Molecular Recognition Features
IUPred	Web Server for The Prediction Of Intrinsically, Unstructured Regions of Proteins
GO	Gene Ontology
BASP1	Brain acid soluble protein 1
TYB4	Thymosin beta-4
MTIX	Metallothionein-1X
PRB2	Basic salivary proline-rich protein 2
PRB4	Basic salivary proline-rich protein 4
HORN	Hornerin
KRA94	Keratin-associated protein 9-4
TYB10	Thymosin beta-10
FILA2	Filaggrin-2
FILA	Filaggrin
KPRP	Keratinocyte proline-rich protein
COL9A2	Collagen alpha-2(IX) chain
VGF	Neurosecretory protein VGF
LLPS	Liquid-liquid phase separation

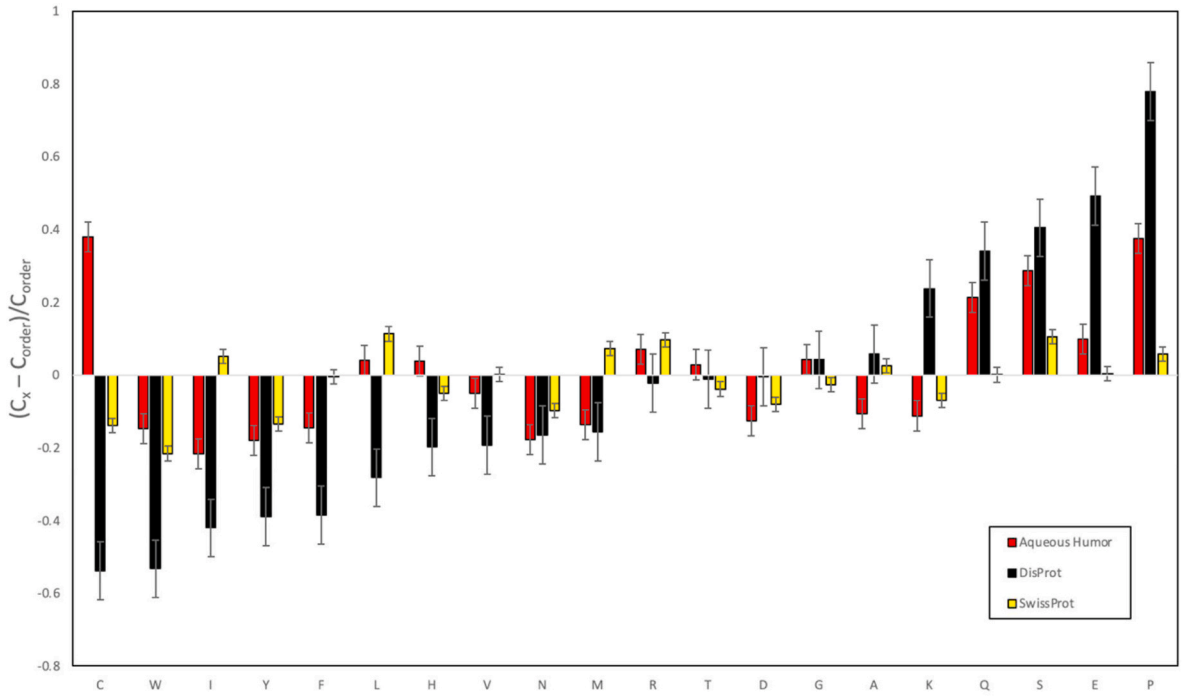


Fig. 1. Amino acid composition profile of 749 humor aqueous proteins (red bars). The fractional difference is calculated as $(C_x - C_{order})/C_{order}$, where C_x is the content of a given amino acid in the query set (749 amino acid sequences of humor aqueous proteins or proteins from SwissProt database) and C_{order} is the content of a given of a given amino acid in the background set (Protein Databank Select 25). The amino acid residues are ranked from most order promoting residue to most disorder promoting residue. Positive values indicate enrichment and negative values indicate depletion of a particular amino acid. Composition profile generated for experimentally validated disordered proteins from DisProt database (black bars) [29] and distribution of amino acids in nature from the SwissProt database (yellow bars) [32] are shown for comparison. In both cases, error bars correspond to standard deviations over 10,000 bootstrap iterations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

protein/region/residue is considered ordered if it has ADS<0.15. When 0.15 ≤ ADS<0.5, is considered as moderately disordered or flexible, whereas ADS≥0.5 correspond to disordered protein/region/residue.

Complementary disorder evaluations together with important disorder-related functional information were retrieved from the D^2P^2 database (http://d2p2.pro/) [13], which is a database of predicted disorder for a large library of proteins from completely sequenced genomes [13]. D^2P^2 database uses outputs of IUPred [14], PONDR® VLXT [15], PrDOS [16], PONDR® VSL2B [17,18], PV2 [13], and ESpritz [19]. The visual console of D^2P^2 displays 9 colored bars representing the location of disordered regions as predicted by these different disorder predictors. In the middle of the D^2P^2 plots, the blue-green-white bar shows the predicted disorder agreement between nine disorder predictors (IUPred, PONDR® VLXT, PONDR® VSL2, PrDOS, PV2, and ESpritz), with blue and green parts corresponding to disordered regions by consensus. Above the disorder consensus bar are two lines with colored and numbered bars that show the positions of the predicted (mostly structured) SCOP domains [20,21] using the SUPERFAMILY predictor [22]. Yellow zigzagged bar shows the location of the predicted disorder-based binding sites (molecular recognition features (MoRF)

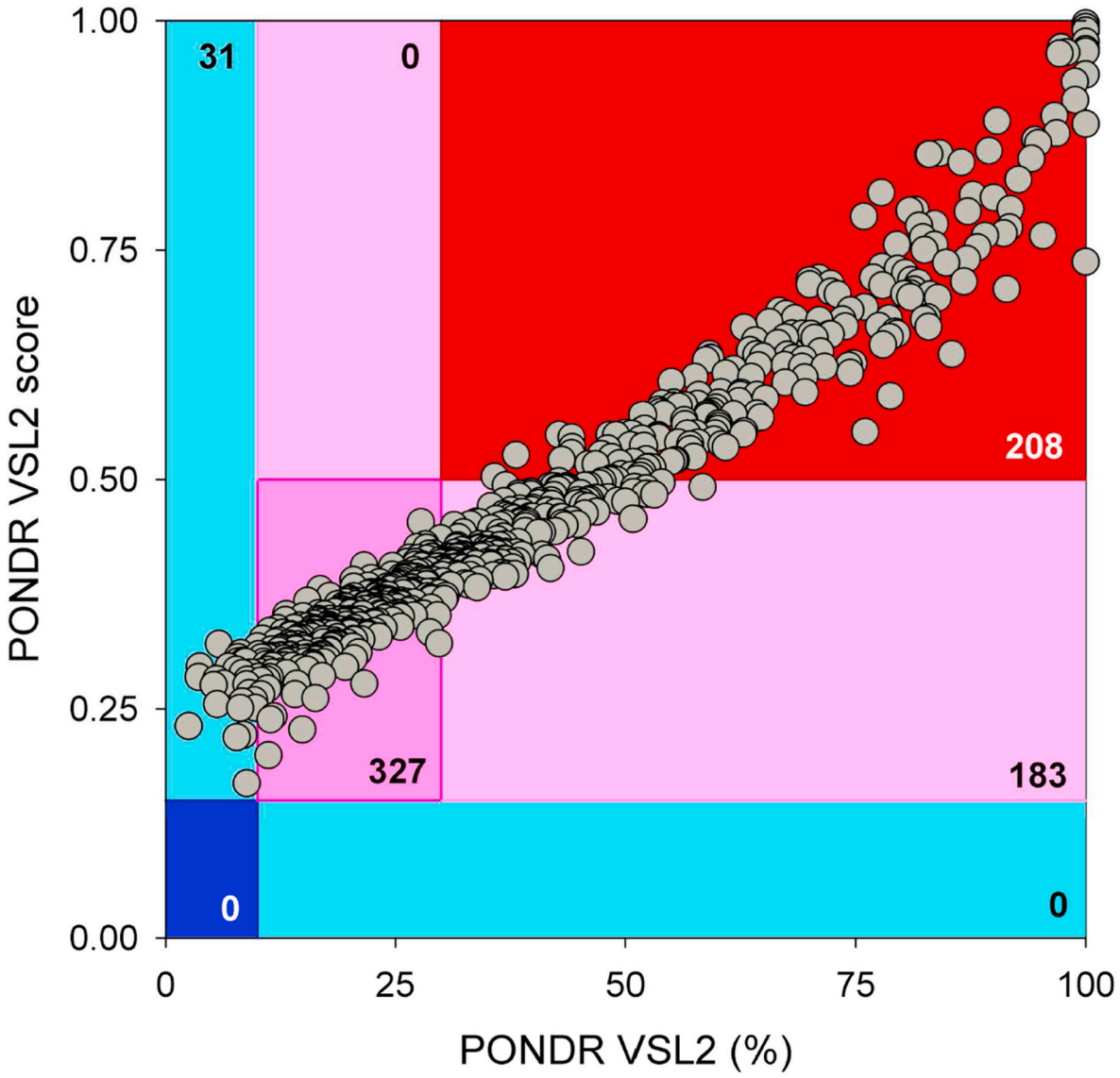


Fig. 2. Prediction of Natural Disordered Regions (PONDR®) VSL2 output for 749 aqueous humor proteins. PONDR® VSL2 score is the average disorder score (ADS) for a protein. PONDR® VSL2 (%) is a percent of predicted disordered residues (PPDR); i.e., residues with disorder scores above 0.5. Color blocks indicate regions in which proteins are mostly ordered (blue and light blue), moderately disordered (pink and light pink), or mostly disordered (red). If the two parameters agree, the corresponding part of background is dark (blue or pink), whereas light blue and light pink reflect areas in which only one of these criteria applies. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

regions) identified by the ANCHOR algorithm [23], whereas differently colored circles at the bottom of the plot show location of various PTMs assigned using the outputs of the PhosphoSitePlus platform [24], which is a comprehensive resource of the experimentally determined post-translational modifications.

We also utilized the power of AlphaFold, a novel deep learning algorithm that incorporates physical and biological knowledge about protein structure to generate highly accurate predictions of protein structures [25,26], to create a gallery of structures for the most disordered aqueous humor proteins.

To determine the degree of interactivity between the aqueous humor proteins, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; available at: https://string-db.org/) [27]. STRING generates a PPI network based on the predicted and experimentally-validated information on the interaction partners for a protein of interest or a set of proteins. This analysis was conducted for 745 human aqueous humor proteins, as no STRING-based information was available for the 4 proteins from the original set. The corresponding protein-protein interaction network was built using the medium confidence of 0.4 as a minimum required interaction score. At the next stage, the top 10 most disordered proteins were used as the query sequences for the focused analysis of their interactivity.

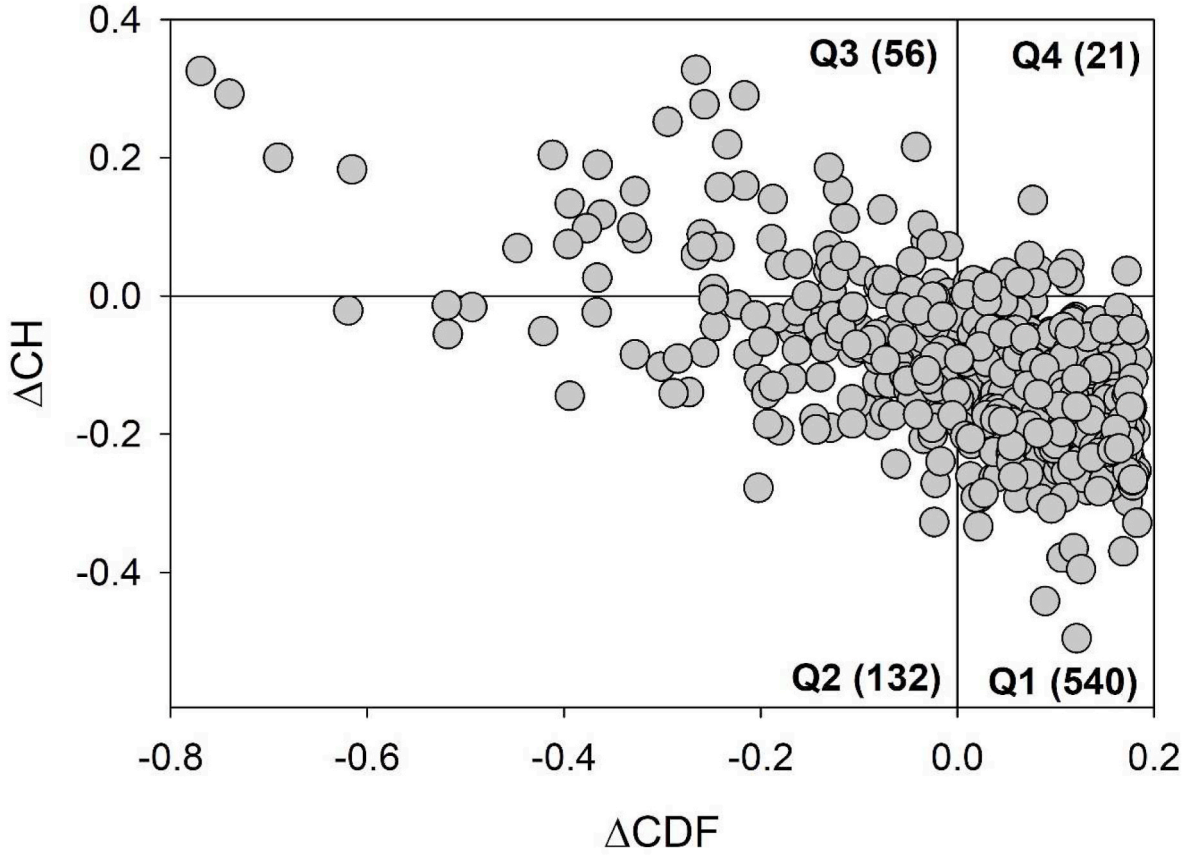


Fig. 3. Charge-hydropathy and cumulative distribution function (CH-CDF) plot for 749 aqueous humor proteins (grey dots). The Y-coordinate is calculated as the distance of the corresponding protein from the boundary in the CH plot. The X-coordinate is calculated as the average distance of the corresponding protein's x curve from the CDF boundary. The quadrant that the protein is located determines its classification. Q1, protein predicted to be disordered by CH-plot and ordered by CDF. Q2, protein predicted to be ordered by CH-plot and CDF. Q3, protein predicted to be ordered to by CH-plot and disordered by CDF-plot. Q4, protein predicted to be disordered by CH-plot and CDF.

3. Results

The amino acid composition of the 749 aqueous humor proteins, the DisProt proteins, and SwissProt proteins were plotted side by side to allow for visualization and appreciation of key differences (Fig. 1). Each amino acid was ranked from most order-promoting residues (i.e., C, W, I, Y, F, L, H, V, N, and M) to most disorder-promoting residues (i.e., R, T, D, G, A, K, Q, S, E, P). Positive numbers indicate enrichment and negative numbers indicate depletion. Only 3 of 10 order-promoting residues (C, L, and H) showed enrichment and 7 of 10 disorder-promoting residues showed enrichment (R, T, G, Q, S, E, P). The composition profile of the aqueous humor matches many amino acid residues as the DisProt protein set. For example, the most disordered residues (Q, S, E, and P) all show enrichment, which is not consistent with the SwissProt protein set. A major exception is cysteine (C) as it highly enriched in the aqueous humor and highly depleted in the DisProt protein set. This key difference may be attributable to the extracellular nature of the aqueous humor, where C is typically used to help stabilize proteins. The initial results of our analysis demonstrate that the proteins found in the aqueous humor have many sequences features characteristic of intrinsically disordered proteins.

At the per-amino acid residue basis, the PONDR® VSL2 output confirms the presence of many intrinsically disordered or partially intrinsically disordered proteins in the aqueous humor (Fig. 2). In fact, of the 749 proteins found in the aqueous humor, 208 are predicted to be highly disordered, 510 show some evidence of moderate disorder or conformational flexibility (327–183), and 31 are predicted to be highly ordered.

We further characterized these proteins with additional PONDR® predictors including PONDR® VL3, PONDR® VL-XT, and PONDR® FIT. The corresponding PPDR and ADS values calculated for all 749 proteins found in the aqueous humor are listed in Supplementary Table S1. This analysis confirmed prevalence of disorder in these proteins. Our PONDR®-centric analysis consistently demonstrated agreement between intrinsic disorder predispositions evaluated by these tools: if the protein was predicted to be highly disordered by one PONDR® predictor then it was likely highly disordered by other PONDR® predictors. We decided to use another predictor of intrinsic disorder, the web server for the prediction of intrinsically unstructured regions of proteins (IUPred), to externally validate our PONDR® findings. To this end, per-residue disorder profiles were generated by IUPred that then were used for calculation of corresponding PPDR and ADS values, which are

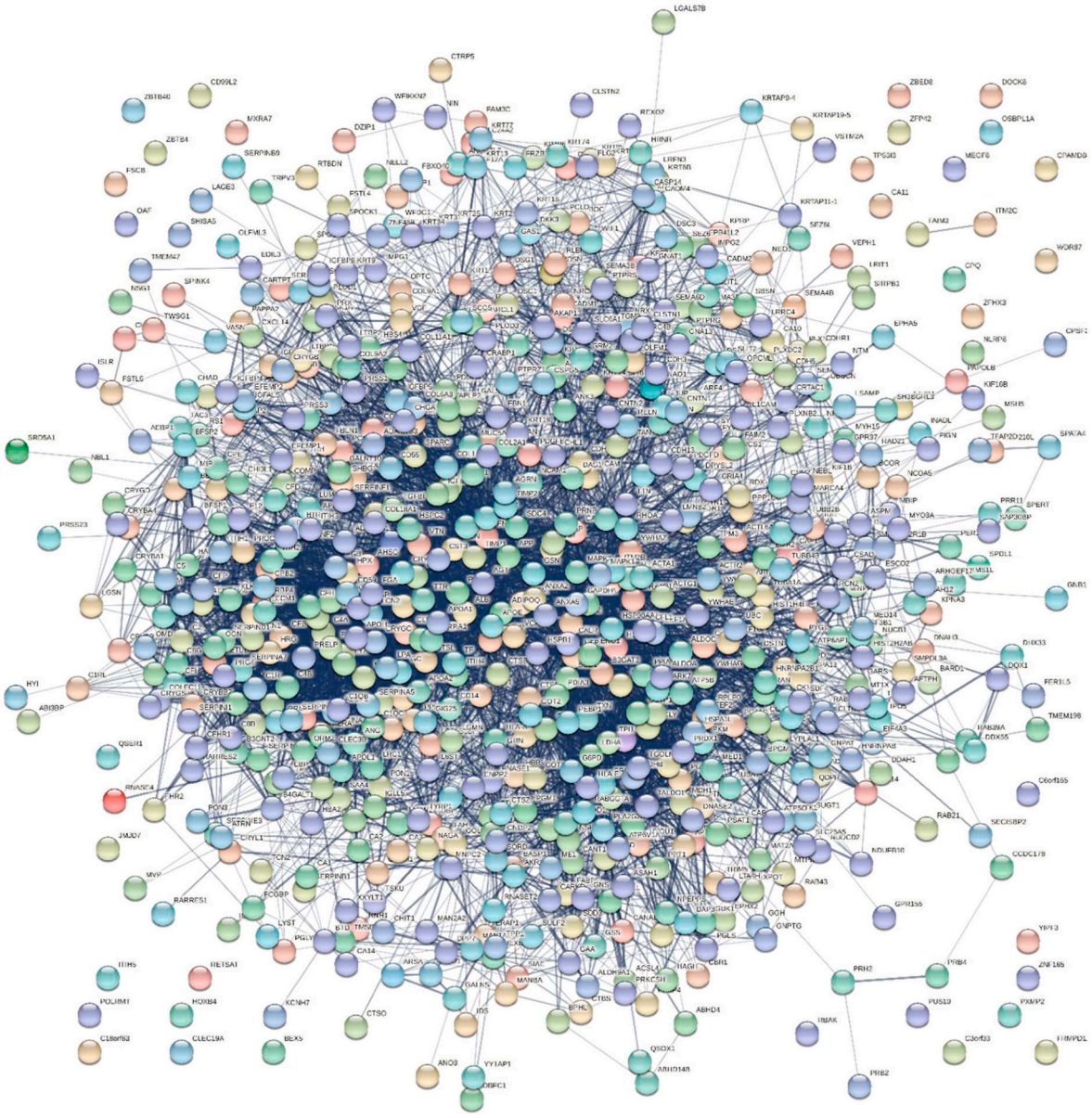


Fig. 4. STRING-based analysis of the inter-set interactivity of 745 human proteins using the medium confidence level of 0.4. This confidence level was selected to ensure maximal inclusion of TPR proteins in the resulting PPI. The nodes correspond to proteins, whereas the edges show predicted or known functional associations. Seven types of evidence are used to build the corresponding network, and are indicated by the differently colored lines: a green line represents neighborhood evidence; a red line – the presence of fusion evidence; a purple line – experimental evidence; a blue line – co-occurrence evidence; a light blue line – database evidence; a yellow line – text mining evidence; and a black line – co-expression evidence [33]. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

summarized in Table S1. This analysis revealed reasonable agreement between the results generated by IUPred run in the long and short modes and the outputs of various PONDRs®. As our last approach for quantitative evaluation of disorder, we generated mean disorder profiles (MDPs) for all the 749 proteins found in the aqueous humor and calculated their corresponding PPDR and ADS values (see Table S1). This additional quantitative characterization of intrinsic disorder strongly supported our original findings that the aqueous humor contains proteins with variable levels of disorder. In fact, MDP-based PPDR analysis demonstrated that 203 and 257 are expected to be highly or moderately disordered, and 289 can be classified as ordered. Furthermore, in a good agreement with data shown in Fig. 2, MDP-based ADS analysis revealed that only 8, 643, and 98 proteins are expected to be highly ordered, moderately disordered/flexible, and disordered, respectively. Therefore, irrespectively of tool used for these analyses, our study indicated that a significant proportion of aqueous humor

Table 1
Name, Universal Protein Resource (UniProt ID), and molecular function as identified on UniProt of the, POND®-VSL2 identified, top 10 most disordered proteins in the aqueous humor. Three pairs of proteins in the top 10 belonged to similar families (e.g. pairs TYB4 and TYB10, PRB4 and PRB2, and FILA2 and FILA), therefore, three additional proteins were considered (e.g. KPRP, COL9A2, VGF).

Protein Name	Abbreviation	Molecular Function	UniProt ID	PONDR® VLS2 Score
Brain acid soluble protein 1	BASP1	Transcription corepressor/regulatory activity	P80723	0.99708
Thymosin beta-4	TYB4	Actin monomer/enzyme/RNA binding	P62328	0.99311
Metallothionein-1X	MT1X	Metal/Zinc ion binding	P80297	0.98922
Basic salivary proline-rich protein 4	PRB4	Major components of parotid and submandibular saliva	P10163	0.97708
Basic salivary proline-rich protein 2	PRB2	Major components of parotid and submandibular saliva	P02812	0.9716
Hornerin	HORN	Calcium/transition metal ion binding	Q86YZ3	0.97107
Keratin-associated protein 9-4	KRA94	Keratinization	Q9BYQ2	0.96893
Thymosin beta-10	TYB10	Actin monomer binding	P63313	0.96642
Filaggrin-2	FILA2	Calcium/transition metal ion binding Structural molecule activity Structural constituent of the skin epidermis	Q5D862	0.9655
Filaggrin	FILA	Calcium/transition metal ion binding Structural molecule activity Structural constituent of the skin epidermis	P20930	0.96438
Keratinocyte proline-rich protein	KPRP	A proline-rich skin protein possibly involved in keratinocyte differentiation	Q5T749	0.9416
Collagen alpha-2(IX) chain	COL9A2	Structural component of hyaline cartilage and vitreous of the eye	Q14055	0.9338
Neurosecretory protein VGF	VGF	VGF and peptides derived from its processing play many roles in neurogenesis and neuroplasticity associated with learning, memory, depression and chronic pain	O15240	0.9137

proteome is expected to be disordered.

Next, the CH-CDF plot (a combined binary predictor of intrinsic disorder) verified that intrinsic disorder is found extensively in the 749 aqueous humor proteins (Fig. 3). The CH-CDF plot allowed us to characterize each protein depending on the quadrant the protein was plotted. 540 proteins are in the quadrant Q1 and are predicted to be ordered by both predictors. 132 proteins in Q2 are molten globular proteins (compact, but without unique 3D structures) and/or hybrid proteins containing comparable levels of ordered and disordered residues; proteins in this quadrant are predicted to be ordered/compact by CH-plot and disordered by CDF. 56 proteins in Q3 are highly disordered, being predicted to be disordered by CH-plot and CDF-plot. Finally, 21 proteins in Q4 are predicted to be disordered by CH-plot and ordered by CDF-plot. Therefore, 209 proteins from human aqueous humor are predicted as containing noticeable levels of disorder. These results further validate the presence of intrinsic disorder in the aqueous humor.

Fig. 4 represents protein-protein interaction network generated for human aqueous humor proteins by using STRING. Most of these proteins are involved in the formation of rather dense interaction network. In fact, of 745 analyzed proteins only 36 (4.5%) were loners, whereas remaining 709 proteins were engaged in 7641 interactions thereby organizing a network with average node degree of 20.5 (i.e., on average, each protein there interacts with 20 partners). This network is characterized by the average local clustering coefficient of 0.39. Average local clustering coefficient defines how close the neighbors are to being a complete clique. If a local clustering coefficient is equal to 1, then every neighbor connected to a given node N_i is also connected to every other node within the neighborhood. If the local clustering coefficient is equal to 0, then no node that is connected to a given node N_i connects to any other node that is connected to N_i.

Since the expected number of interactions in a similar size set of proteins randomly selected from human proteome is equal to 3,060, this STRING-generated PPI network has significantly more interactions than expected, being characterized by a PPI enrichment p-value of $<10^{-16}$. This observation indicates that the query proteins in the analyzed PPI network have more interactions among themselves than what would be expected for a random set of proteins of similar size. Therefore, such an enrichment indicates that the proteins are at least partially biologically connected, as a group. Analysis of this PPI network for GO-centered functional enrichment revealed that most prominent biological processes (a), molecular function (b) and cellular components are: (a) Regulated exocytosis (GO:0045,055; $p = 3.37 \times 10^{-49}$), Exocytosis (GO:0006887; $p = 1.39 \times 10^{-44}$), Secretion by cell (GO:0032,940; $p = 1.20 \times 10^{-42}$), Leukocyte mediated immunity (GO:0002443; $p = 6.90 \times 10^{-42}$), and Secretion (GO:0046,903; $p = 1.07 \times 10^{-41}$); (b) Endopeptidase inhibitor activity (GO:0004866; $p = 3.08 \times 10^{-24}$), Endopeptidase regulator activity (GO:0061,135; $p = 3.08 \times 10^{-24}$), Peptidase regulator activity (GO:0061,134; $p = 3.08 \times 10^{-24}$), Structural molecule activity (GO:0005198; $p = 3.08 \times 10^{-24}$), Serine-type endopeptidase inhibitor activity (GO:0004867; $p = 2.42 \times 10^{-20}$); and (c) Extracellular region (GO:0005576; $p = 2.69 \times 10^{-172}$), Extracellular space (GO:0005615; $p = 1.47 \times 10^{-171}$), Extracellular vesicle (GO:1903561; $p = 8.52 \times 10^{-148}$), Extracellular exosome (GO:0070,062; $p = 6.20 \times 10^{-146}$), and Vesicle (GO:0031,982; $p = 9.82 \times 10^{-113}$).

Next, we decided to have a closer look at the most disordered members from the set of human aqueous humor proteome. For this purpose, we selected proteins characterize by the highest average disorder scores predicted by PONDR® VSL2 ($ADS_{PONDR\text{-}VSL2}$). This predictor was selected for this analysis based on its exceptional performance at the recent Critical Assessment of protein Intrinsic

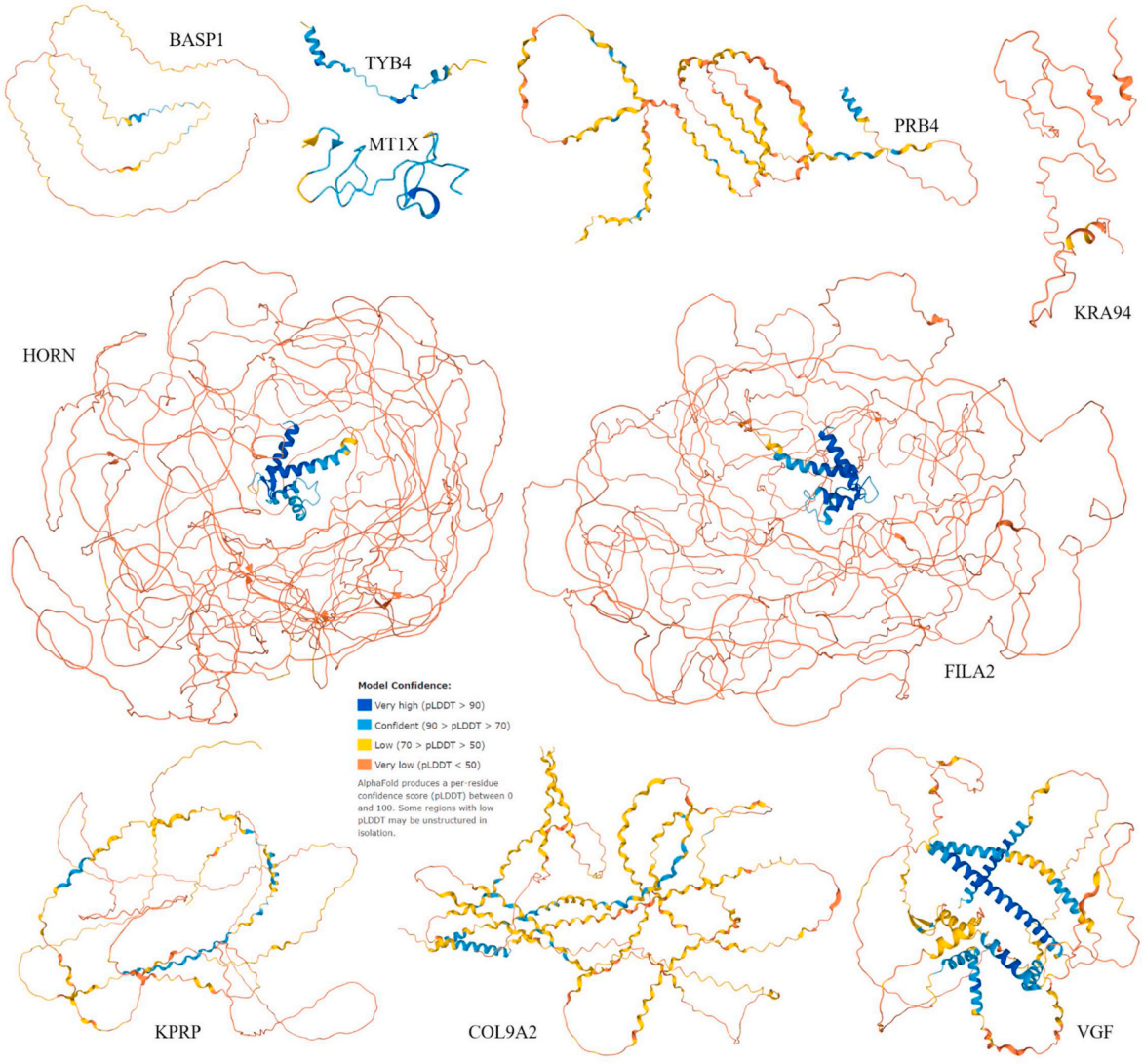


Fig. 5. Structures generated for the most disordered human aqueous humor protein by AlphaFold2. Note that the most structures are characterized by the presence of regions with low per-residue confidence scores suggesting that these regions are unstructured (corresponding parts are shown by yellow and orange colors). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Disorder prediction (CAID) experiment [28], where it was ranked #3 of 43 methods evaluated on a dataset of 646 proteins from DisProt [29]. The 10 most disordered proteins as determined by PONDR® VSL2 were Brain acid soluble protein 1 (BASP1; $ADS_{PONDR\text{-}VSL2}$: 0.99708), Thymosin beta-4 (TYB4; 0.99311), Metallothionein-1X (MTIX; 0.98922), Basic salivary proline-rich protein 4 (PRB4; 0.97708), Basic salivary proline-rich protein 2 (PRB2; 0.9716), Hornerin (HORN; 0.97107), Keratin-associated protein 9-4 (KRA94; 0.96893), Thymosin beta-10 (TYB10; 0.96642), Filaggrin-2 (FILA2; 0.9655), and Filaggrin (FILA; 0.96438) (Table 1). Since some of the members of these subset belonged to similar families (e.g. pairs TYB4 and TYB10, PRB4 and PRB2, and FILA2 and FILA) we decided to keep only one representative of each pair and included three additional proteins: Keratinocyte proline-rich protein (KPRP; 0.9416), Collagen alpha-2(IX) chain (COL9A2; 0.9338), and Neurosecretory protein VGF (VGF: 0.9137). Subsequent sections represent some interesting observations related to these proteins. Amino acid sequences of these proteins are shown in Supplementary materials (Fig. S1).

Fig. 5 shows corresponding structure and clearly illustrates that all these proteins are highly disordered, which is in complete agreement with the results of the evaluation of their global disorder predisposition (see Table 1).

Per-residue disorder profiles generated for 10 most disordered proteins by several commonly used predictors are assembled in Fig. 6. Remarkable agreement between the outputs of six tools used for evaluation of disorder predisposition is observed for seven proteins. The noticeable exceptions are MT1X, KRA94, and KPRP because their disorder profiles of which are characterized by the dramatic variability of

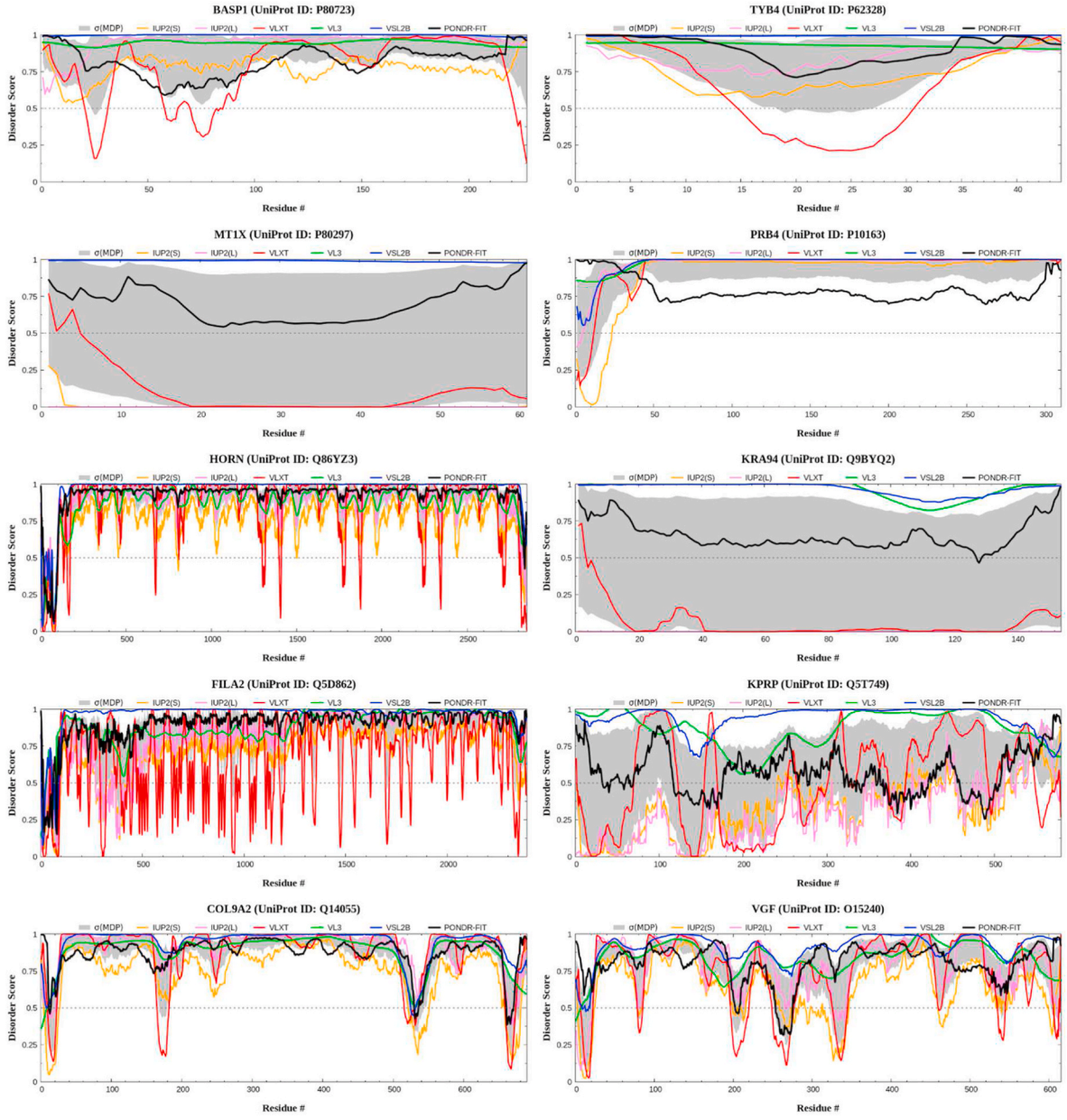


Fig. 6. Per-residue disorder profiles generated for 10 most disordered human aqueous humor proteins. Profiles were generated by the DiSpi web crawler that was designed for the rapid prediction and comparison of protein disorder profiles. It aggregates the results from a number of well-known disorder predictors: PONDR® VLXT [15], PONDR® VL3 [18], PONDR® VLS2B [34], PONDR® FIT [35], IUPred2 (Short) and IUPred2 (Long) [14,36]. The outputs of the evaluation of the per-residue disorder propensity by these tools are represented as real numbers between 0 (ideal prediction of order) and 1 (ideal prediction of disorder). A threshold of $\geq$0.5 is used to identify disordered residues and regions in query proteins.

the outputs of individual predictors. In fact, for these proteins, difference between the outputs of the tools is approaching 100%. This high degree of confusion is defined by the highly biased amino acid compositions of these three proteins and their high content of cysteine residues, which is mounting to 9.3%, 32.5%, and 32.8% in KPRP, KRA94, and MT1X, respectively (note that the average cysteine content in the UniProtKB/SwissProt databank is 1.37%). Since intrinsic disorder predisposition of multi-cysteine-containing proteins depends on the redox potential of the environment, a specialized tool, IUPred2A_redox, was elaborated to find redox-sensitive regions in proteins based on the energy estimation method [18]. Results of application of this tool to the human KPRP, KRA94, and MT1X are shown in Fig. 7. Based on this analysis, we can conclude that all three proteins represent conditionally disordered redox-sensitive proteins, as their redox-plus and redox-minus profiles are significantly separated.

Fig. 8 represents functional disorder profiles generated for 8 human aqueous tumor proteins (HORN, VGF, FILA2, COL9A2, BASP1, KPRP, KRA94, and MT1X) by the D^2P^2 platform. This analysis not only supported highly disordered status of these proteins, but also indicated that intrinsically disordered regions are used by these proteins as they

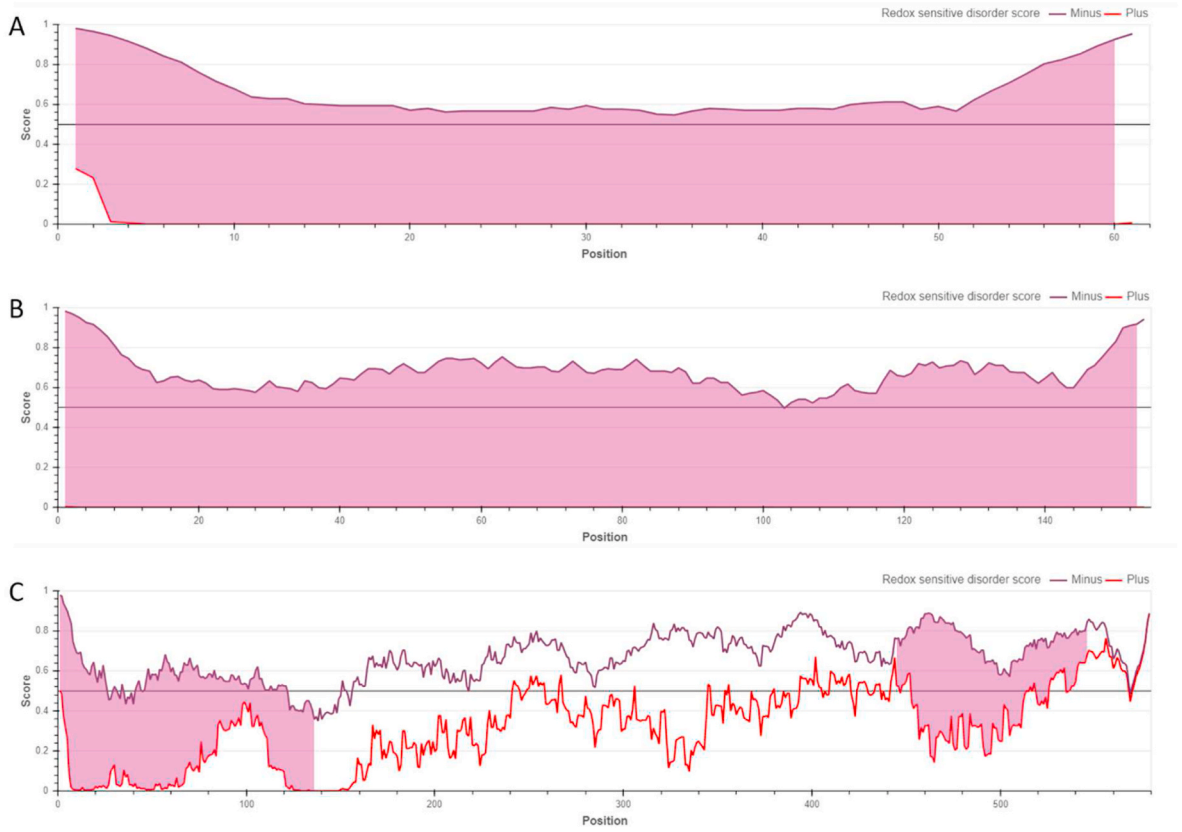
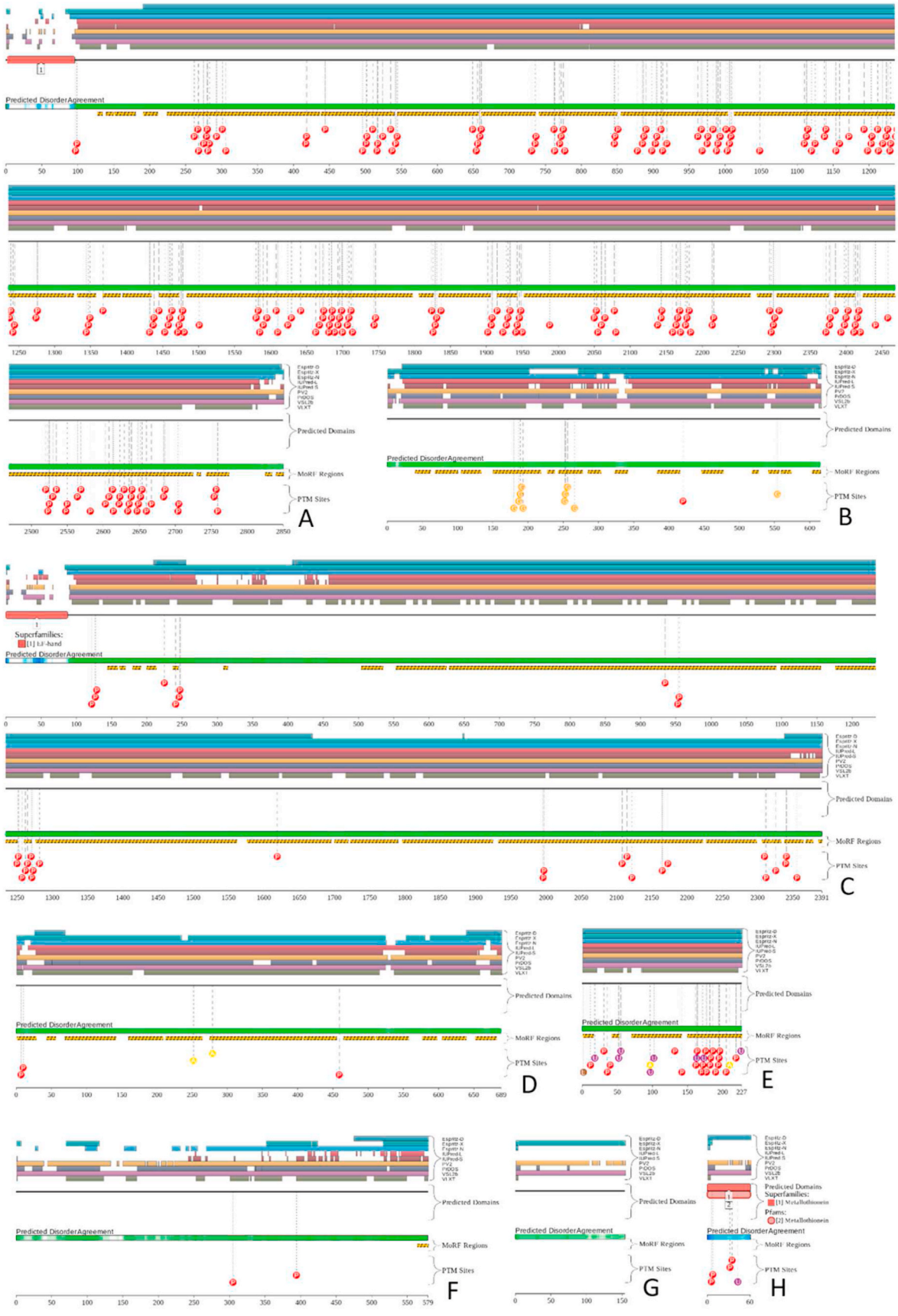


Fig. 7. Evaluation of redox-sensitive disorder predispositions of MT1X (**A**), KRA94 (**B**), and KPRP (**C**) by IUPred2A_redox. In corresponding plots, disorder profiles for the reduced (redox-minus) and oxidized (redox-plus) forms of proteins are shown by violet and red lines correspondingly, whereas shaded areas show redox sensitive regions. It is seen that the entire MT1X, KRA94, and KPRP can be considered as conditionally disordered redox-sensitive proteins. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

contain multiple sites of different posttranslational modifications and numerous molecular recognition features (MoRFs), which are the disorder-based protein-protein interaction sites that can fold at interaction with the binding partners. In fact, BASP1, COL9A2, FILA2, HORN, and VGF are exceptionally enriched in MoRFs, which occupy almost entire amino acid sequences of these five proteins.

We next turned our attention to understanding the intra-set and set-centric interactivity of 10 most disordered proteins from aqueous humor. Our STRING analysis demonstrated only 3 of 10 most intrinsically disordered proteins of the aqueous humor (FLG2, KPRP, and HRNR) are interacting with each other (Fig. 9A). However, when a first shell interactors (i.e., proteins interacting with these 10 proteins) were included into consideration, a rather connected common protein-protein interaction network was formed (see Fig. 9B) via overlapping of the local networks centered at the 10 individual proteins. With the parameters used for this analysis (minimum required interaction score of 0.550), only cluster of the KRA94 interacting proteins is not included into the common network. The resulting network has 195 nodes and 1358 edges, therefore being characterized by the average node degree of 13.9. This network is rather well connected showing the average local clustering coefficient of 0.674. As only 349 edges were expected for the similarly sized random set of proteins, the PPI enrichment score was a p-value $< 10^{-16}$. Therefore, this analysis revealed that there is an extensive network of PPIs that involve the 9 of the 10 most disordered proteins from aqueous humor. The ability of these proteins to interact with many unique binding partners is likely due to intrinsic disorder embedded into their structural framework. Analysis of this PPI network for GO-centered functional enrichment revealed that most prominent biological processes are Extracellular matrix organization (GO:0030,198; $p = 7.91 \times 10^{-35}$), Tissue development (GO:0009888; $p = 1.04 \times 10^{-34}$), System development (GO:0048,731; $p = 1.02 \times 10^{-23}$), Integrin-mediated signaling pathway (GO:0007229; $p = 1.87 \times 10^{-21}$), and Skin development (GO:0043,588; $p = 3.37 \times 10^{-21}$). Most common biological functions of these proteins are Collagen binding (GO:0005518; $p = 3.65 \times 10^{-14}$), Integrin binding (GO:0005178; $p = 1.29 \times 10^{-10}$), Protein-containing complex binding (GO:0044,877; $p = 2.01 \times 10^{-10}$), L-ascorbic acid binding (GO:0031,418; $p = 8.13 \times 10^{-09}$) and Transition metal ion binding (GO:0046,914; $p = 1.58 \times 10^{-08}$), and their most common cellular compartments are Integrin complex (GO:0008305; $p = 2.53 \times 10^{-24}$), Extracellular region (GO:0005576; $p = 1.70 \times 10^{-19}$), Endoplasmic reticulum lumen (GO:0005788; $p = 2.48 \times 10^{-16}$), Collagen-containing extracellular matrix (GO:0062,023; $p = 1.48 \times 10^{-15}$) and Focal adhesion (GO:0005925; $p = 2.24 \times 10^{-15}$).



(caption on next page)

Fig. 8. Functional disorder profiles of human HORN (**A**), VGF (**B**), FILA2 (**C**), COL9A2 (**D**), BASP1 (**E**), KPRP (**F**), KRA94 (**G**), and MT1X (**H**) generated by D^2P^2 platform (http://d2p2.pro/) [13], which is a database of predicted disorder for a large library of proteins from completely sequenced genomes [13]. D^2P^2 database uses outputs of IUPred [14], PONDR® VLXT [15], PrDOS [16], PONDR® VSL2B [17,18], PV2 [13], and ESpritz [19]. The visual console of D^2P^2 displays 9 colored bars representing the location of disordered regions as predicted by these different disorder predictors. In the middle of the D^2P^2 plots, the blue-green-white bar shows the predicted disorder agreement between nine disorder predictors (IUPred, PONDR® VLXT, PONDR® VSL2, PrDOS, PV2, and ESpritz), with blue and green parts corresponding to disordered regions by consensus. Above the disorder consensus bar are two lines with colored and numbered bars that show the positions of the predicted (mostly structured) SCOP domains [20,21] using the SUPERFAMILY predictor [22]. Yellow zigzagged bar shows the location of the predicted disorder-based binding sites (MoRF regions) identified by the ANCHOR algorithm [23], whereas differently colored circles at the bottom of the plot show location of various PTMs assigned using the outputs of the PhosphoSitePlus platform [24], which is a comprehensive resource of the experimentally determined post-translational modifications. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

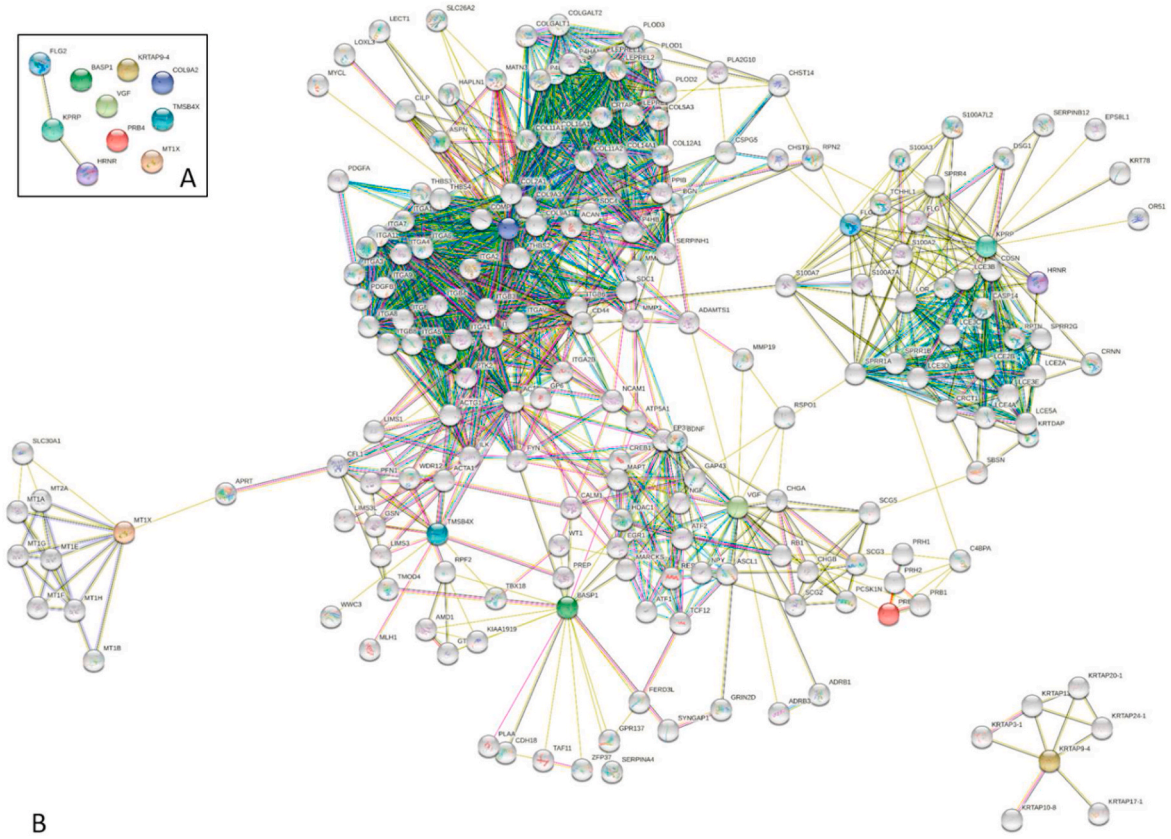


Fig. 9. Interactability of 10 most disordered proteins in aqueous humor analyzed by Search Tool for the Retrieval of Interacting Genes (STRING) platform. **A.** Intraset protein-protein interaction network. **B.** Network representing the first shell of interactors; i.e., proteins interacting with 10 most disordered proteins in aqueous humor. Both networks were generated using confidence level of 0.55 as minimum required interaction score.

4. Discussion

Our study demonstrates that intrinsic disorder is abundant in the 749 aqueous humor proteins. The binary predictors, charge hydropathy and cumulative distribution function, showed that many proteins in the aqueous humor can be considered highly disordered. 208 proteins were found to be highly disordered as measured by PONDR® VSL2, and MDP-based analysis showed that 203 and 257 proteins are expected to be highly or moderately disordered. The 10 most disordered proteins as determined by PONDR®-VSL2 score were BASP1 (PONDR®-VSL2 score: 0.99708), TYB4 (0.99311), MTIX (0.98922), PRB4 (0.97708), PRB2 (0.9716), HORN (0.97107), KRA94 (0.96893), TYB10 (0.96642), FILA2 (0.9655), and FILA (0.96438). In subsequent in-depth analyses, this list was extended to include KPRP (0.9416), COL9A2 (0.9338), and VGF (0.9137) due to the presence of homologous pairs TYB4-TYB10, PRB4-PRB2, and FILA2-FILA in the original set. Focused analysis of these 10 highly disordered proteins suggested that the presence of intrinsic disorder in the aqueous humor proteins is crucial for their functionality. We were able to demonstrate that intrinsic disorder allows the proteins of the aqueous humor to interact with many different binding partners and to be regulated by multiple posttranslational modifications, which gives us hints to the complex nature of the fluid. One possible function these disordered proteins have is liquid-liquid phase separation (LLPS). IDPs have been linked to LLPS in intracellular membrane-less organelles and extracellular tissue [30]. It is possible that the most disordered proteins of the aqueous humor participate in LLPS, providing insights into its molecular behavior and may be an attractive target for biomaterials and drug discovery [31].

To the best of our knowledge, this report is the first study to characterize the aqueous humor proteins in the context of intrinsic disorder. Our findings are crucial for better understanding of the aqueous humor. The better we understand the molecular behavior of the aqueous humor, the more likely we will be able to intervene when pathology ensues. There are several different diseases that are associated with the aqueous humor. Novus Biologics has an interactive database that identifies diseases linked to aqueous humor (see https://www.novusbio.com/diseases/aqueous-humor-disorders). These diseases include intraocular pressure disorder, glaucoma, corneal diseases, cataract, uveitis, open-angle

glaucoma, endophthalmitis, ocular hypertension, and retina disease. If we can identify intrinsically disordered proteins that show aberrant behavior in the aqueous humor in different pathologies, then we may be able to target these disorders in novel ways. These diseases can have a rapid or insidious courses and can cause patients a lot of distress. Many times, an ophthalmologist can intervene medically or surgically, but some of these diseases can remain refractory to either type of treatment modality. Therefore, it is important for the ophthalmology research community to consider intrinsic disorder in the aqueous humor proteome as a potential reason for pathology, which may be able to be leverage for treatment in the future.

As with all studies, this work has limitations. We were able to show that intrinsic disorder is abundant in the aqueous humor. Our functional characterization of these proteins is limited as our analysis was purely computational. A large-scale effort of the multi-omics wet labs will likely be needed to translate these findings to patient care. For now, we believe that we helped to provide a basic framework that can be used to better understand the molecular interactions that occur in the proteome of the aqueous humor.

5. Conclusions

Our computational analysis reported in this study demonstrated that intrinsic disorder is abundant in the aqueous humor. We showed that highly disordered proteins are consistently predicted as highly disordered as measured by various predictors. It is important to consider these intrinsic disorder-based molecular features, when thinking about any diseases that involve the aqueous humor, as misregulation of these protein entities can promote pathogenesis of various maladies. While there are currently not intrinsically disorder protein targets in the aqueous humor, these proteins should be considered as future targets as this may lead to novel therapeutics. More work will be needed to determine the significance of these findings as it relates to the patient care.

Support

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2022.101202.

References

[1] K.R. Murthy, et al., Proteomics of human aqueous humor, Omics 19 (5) (2015) 283–293.
[2] G.W. Daughdrill, et al., Natively disordered proteins, in: Protein Folding Handbook, 2008, pp. 275–357.
[3] V.N. Uversky, A decade and a half of protein intrinsic disorder: biology still waits for physics, Protein Sci. : Publ. Protein Soc. 22 (6) (2013) 693–724.
[4] V.N. Uversky, Unusual biophysics of intrinsically disordered proteins, Biochim. Biophys. Acta Protein Proteonomics 1834 (5) (2013) 932–951.
[5] U. Consortium, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (D1) (2019) D506–D515.
[6] V. Vacic, et al., Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinf. 8 (1) (2007) 1–7.
[7] K. Peng, et al., Optimizing long intrinsic disorder predictors with protein evolutionary information, J. Bioinf. Comput. Biol. 3 (1) (2005) 35–60.
[8] B. Xue, et al., CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions, FEBS Lett. 583 (9) (2009) 1469–1474.
[9] F. Huang, et al., Subclassifying disordered proteins by the CH-CDF plot method, in: Biocomputing 2012, World Scientific, 2012, pp. 128–139.
[10] P. Romero, et al., Sequence complexity of disordered protein, Proteins: Struct. Funct. Bioinf. 42 (1) (2001) 38–48.
[11] B. Xue, et al., PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta Protein Proteonomics 1804 (4) (2010) 996–1010.
[12] K. Rajagopalan, et al., A majority of the cancer/testis antigens are intrinsically disordered proteins, J. Cell. Biochem. 112 (11) (2011) 3256–3267.
[13] M.E. Oates, et al., D(2)P(2): database of disordered protein predictions, Nucleic Acids Res. 41 (2013) D508–D516 (Database issue).
[14] Z. Dosztányi, et al., IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (16) (2005) 3433–3434.
[15] P. Romero, et al., Sequence complexity of disordered protein, Proteins 42 (1) (2001) 38–48.
[16] T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, Nucleic Acids Res. 35 (2007) W460–W464 (Web Server issue).
[17] Z. Obradovic, et al., Exploiting heterogeneous sequence properties improves prediction of protein disorder, Proteins: Struct. Funct. Bioinf. 61 (S7) (2005) 176–182.
[18] K. Peng, et al., Length-dependent prediction of protein intrinsic disorder, BMC Bioinf. 7 (2006) 208.
[19] I. Walsh, et al., ESpritz: accurate and fast prediction of protein disorder, Bioinformatics 28 (4) (2012) 503–509.
[20] A. Andreeva, et al., SCOP database in 2004: refinements integrate structure and sequence family data, Nucleic Acids Res. 32 (2004) D226–D229 (Database issue).
[21] A.G. Murzin, et al., SCOP: a structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247 (4) (1995) 536–540.
[22] D.A. de Lima Morais, et al., SUPERFAMILY 1.75 including a domain-centric gene ontology method, Nucleic Acids Res. 39 (2011) D427–D434 (Database issue).
[23] B. Meszaros, I. Simon, Z. Dosztanyi, Prediction of protein binding regions in disordered proteins, PLoS Comput. Biol. 5 (5) (2009) e1000376.
[24] P.V. Hornbeck, et al., PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, Nucleic Acids Res. 40 (2012) D261–D270 (Database issue).
[25] M. Varadi, et al., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, Nucleic Acids Res. (2021).
[26] J. Jumper, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[27] D. Szklarczyk, et al., STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (D1) (2019) D607–D613.
[28] M. Necci, et al., Critical assessment of protein intrinsic disorder prediction, Nat. Methods 18 (5) (2021) 472–481.
[29] M. Sickmeier, et al., DisProt: the database of disordered proteins, Nucleic Acids Res. 35 (2007) D786–D793 (Database issue).
[30] B. Gabryelczyk, et al., Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides, Nat. Commun. 10 (1) (2019) 5465.
[31] G.L. Dignon, et al., Temperature-controlled liquid–liquid phase separation of disordered proteins, ACS Cent. Sci. 5 (5) (2019) 821–830.
[32] A. Bairoch, et al., The universal protein resource (UniProt), Nucleic Acids Res. 33 (2005) D154–D159 (Database issue).
[33] D. Szklarczyk, et al., The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored, Nucleic Acids Res. 39 (2011) D561–D568 (Database issue).
[34] K. Peng, et al., Optimizing long intrinsic disorder predictors with protein evolutionary information, J. Bioinf. Comput. Biol. 3 (1) (2005) 35–60.
[35] B. Xue, et al., PONDR-FIT: a meta-predictor of intrinsically disordered amino acids, Biochim. Biophys. Acta 1804 (4) (2010) 996–1010.
[36] Z. Dosztanyi, et al., The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, J. Mol. Biol. 347 (4) (2005) 827–839.