

Research Project

Intrinsic Disorder Analysis of Cortactin and Its Role in Invadopodia Formation

Pavan Vishnu Sai Bestha (3804-3186)

Abhinav Reddy Pannala (7031-4901)

1. Introduction:

For several decades, the structural biology community focused on the single-sequence-structure function theory as its base of operation. This theory suggests that a protein's amino acid sequence encodes for a specific native state structure, which in turn fulfils a designated function begins. Over the last 20 years, a variety of findings have disproved this idea to the point the fact that it is no longer valid. Numerous novel protein types, including intrinsic disease proteins and moonlighting proteins, have been found. Furthermore, we now know that various mescales of biomolecular dynamics in addition to structure, supervise biomolecular function, contrary to what had previously been believed to be the case. Consider that metamorphic proteins, that could fold into a wide range of distinctive structural topologies, have merely developed.

Cortactin modulates actin polymerization by interacting with actin-binding protein. This is an important regulator of the cytoskeletal dynamic. Dysregulation in cortactin was strongly associated with cancer metastasis. Therefore, it is a crucial therapeutic target. This research investigates structural dynamics and intrinsically disordered regions (IDRs) of cortactin by using advanced tools in bioinformatics such as: RIDAO D2P2, FUZDROP ALPHAFOLD & STRING. These analyses reveal cortactin IDRs to be functionally adaptable, essential for cellular functions. They also present risks associated with misfolding and accumulation implicated in metastasis related pathways. Key findings included the identification of IDRs which are susceptible to liquid/liquid phase separation as well structural insights into cortactin networks. These findings highlight the dual nature inherent disorder plays in both enabling normal cell function and pathological outcome. This study advances the understanding of cortactin's function in actin-cytoskeleton remodelling. The future work of this research establishes an effective framework for targeting IDRs in cancer related proteins.

2. Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) and regions (IDPRs) occupy secondary or tertiary structure levels. Conformations toward which the system is rapidly moving as highly dynamic ensembles of rapidly approaching conformations (or highly dynamic/set of). One building within another, as opposed to enclosing a three-dimensional building. They might have been discovered in two distinct forms: rounded smooth (as in the molten blobs) and wire-like (like the coils). Pre-molten globules also form part of these structures. Many ordered proteins and domains provide a complementary service that is completed by inhabitants of areas inside IDPs and inhabitants of areas inside IDRs. Promiscuous binders, referred to as inclusion-dependent polymers and inclusion-dependent receptors, are involved in disordering and coordinating a wide spectrum of cellular operations. Molecular recognition, entropic chain assemblies, molecular assembly, protein modifications, and RNA and protein pseudoknots chaperones can be stated as some of the general functional classes into which the IDPs/IDRs can be subdivided. It is probable that IDPs represent an issue compared with other groups of population. Men IDPs have a way of being in existence that is similar to IDRs and multiple human disorders, and they

have promiscuous interactions on a reasonably frequent basis. It was observed that 79% of cancer-related proteins tended to have IDRs of length 30 residues or beyond. All misfolded proteins or regions of proteins must therefore be analyzed. Regarding the disorder areas, there are several predictors available to look at. In this section, it is important for us to express about two predictors that we utilized in our study to identify the disorder region: Rapid Intrinsic Disorder Analysis (RIDAO) and the D2P2.

3. Tools we used:

- **STRING:**

STRING (Search tool for the retrieval of interacting genes/proteins) is an informational bioinformatics source that provides data about protein-protein interaction. It integrates and annotations known and predicted interactions between proteins from various sources including experimental data and computational predictions. STRING is used by researchers to identify complexes of proteins, explore functional associations, and predict functions using interaction networks.

- **RIDAO:**

RIDAO is a uniquely effective web tool for the prediction of disorders allowing researchers to put into practice protein intrinsic disorder analysis for comparative genomics, proteomics, and genome-scale structural bioinformatics as a much easier task. At a per-residue rate that is expected to be contained in the input data set. RIDAO builds one single model which, all things considered, looks very much like all of the five famous predictors of disorder.

- **D2P2:**

The D2P2 database, that generated an average disorder score by integrating several state-of-art disorder prediction tools, improves also the precision as well as the reliability of the predictions. Concurrently, the D2P2 data is to share with the users how the structured and disorder are interrelated by giving the users access to the SCOP domain on the premise of concurrent disorders. The D2P2 data also reflects the presence of the converse of the folded structure and may also give some useful data for in enhancing the fold forecasting methods.

- **FUZDROP:**

FuzDrop is a bioinformatics and computer programmed tool that uses a different approach to calculate the probability of liquid-liquid phase separation for specific proteins or other molecules. FuzDrop also provides binding scores, aggregation of hotspots and binding scores that match to specific protein sequences.

- **ALPHAFOLD:**

AlphaFold is an Alphabet Inc. subsidiary's deep-learning tool. It has made precise predictions of 3D structures of proteins from the explicit amino acid sequencing sequence. It is vital because the 3D structural form of a particular protein greatly determines its functionality. Predicting protein structures accurately can help us to improve our understanding of biology, as well as speed up drug development and discovery processes.

4. Methodology:

To perform our analysis, we first downloaded the CTTN's (human cortactin gene) canonical FASTA sequence from the UniProt database. This sequence formed the basis from which

all subsequent analyses were derived. To find out the functional association of cortactin, we used the tool STRING, and found 500 proteins interacting with cortactin. These interactions showed its strong association with some of important proteins involved in cell signaling, actin cytoskeleton and cell architecture such as WIPF1, Synaptopodin, Caldesmon and others. For the interacting proteins, we used STRING to export the FASTA format of multiple sequence alignment format.

Out of the identified 400-500 PPIs that physically interact with cortactin, we have chosen 10 most relevant for the analysis of intrinsic disorder, which are implicated within cortactin's functional networks. The FASTA sequences of these proteins were also retrieved from UniProt database of protein sequence and function. To analyze the structural dynamics and intrinsic disorder regions (IDRs), we utilized advanced bioinformatics tools: RIDAO, D2P2, FuzDrop and AlphaFold. RIDAO and D2P2 were used for disorder region and functional feature, FuzDrop was applied for LLPS potential, and AlphaFold was applied to predict protein structures.

5. Contributions:

Pavan concentrated on the following UniProt protein's IDs and run through RIDAO and D2P2; Figures illustrates the intrinsically disordered regions (IDRs) of the selected proteins. RIDAO presented graphical displays of disorder probability distribution over protein sequences; the overlapped disorder probability was derived from the stacked predictions of VLXT, VSL2B, VL3, IUP_S, IUP_L, P_fit, and MDP. Protein flexibility was identified from these predictions pointing to large zones of disorder. D2P2 even finer tuned the study by providing consensus disordered regions, functional domains, post-translational modifications (PTMs) and molecular recognition features (MoRFs). This gave a better feeling about structural propensity and biological roles of the disordered segments.

Further, Abhinav focused on identifying the behaviour of LLPS and an accurate estimation of protein 3D structures. When performing the experiment with FuzDrop, he determined areas of phase separation and aggregation hotspots which are critical for signaling and compartmentalization. Additionally, for each of the identified proteins, he used AlphaFold for a 3-D protein structure prediction. AlphaFold's ability to combine the predicted disorder regions into actual spatial conformations offered outcomes into the relation between structural flexibility and protein function.

By these contributions, we were able to identify disorder regions and assess the propensity for phase separation, as well as propose structural models in the context of cortactin and cellular dynamics and metastatic cancer.

6. Proteins we used:

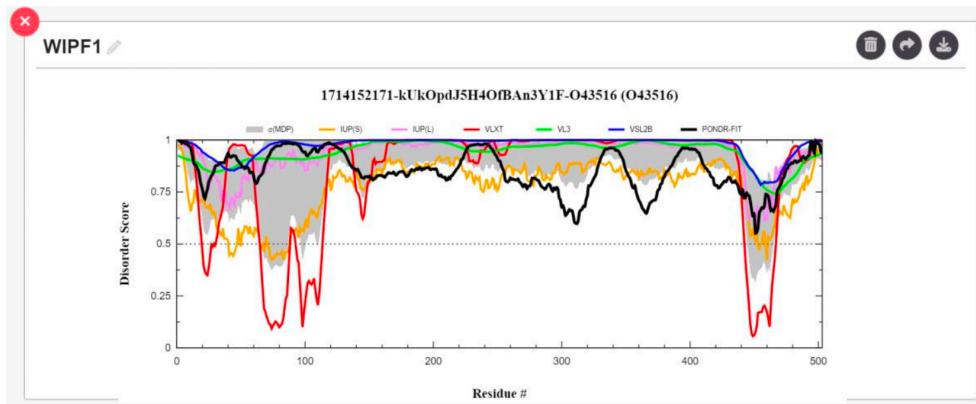
- WIPFI (WASP- Interacting Family Member -1) (O43516): A protein which interacts with the actin cytoskeleton and contributes to cellular processes such as motility.
- Synaptopodin (Q8N3V7): It is an actin-associated proteins involved in the regulation cytoskeletal dynamic in renal podocytes.

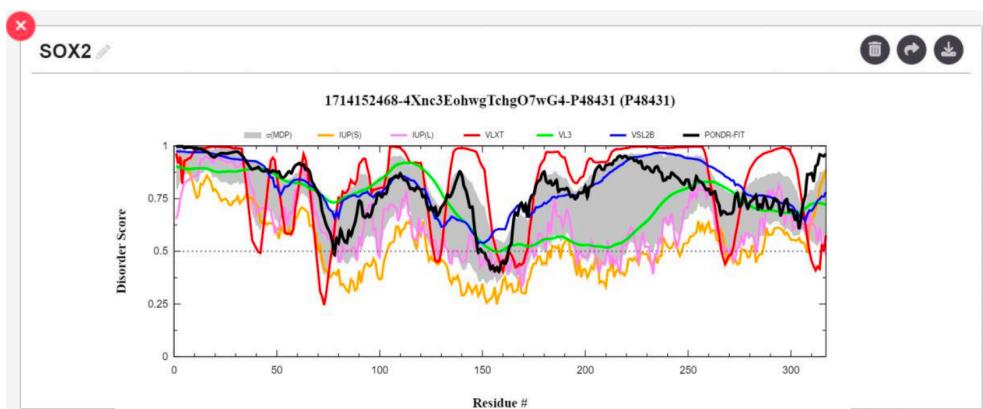
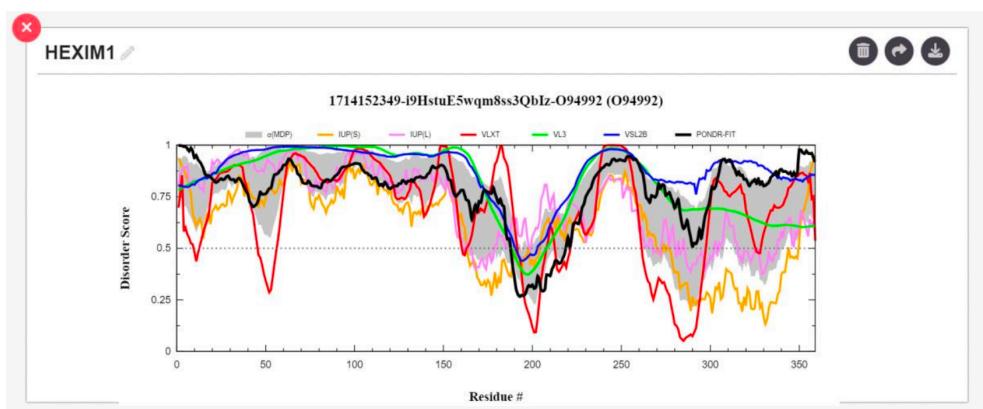
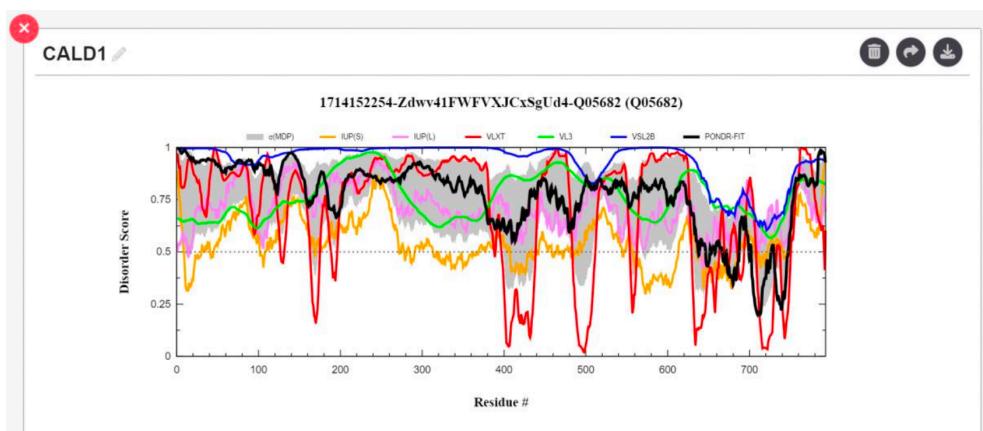
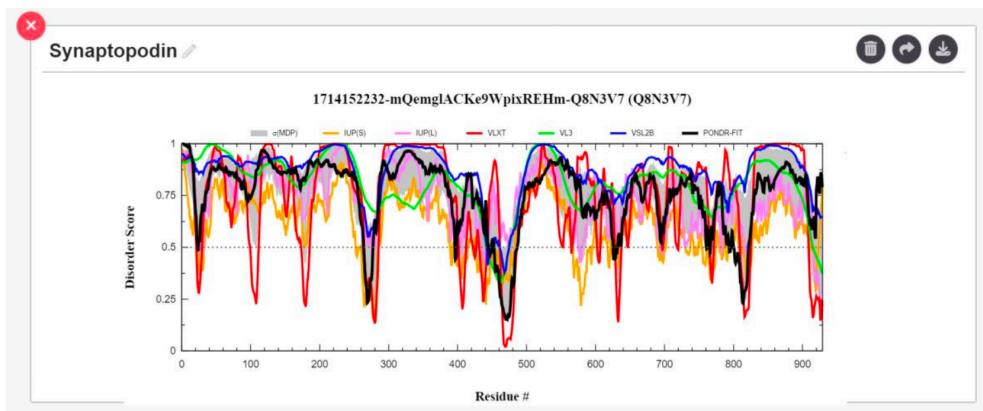
- Caldesmon (Q05682): A regulatory protein which links actin and Myosin filaments. It also controls the actomyosin-actin interaction within muscle and non-muscle cells.
- GAP-43/Neuromodulin (Q5U058): Protein associated with nerve development, which plays a part in axonal growth and synaptic plasticity.
- HEXIM 1(O94992): A protein that regulates transcription of RNA Polymerase II, modulating transcriptional length and innate immunity responses.
- SOX2 (P48431): A transcriptional element essential for stem cell pluripotency.
- Cyclin dependent kinase (P50750): It is one of the family of proteins that regulates cell cycle, especially in transitions, such as G1-S phase. This protein play critical roles in the proliferation of cells.
- Angiomotin-Like 2 (AMOLT2) (Q9Y2J4): This protein is involved in angiogenesis as well cell polarity. It interacts with signaling pathways, such as Wnt/b Catenin.
- HCLS1 (O00165): A protein which regulates the actin cytoskeleton, and contributes to cell movement and immune responses.
- Actin-Binding proteins (Q9UPY6): Actin filaments are interacted with by these proteins, which are necessary for cytoskeletal structure and cell shape maintenance.

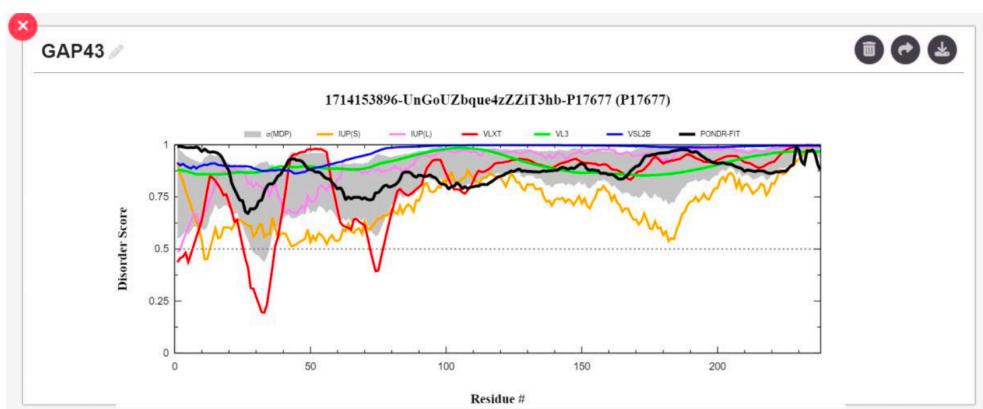
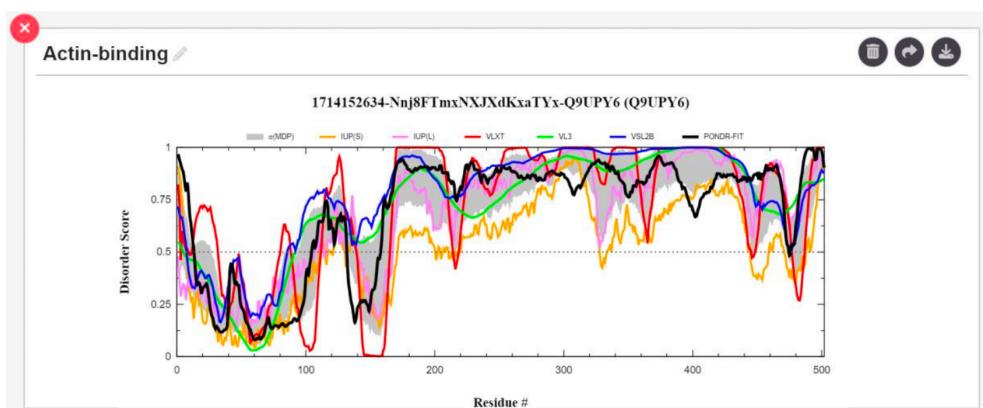
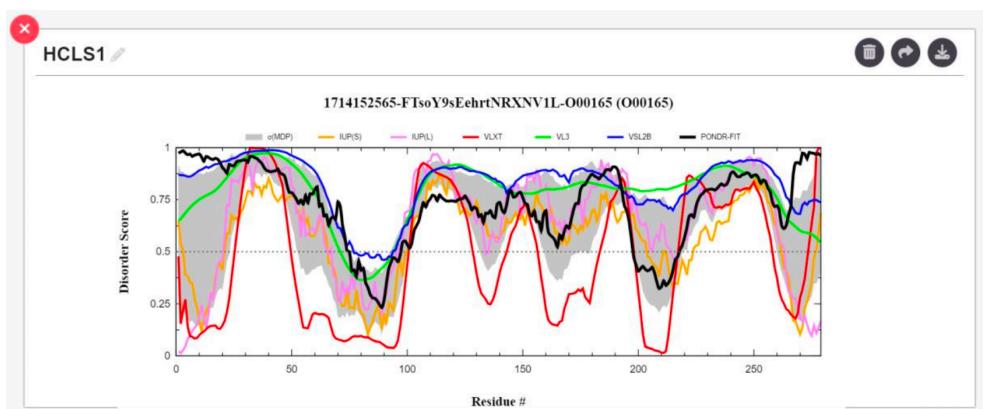
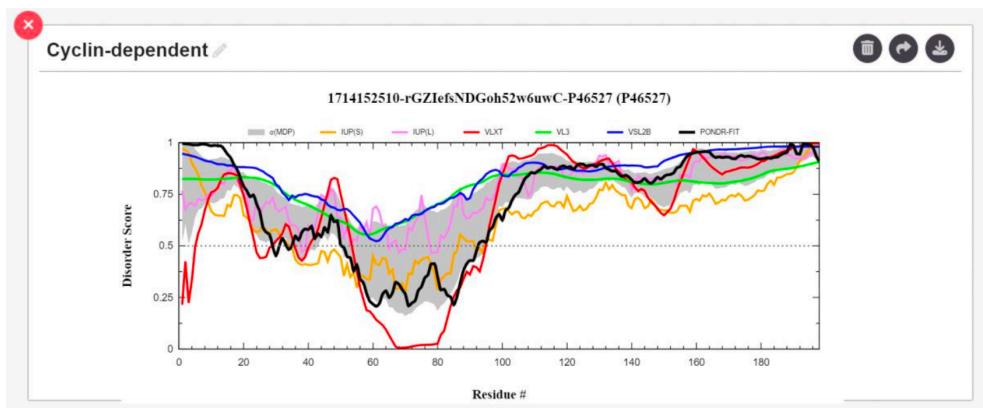
7. Results:

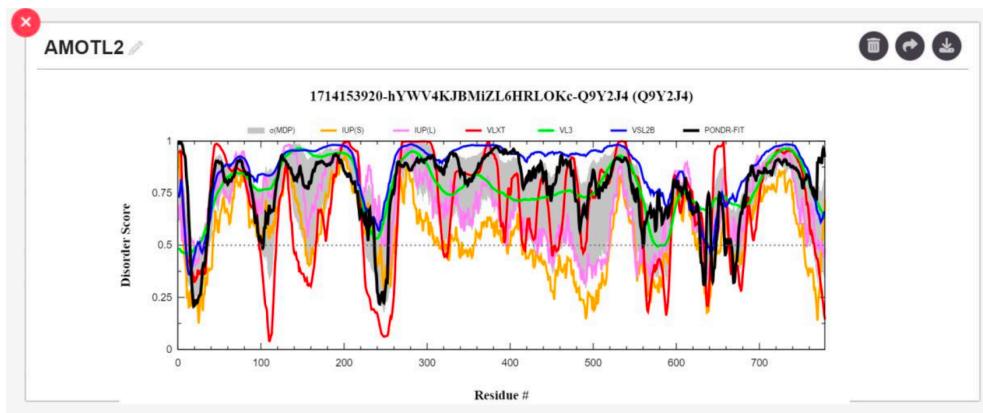
- **RIDAO:**

We used the RIDAO tool to analyze the disordered regions of our selected proteins. By inputting each protein's FASTA sequence, RIDAO generated a consolidated graph showing disorder regions and their scores, along with the percentage of disordered residues. It combined predictions from multiple advanced algorithms, including VLXT, VSL2B, and IUP_S. The following outputs provided valuable insights into protein disorder dynamics and highlighted regions critical for cellular processes and disease pathways like metastasis. These results were downloaded and systematically analyzed to support our study.





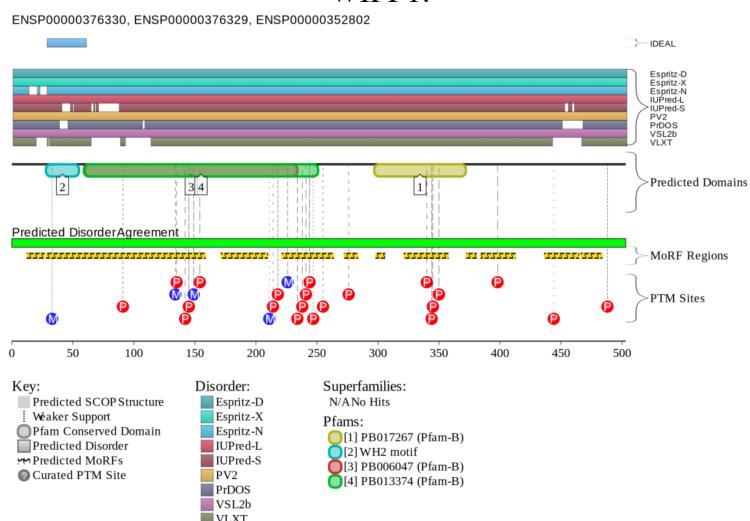




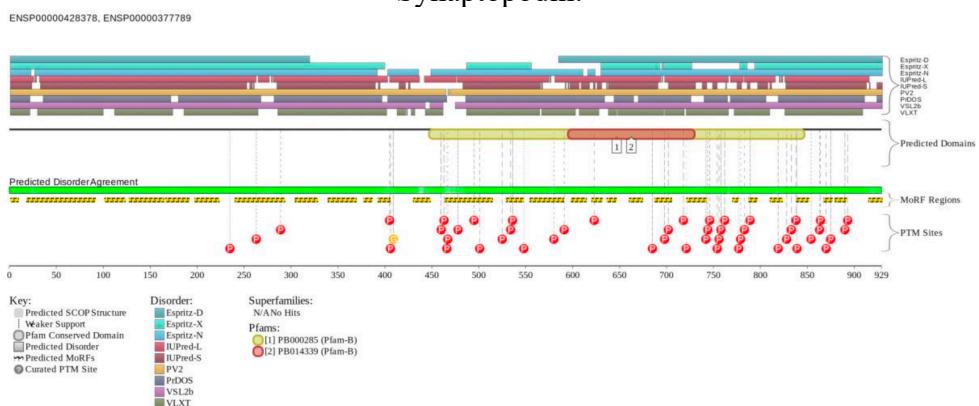
- **D2P2 predictor tool:**

We examined our chosen protein's intrinsic disorder regions (IDRs) using the D2P2 tool. Each protein's FASTA sequence was entered into the program, and consensus disorder scores produced by sophisticated algorithms such as VLXT, VSL2B, PrDOS, PV2, IUPred, and Espritz were acquired. Post Translational Modifications (PTMs), Molecular Recognition Features (MoRFs), and the locations of functional domains were among the other insights that the D2P2 tool offered. Protein interactions and dynamic behavior depend on these outputs, which also enabled us to identify functional domains that overlap with disordered areas and regions with a high susceptibility for disorder. We downloaded and methodically arranged the graphical results, which contained annotated areas for each protein and disorder profiles.

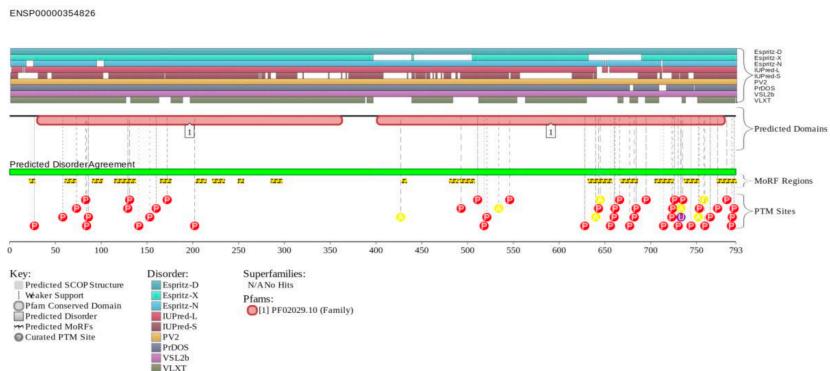
WIPF1:



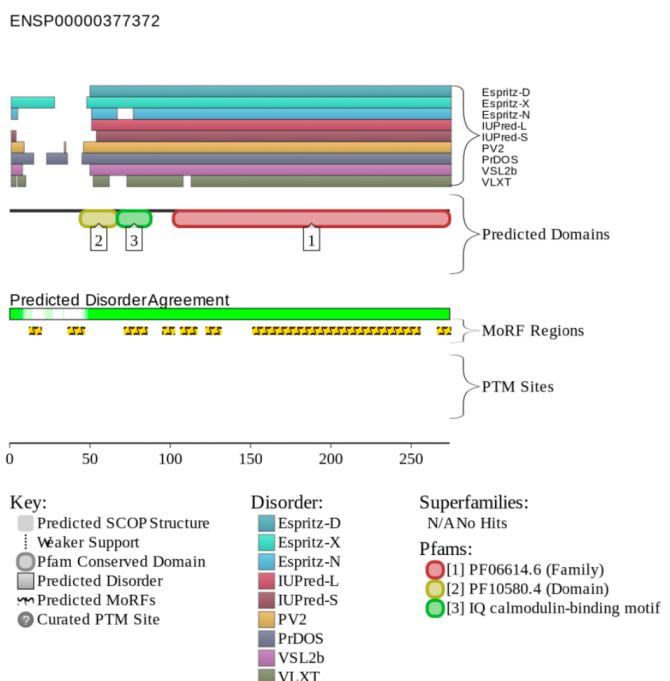
Synaptopodin:



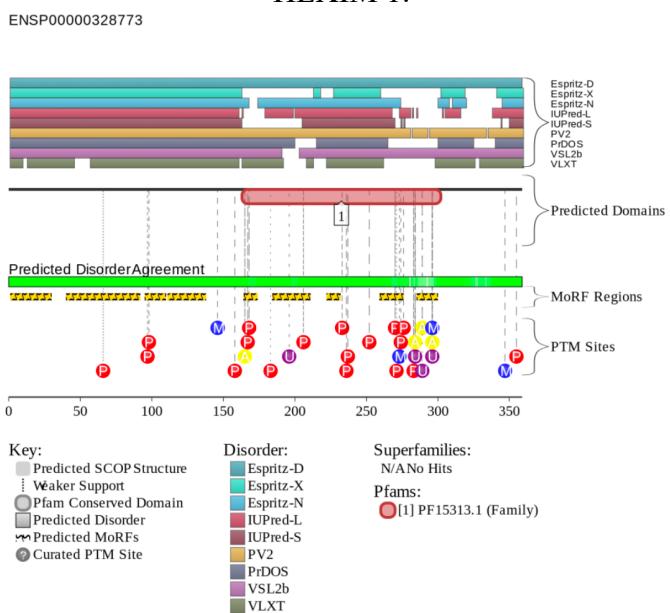
Caldesmon:



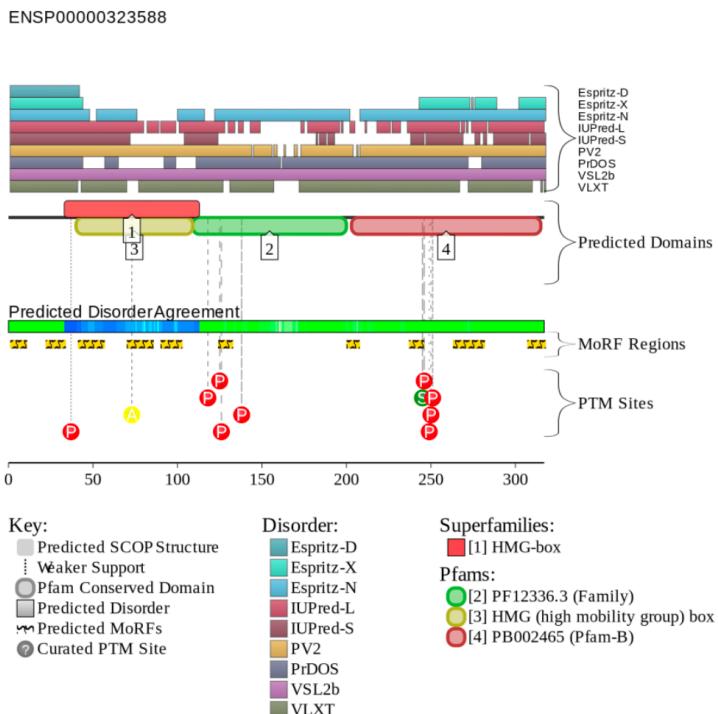
Neuromodulin:



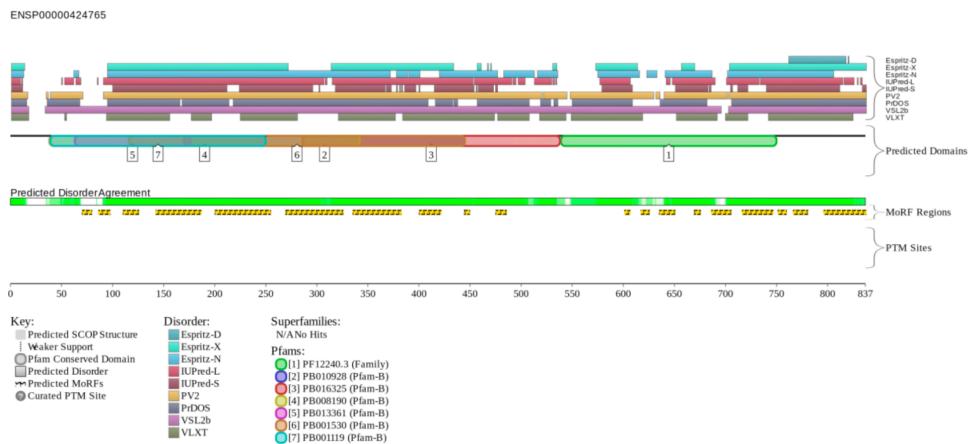
HEXIM 1:



SOX 2:



AMOTL 2:

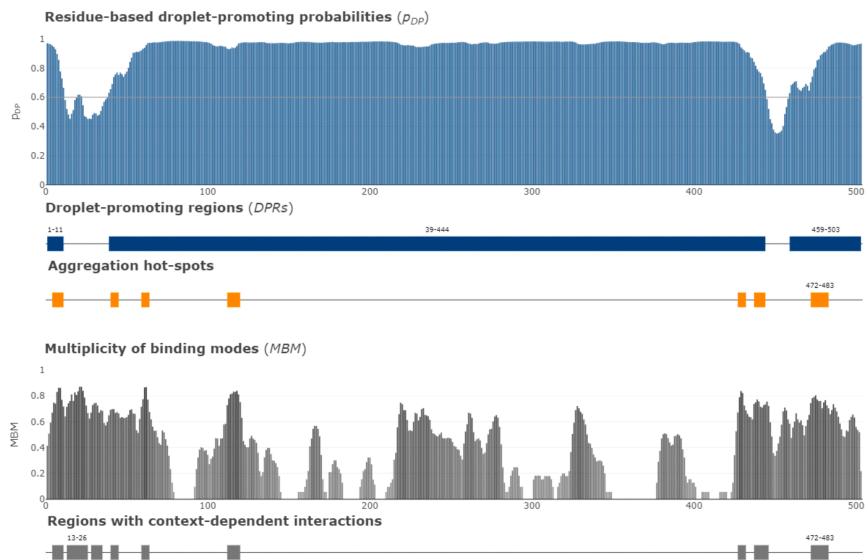


- **FuzDrop:**

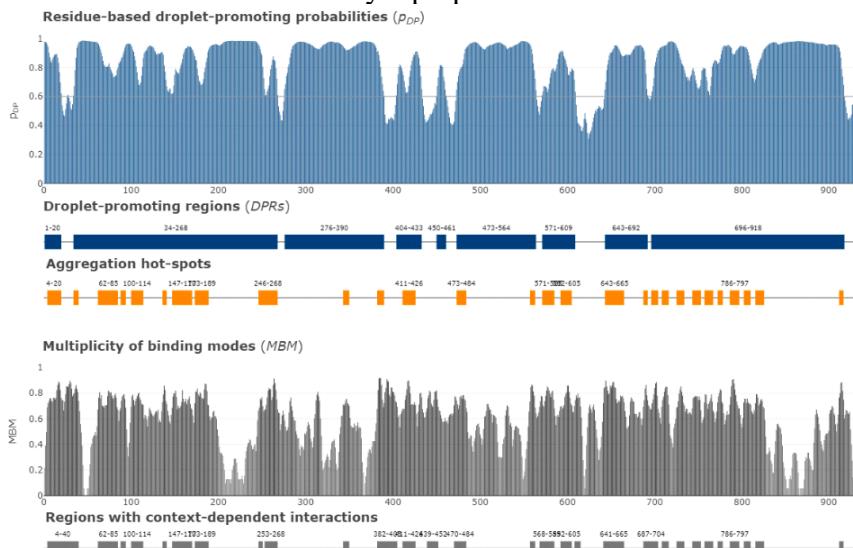
We employed the FuzDrop tool to analyze the liquid-liquid phase separation (LLPS) potential of our selected proteins. For each protein, we input the corresponding FASTA sequence into the tool, which calculated the probability of spontaneous phase separation and provided cellular context dependence scores. These predictions offered insights into how the intrinsically disordered regions (IDRs) of the proteins might facilitate dynamic interactions essential for cellular organization and function.

The outputs were downloaded and systematically organized, including probability scores, graphical depictions of phase separation potential, and annotated regions prone to aggregation or droplet formation. These results highlighted critical IDRs involved in phase separation, further linking structural disorder to functional adaptability and pathological implications, such as metastasis.

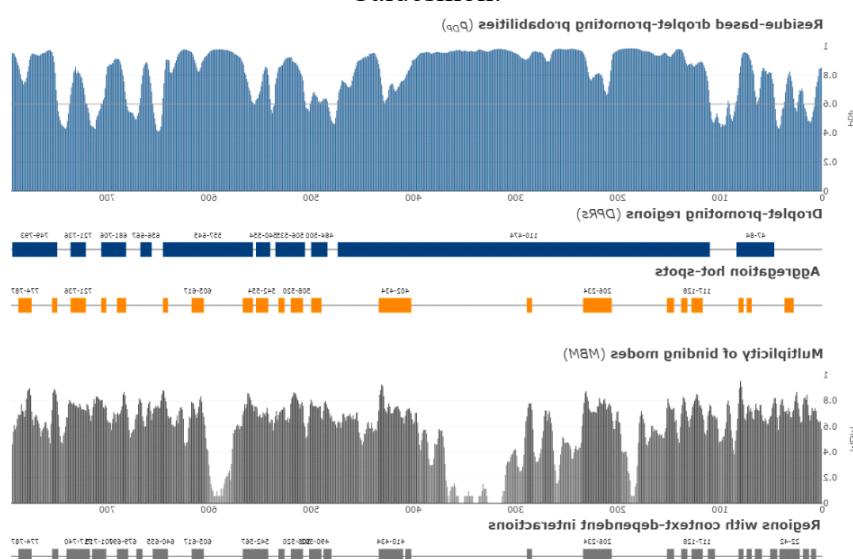
WIPF 1:



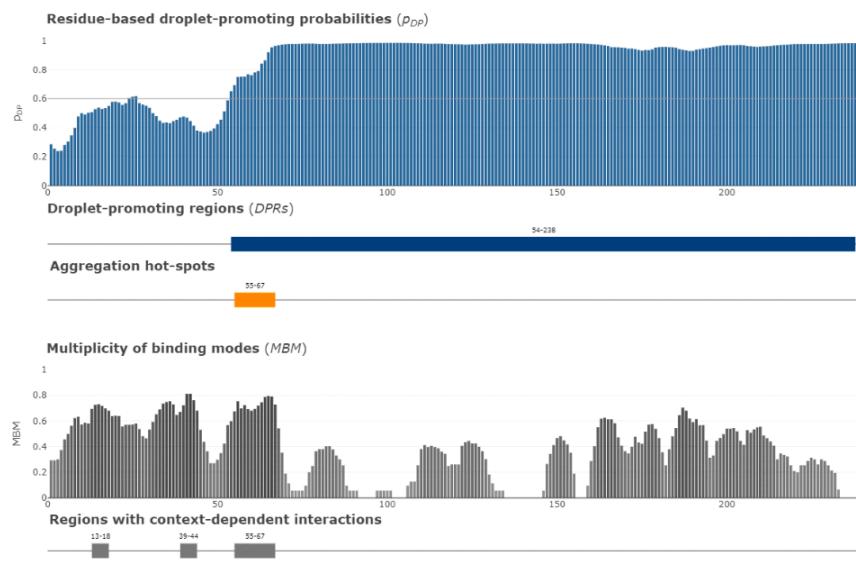
Synaptopodin:



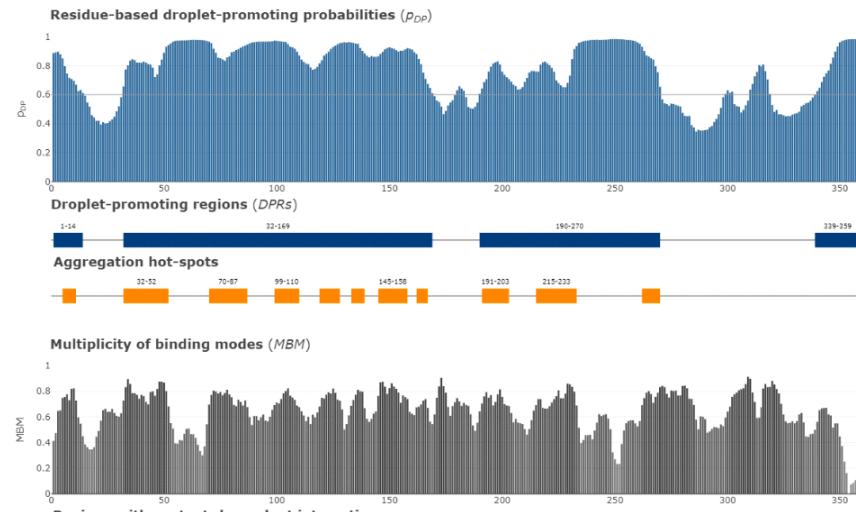
Caldesmon:



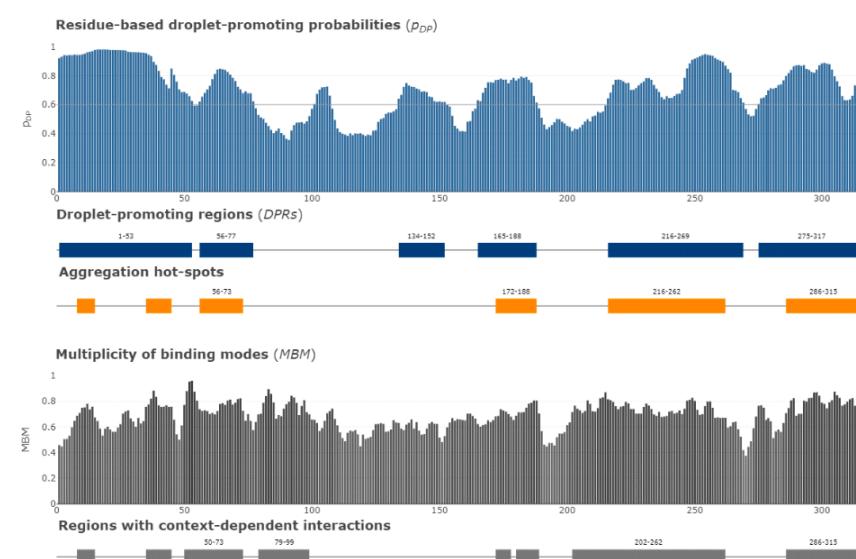
Neuromodulin:



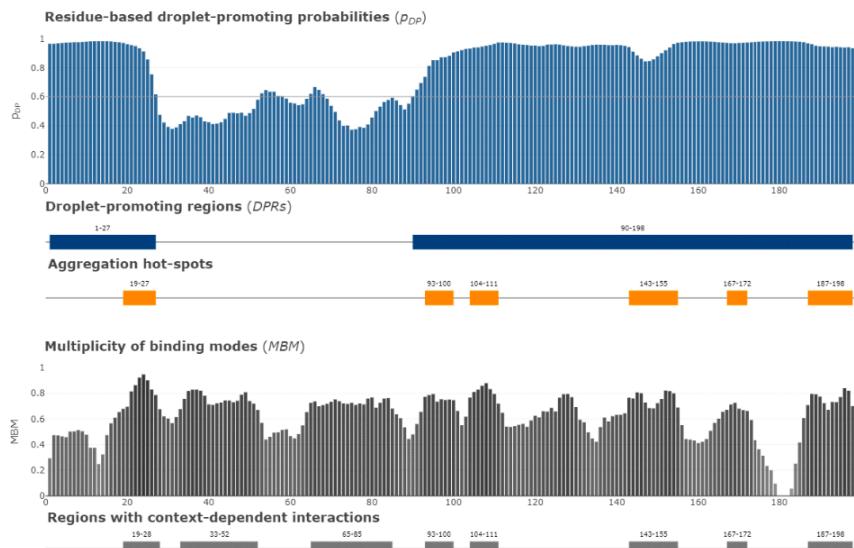
HEXIM 1:



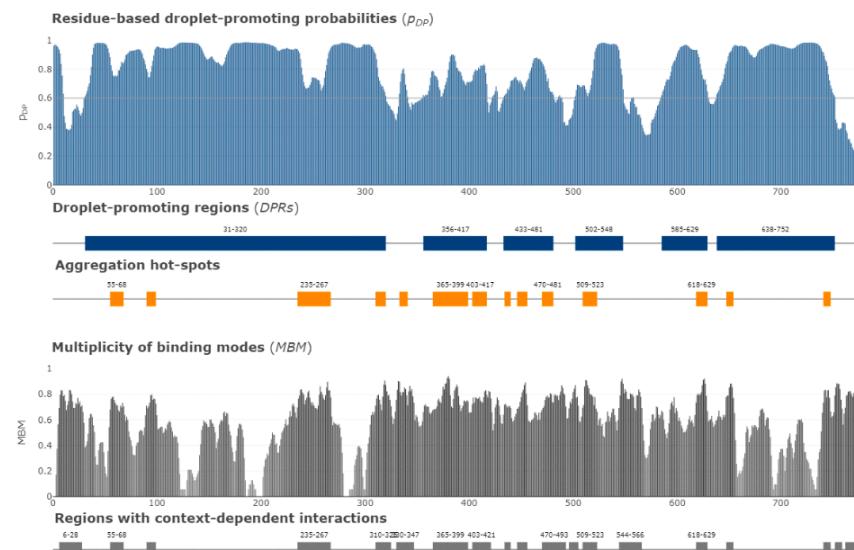
SOX 2:



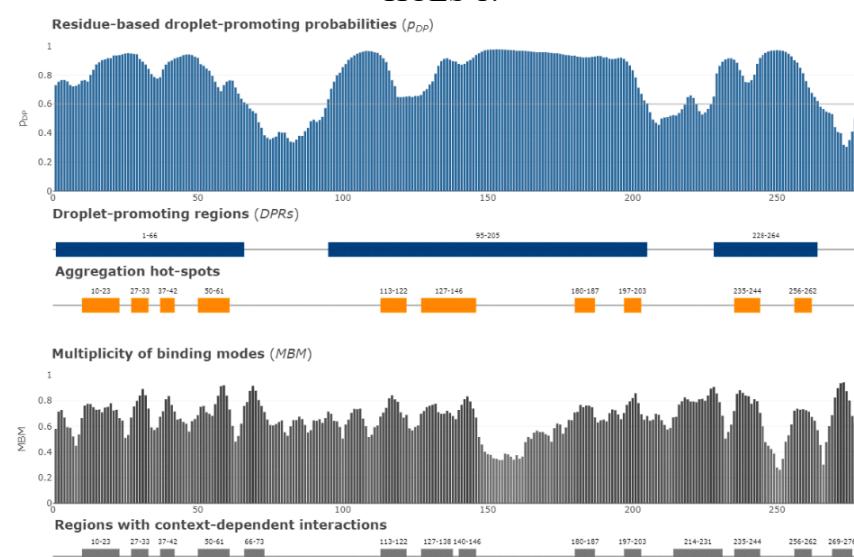
Cyclin DK:

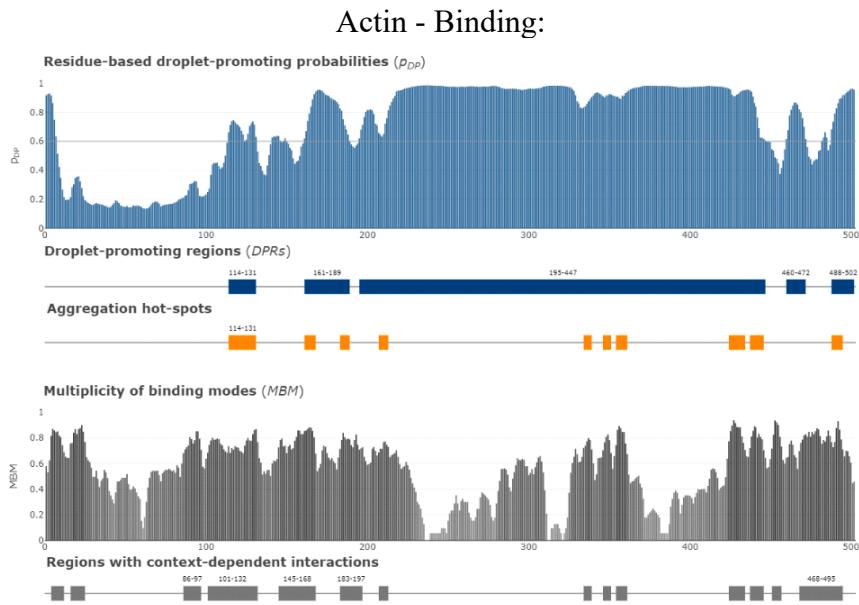


AMOTL 2:



HCLS 1:

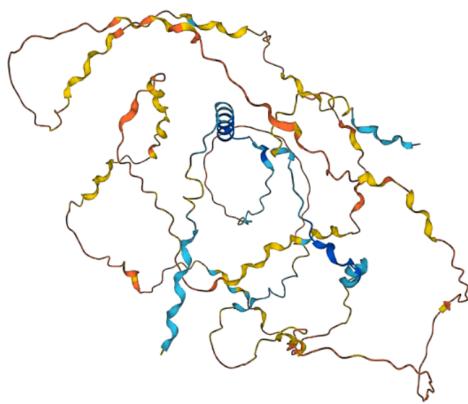




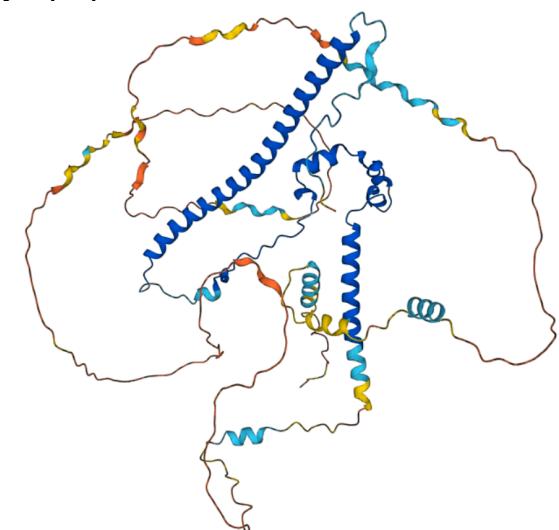
- **AlphaFold:**

To predict the 3D structures of our selected proteins we utilized AlphaFold available in Google Colab. Lastly, each protein sequence was inputted for prediction after the environment setup where dependencies and the AlphaFold package were downloaded. Essentially, the tool produced MSA and near-ideal 3D structural models of each protein, essential information regarding the spatial organization of the proteins and the presence of functional domains. The highly sophisticated neural network structure of AlphaFold correctly identified complex atomic configurations based on data sets of proteins. The predictions outlined regions that can adapt and interface with other proteins and parts of a cell, which aided understanding of the proteins part in cellular events and disease routes such as metastasis. This tool has taken a long time to install, run all the required code for the sequences we took.

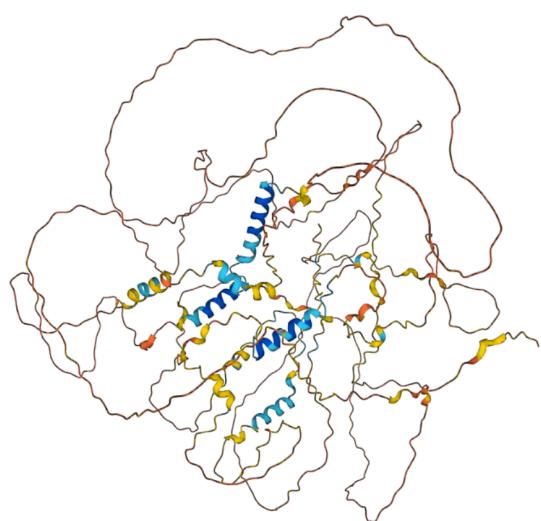
WIPF 1:



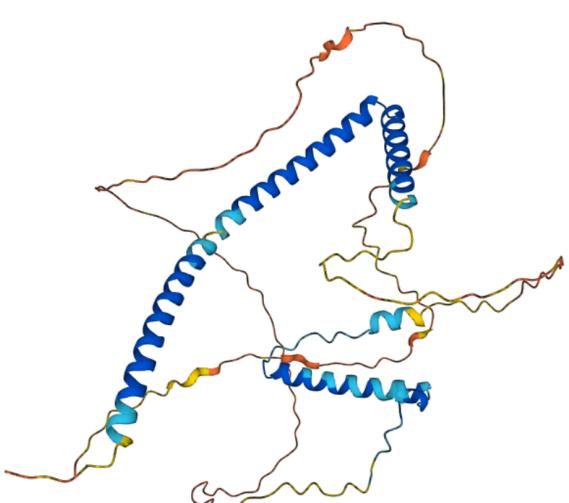
Synaptopodin:



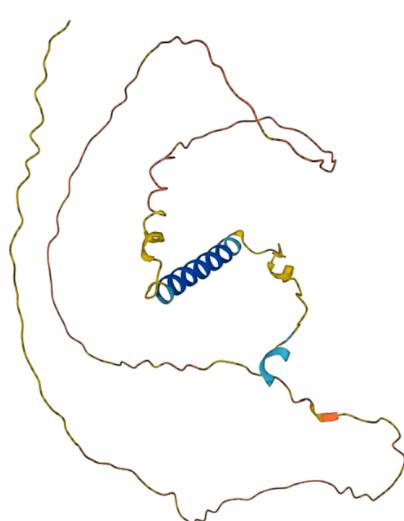
Caldesmon:



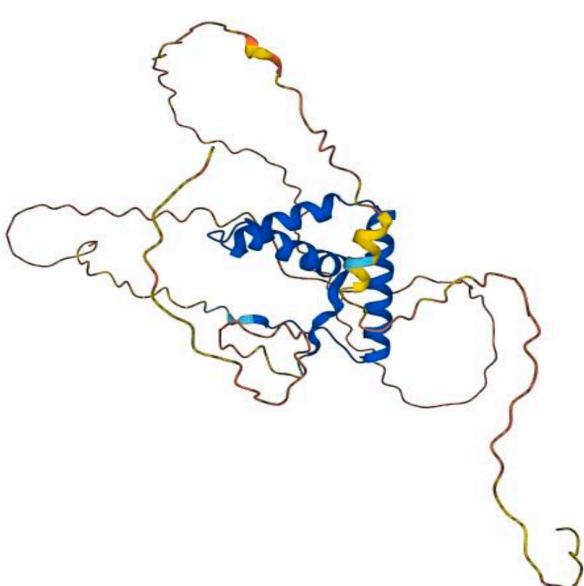
SOX 2:



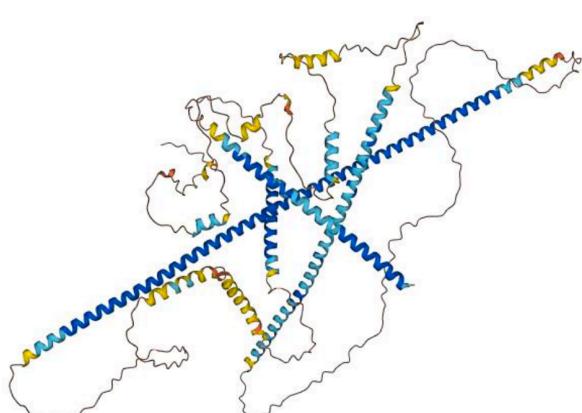
Neuromodulin:



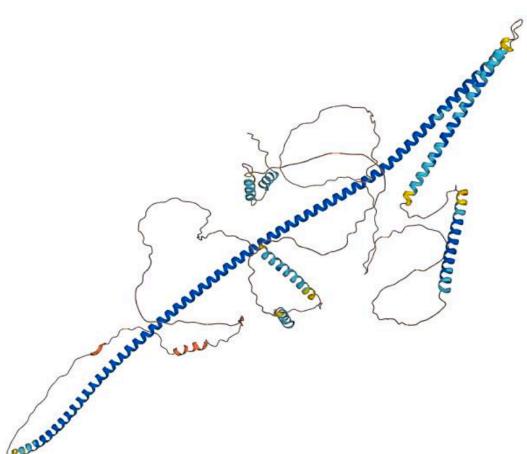
Cyclin DK:



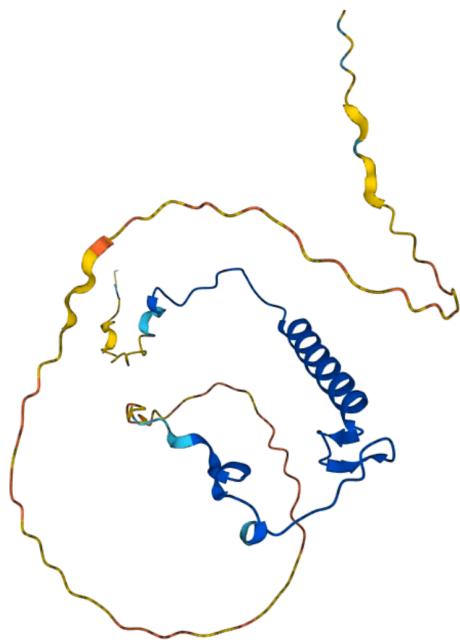
HEXIM 1:



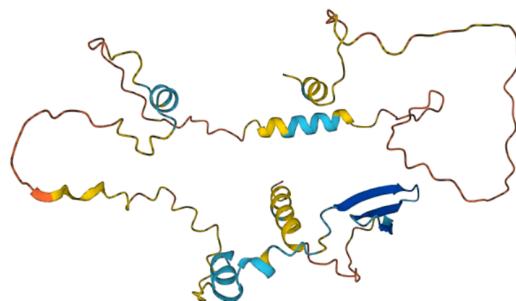
AMOTL 2:



HCLS 1:



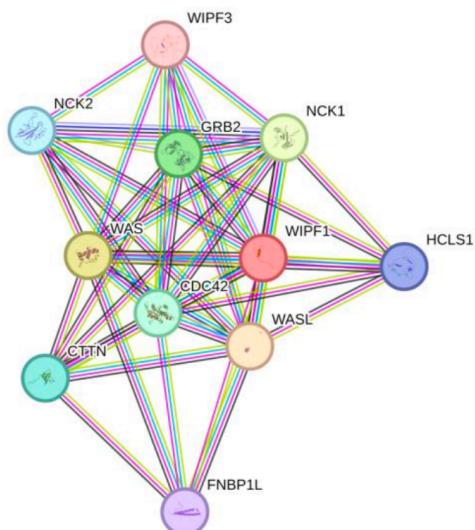
Actin – Binding:



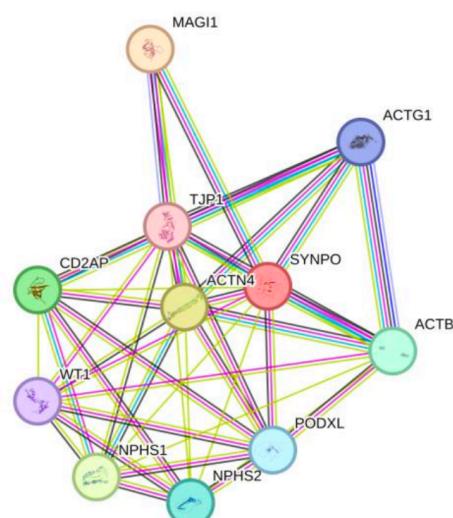
- **STRING:**

By directly entering the FASTA sequences of the proteins we had chosen, we were able to utilize the STRING program to examine their interaction networks. After submission, STRING analyzed the sequences to find networks of protein-protein interactions by combining information from curated databases, computational predictions, and experimental research. Our target proteins' direct and indirect relationships with their functional partners were emphasized by the tool's comprehensive interaction maps. The results provide important new information on biological pathways, protein activities, and how cellular processes are organized. To find important hubs and routes pertinent to our research, these interaction networks were downloaded and examined.

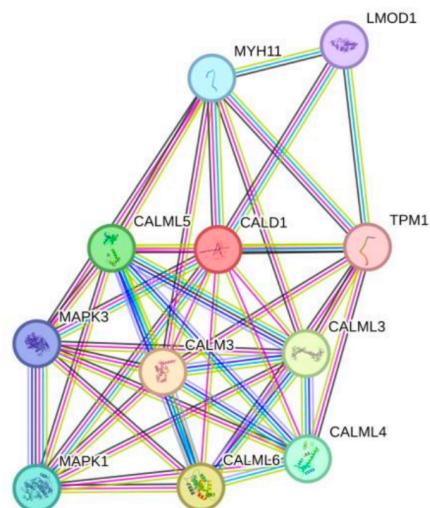
WIPF 1:



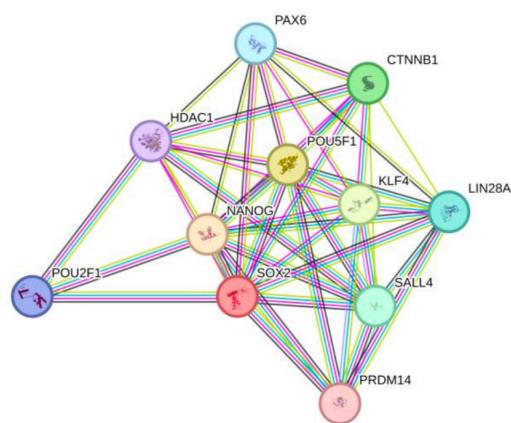
Synaptopodin:



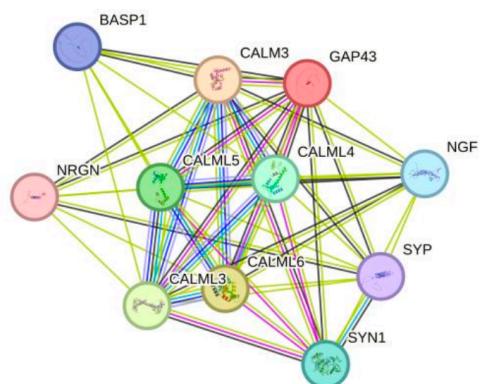
Caldesmon:



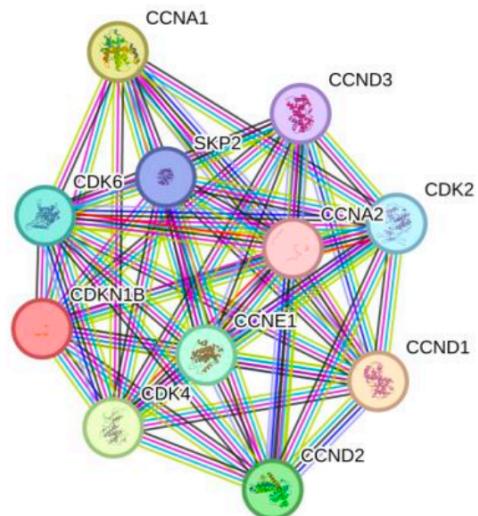
SOX 2:



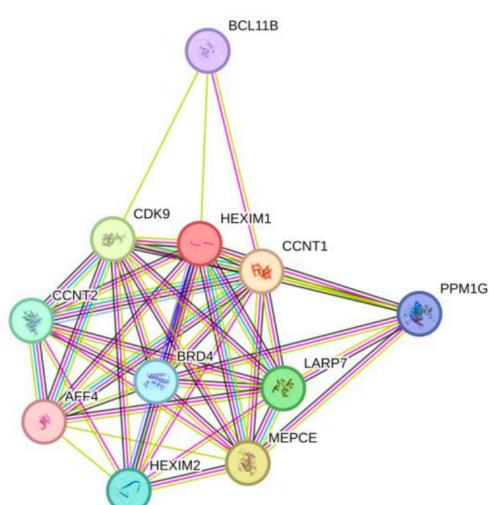
Neuromodulin:



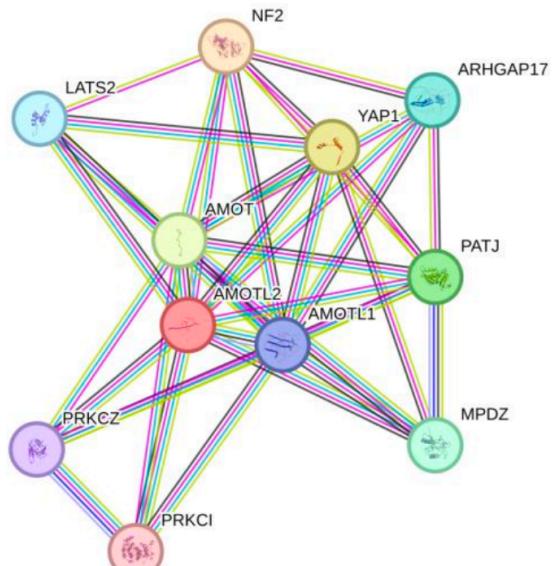
Cyclin DK:



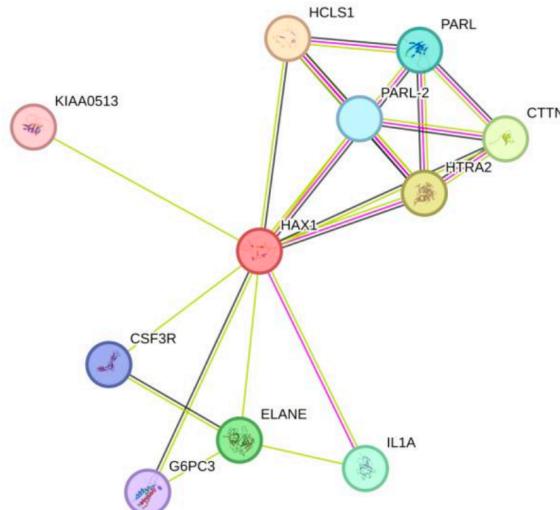
HEXIM 1:



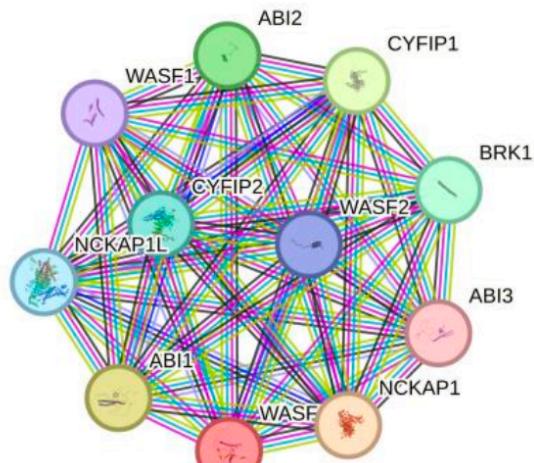
AMOTL 2:



HCLS 1:



Actin - Binding:



8. Discussion:

We have used 5 tools to predict the results. RIDAO predictor results display disorder profiles of each derived protein as graphs. This graph also displays the percentage of disorder across all proteins using various predictors including VLXT VSL2B VL3, IUP_S IUP_L MDP, and PFIT. This tool also shows the disorder regions on all proteins using graphs. D2P2 data shows functional disorder profiles in two proteins using different predictors. This includes IUPred PV2, PrDOS VLXT and VSL2B. D2P2's outcomes then show pictures of functional disorders for seven different proteins. These are predicted by VLXT and VSL2B. They also use PV2, IUPred and Espritz. The position of the functional domains as well as MoRFs and PTMS is also shown. Additionally, results are shown to show where MoRFs and functional domains such as PTMs are situated. We both implemented AlphaFold and STRING tools which have been run simultaneously for each protein in two pc's, where AlphaFold took a long time to show the results. FuzDrop has been used to calculate the probability for proteins that are involved in liquid-liquid phases separation.

For all proteins analysed;

- STRING: Predicted a network of 500 protein-protein interactions for Cortactin and key interacting proteins as these 10 proteins we used in this research.
- RIDAO: Predicted that large portions of protein sequences have scores above 0.5, particularly from residues ~100 to ~450. About ~60% to ~80% disorder in Cortactin and interactors.
- D2P2: Predicted about ~50% to ~70% consensus disorder score for disordered regions across protein sequences. Regions which have MoRF's and PTMs were found indicating functional activities in flexible, binding-prone areas.
- FuzDrop: Predicted LLP's potential in disordered regions, particularly in central parts of the protein sequences and high droplet-promoting scores suggesting a role in phase separation.
- AlphaFold: Predicted 3-D structures of all proteins with low confidence scores of regions which are identified as disordered, particularly about ~100 to ~400 residues.

9. Conclusion:

The intrinsically disordered regions (IDRs) of cortactin and its interactors were thoroughly examined in this study, which demonstrated their critical involvement in cellular processes and the pathways leading to cancer metastasis. We detected disorder-prone areas, post-translational modifications (PTMs), and liquid-liquid phase separation (LLPS) potential using sophisticated bioinformatics techniques, including RIDAO, D2P2, FuzDrop, AlphaFold, and STRING. The results demonstrate how IDRs' structural flexibility allows for dynamic interactions that are necessary for adhesion, migration, and actin polymerization. While STRING highlighted the important interaction networks of cortactin, AlphaFold validated the flexibility of these areas. Our knowledge of cortactin's function in cytoskeletal remodeling and disease progression is improved by these findings.

10. Future Work:

Future studies should focus on developing targeted therapies to disrupt cortactin-related pathways in cancer. Experimental cortactin of predicted IDRs using NMR spectroscopy and molecular dynamics simulations will confirm their functional roles. Deeper understanding of cellular organization may be possible with additional research into liquid-liquid phase separation using sophisticated imaging techniques. Furthermore, cortactin may be examined as a therapeutic target and biomarker for cancer metastasis, bridging the gap between clinical applications and computer forecasts.

11. References:

- [1] Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C. J., Muller, J., Doerks, T., Julien, P., Röth, A., Simonovic, M., Bork, P., & Von Mering, C. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, 37(Database), D412–D416. <https://doi.org/10.1093/nar/gkn760>
- [2] Jin, J., Lu, P., Xu, Y., Li, Z., Yu, S., Li, J., Wang, H., Chua, N., & Cao, P. (2020). PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Research*, 49(D1), D1489– D1495. <https://doi.org/10.1093/nar/gkaa910>
- [3] Madhurima, K., Nandi, B., & Sekhar, A. (2021). Metamorphic proteins: the Janus proteins of structural biology. *Open Biology*, 11(4). <https://doi.org/10.1098/rsob.210012>
- [4] Inohara, N., & Núñez, G. (2001). The NOD: a signaling module that regulates apoptosis and host defence against pathogens. *Oncogene*, 20(44), 6473-6481. <https://doi.org/10.1038/sj.onc.1204787>
- [5] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Röth, A., Bork, P., Jensen, L. J., & Von Mering, C. (2016). The STRING database in 2017: qualitycontrolled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, 45(D1), D362–D368. <https://doi.org/10.1093/nar/gkw937>
- [6] Dayhoff, G. W., & Uversky, V. N. (2022). Rapid prediction and analysis of protein intrinsic disorder. *Protein Science*, 31(12). <https://doi.org/10.1002/pro.4496>
- [7] W, A., Senior, Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S.,

Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>

[8] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A. J., Cowie, A., RomeraParedes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

[9] Farah, C. A., & Leclerc, N. (2008). HMWMAP2: New perspectives on a pathway to dendritic identity. *Cell Motility and the Cytoskeleton*, 65(7), 515–527. <https://doi.org/10.1002/cm.20284>

[10] Hatos, A., Toso, S. C. E., Vendruscolo, M., & Fuxreiter, M. (2022). FuzDrop on AlphaFold: visualizing the sequence-dependent propensity of liquid–liquid phase separation and aggregation of proteins. *Nucleic Acids Research*, 50(W1), W337–W344. <https://doi.org/10.1093/nar/gkac386>

[11] Liu, Z., & Huang, Y. (2014). Advantages of proteins being disordered. *ProteinScience*, 23(5), 539–550. <https://doi.org/10.1002/pro.2443>

[12] Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztányi, Z., Uversky, V. N., Obradovic, Z., Kurgan, L., Dunker, A. K., & Gough, J. (2012). D2P2: database of disordered protein predictions. *Nucleic Acids Research*, 41(D1), D508–D516. <https://doi.org/10.1093/nar/gks1226>