Dissecting biophysics and biology of intrinsically disordered proteins

Authors: Priya R. Banerjee*[1], Alex S. Holehouse[2], Richard Kriwacki[3], Paul Robustelli[4], Hao Jiang[5], Alexander Sobolevsky[6], Jennifer Hurley[7] & Joshua T. Mendell[8]

Affiliations: [1]Department of Physics, University at Buffalo, NY 14260; [2]Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis; [3]Structural Biology Department, St. Jude Children's Research Hospital; [4]Department of Chemistry, Dartmouth College; [5]Department of Biochemistry and Molecular Genetics, University of Virginia; [6] Department of Biochemistry and Molecular Biophysics, Columbia University; [7]Department of Biological Sciences, Rensselaer Polytechnic Institute; [8]UT Southwestern Medical Center and Howard Hughes Medical Institute

*__Corresponding Author__: Priya R. Banerjee. Email: prbanerj@buffalo.edu

Dissecting biophysics and biology of intrinsically disordered proteins

Abstract

Intrinsically disordered regions within human proteins play critical roles in cellular information processing, including signaling, transcription, stress response, DNA repair, genome organization, and RNA processing. Here, we summarize current challenges in the field and propose cutting-edge approaches to address them in normal physiology and disease, with a focus on cancer.

Keywords: Phase Separation, Cancer, Condensates, Single molecule

Dissecting biophysics and biology of intrinsically disordered proteins

Abundant evidence supports that intrinsically disordered regions (IDRs) in proteins play critical roles in normal cellular functions and many disease processes, including cancer [1]. Despite tremendous progress in our understanding of how IDRs regulate a myriad of biological processes, such as gene regulation and intracellular signaling, there are many open questions and challenges. Further, IDRs are now widely recognized as drivers and modulators of biomolecular condensates, which are membrane-less subcellular hubs that play important roles in the dynamic compartmentalization of biochemical processes in living cells [1]. Mutations in IDRs have been shown to result in aberrant behaviors of condensates, resulting in dysregulation of signaling events in the cytoplasm as well as activation of oncogenic transcriptional programs in the nucleus [2, 3]. Therefore, there is a pressing need to understand the mechanistic principles underlying the biological functions of IDRs and leverage this knowledge to target their aberrant behaviors in disease processes.

Recently, leading researchers in the intrinsically disordered protein (IDP) field gathered in a symposium organized by the National Cancer Institute at the National Institutes of Health[1]. They discussed the current state of the field, key open questions, and emerging new approaches critical to furthering our understanding of how biological functions of IDRs and IDPs are encoded in their primary sequences. Below we summarize these open questions and discuss integrative approaches to address them (Fig. 1).

Understanding how sequence-encoded grammar governs IDR functions

IDRs differ from folded domains in that they lack a fixed three-dimensional structure, and instead populate a collection of conformations referred to as conformational ensemble. As such, IDRs do not conform to the classical sequence-structure paradigm, challenging existing experimental, computational, and conceptual tools to understand how they function.

In some IDRs, the primary sequence exhibits low complexity, meaning the sequence composition is biased in amino acid content, with enrichment of a few amino acids. For example, prion-like low-complexity domains (PLCDs) display an overrepresentation of aromatic (Y/F) and polar amino acids (G/S/Q/N) and depletion of charged residues. In contrast, other IDRs show sequence complexities equivalent to folded domains, suggesting that complex chemical patterns can be readily encoded. More broadly, IDRs from across the human proteome exhibit diverse sequence features as quantified based on, for example, amino acid composition, residue patterning, and the presence of linear motifs. A major question is how we can decipher IDR sequence-specific grammars, which presumably encode their functions in cell signaling, gene regulation, and other biological processes.

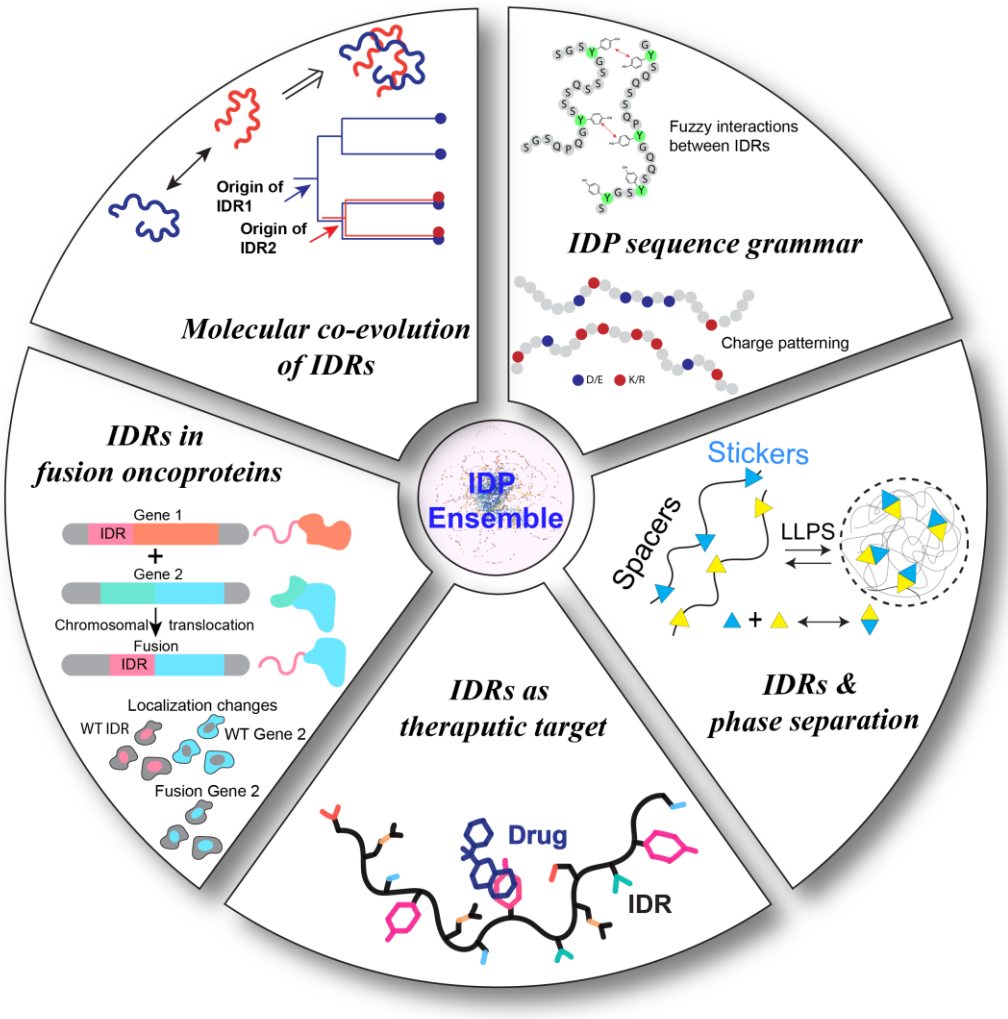

Figure 1: Emerging paradigm and current challenges in sequence-ensemble-function relationships of intrinisically disordered proteins.

A growing body of work suggests that IDR function can be viewed through a sequence-ensemble-function paradigm [4] (Fig. 1). That is, IDR function depends on a combination of sequence features and the emergent biophysical consequences those features have on an IDR's ensemble. While much work has focussed on mapping relationships between sequence and ensemble, extending this mapping from sequence and ensemble to function is incomplete. Based on extant results, IDR functions can be driven by different properties, including (i) conformational plasticity, (ii) multivalent sequence-encoded molecular interactions and/or phase separation, (iii) post-translational modifications that dynamically re-wire IDR-mediated interactions, and (iv) short linear motifs that enable sequence-specific molecular recognition. However, in general, we lack a holistic understanding of the mapping between IDR primary sequence and function. One route to understanding sequence-ensemble-function relationships for specific proteins is to integrate

systematic high-resolution *in vitro* biophysical measurements of IDRs and IDPs stretching from single molecules to multi-component assemblies with measurements of IDR functions in live cells and/or animal models.

Molecular co-evolution of IDP sequences

When assessed by sequence alignment-based metrics, IDRs often appear less well-conserved than their folded counterparts, suggesting a lower level of evolutionary conservation. However, conservation may operate in different ways for IDRs than folded domains [5]. Specifically, conservation can operate at the level of interactions with evolutionarily conserved binding partners and ensemble properties, as opposed to alignable sequences. This realization embraces the sequence-ensemble-function paradigm described above and offers a conceptual framework through which disease-causing mutations can be interpreted. When regions in IDRs are well-conserved in terms of primary sequence, this may be a signature of co-evolutionary coupling between a region that folds upon binding and the partner it binds. Our emerging understanding of the compensatory and labile evolutionary dynamics of IDRs offers a tremendous opportunity to better understand sequence-ensemble-function relationships for IDRs as well as their interaction networks. Given this potential, developing new computational tools to interpret these relationships in an evolutionary context is an active area of investigation.

IDRs as drivers and modulators of intracellular phase transitions

Dynamic, membrane-free compartmentalization of subcellular processes via macromolecular phase separation is ubiquitous in living systems. An intracellular phase transition (e.g., phase separation) is a process in which a protein, and/or other biomolecules, assemble into a condensed state [6]. While IDRs are (in general) neither necessary nor sufficient for phase transitions or condensate formation, many of the proteins studied *in vitro* and *in vivo* that undergo phase separation have been found to possess IDRs that are, in specific cases, necessary and sufficient. With this in mind, there continues to be a great interest in understanding how and why IDRs influence the formation, maintenance, and regulation of condensates.

Given that phase separation is a concentration-dependent phenomenon, a major challenge is to characterize condensates under endogenous cellular protein levels. This can be addressed by fixed cell imaging and live cell knock-in based-approaches probing condensates at endogeneous protein levels. Further, quantitative concentration titrations *in vitro* and in living cells can help characterizing how IDR sequence features enocde driving forces for phase separation. However, a key challenge is relating *in vitro* and cell culture model conditions to the native cell-specific microenvironment of the protein of interest.

An equally important task is determining where condensates (driven by phase separation or otherwise) are functionally relevant vs. an epiphenomenon that unavoidably accompanies multivalent molecular interactions, which themselves drive function. Separation of function experiments offer one potential route; e.g., demonstrating that phase separation via several chemically or physically distinct routes phenocopy one another in terms of a functional readout.

Engineering stimuli-responsive pathways to induce or disrupt phase separation without changing protein concentrations in live cells can further help dissect the role of phase separation in biological functions [7]. Importantly, the physics of phase transitions [6, 8] (e.g., viscoelasticity, distinct intra-condensate chemical microenvironments, liquid-liquid interfaces, *etc.*) encodes emergent mesoscale properties that are not achievable via alternative modes of biomolecular assembly (e.g., isodesmic polymerization). A common theme of these challenges is the need for integrative, multifaceted approaches that combine molecular and cellular engineering, statistical physics, and novel experimental approaches, including opto- and chemo-genetics, mass photometry, and live-cell super-resolution nanoscopy to address these important questions.

IDRs in fusion oncoproteins

IDRs of signaling proteins have been extensively implicated in the formation and regulation of phase-separated hubs in living cells. In the context of fusion oncoproteins that arise via chromosomal translocations, phase separation has been implicated to be a major driver in aberrant chromatin remodeling and transcriptional reprogramming in certain cases [3, 9]. A major open question is what features of an IDR contribute to phase separation and modulate the complex folding landscape of chromatin? Further, since the database of fusion genes linked to cancer is rapidly growing[2,3], a key challenge is to create an effective platform that will allow integratation of large-scale information from live-cell experiments with machine learning-based approaches. Performing large-scale proteome-level characterization to identify phase separation driven by IDRs, and subsequently, implementing machine learning to link IDR sequence features with their functions will be a valuable approach, as has recently been demonstrated for 166 fusion oncoproteins [10].

IDRs as therapeutic targets

Owing to the complex roles of IDRs in signaling events, transcriptional regulation, and oncogenesis, IDRs have emerged as prime therapeutic targets [11]. A small molecule-based approach to target an IDR's conformational ensemble and molecular interactome presents challenges, but recent studies have shown some promise. Further, anti-cancer drugs have been observed to selectively enrich within certain nuclear condensates and may perturb aberrant phase separation of transcription factors. Therefore, dissecting the interactions between small molecule therapeutics and cancer-associated biomolecular condensates enriched in regulatory proteins containing IDRs offers potential new directions for anti-cancer drug discovery.

Future Outlook

Although our discussion here narrowly focused on five key aspects of the biology and biophysics of IDPs (Fig. 1), we acknowledge that the functional roles played by IDRs in human proteins are much broader and more diverse (Box 1). The discussions at the NCI symposium highlighted that no single research group could deploy all the techniques and expertise needed to effectively address these questions. We envision that a multi-group collaborative effort spanning chemical screening, biophysics, structural biology, cell biology, genetics, advanced

microscopy, and data science methods is needed to make progress on these key challenges. Such a multidisciplinary approach will enhance our broad understanding of relationships between the dynamic conformational features of IDPs and their biochemical and biological functions. The knowledge gained will provide a foundation for understanding how the biophysical properties and functions of IDPs are altered in human diseases, including diverse cancers.

Acknowledgements:
Research in the Banerjee Lab is supported by grant awards from the National Institute of General Medical Sciences (R35 GM138186), National Institute of Aging (R03 AG070510), National Center for Advancing Translational Sciences (R03 TR003387), St. Jude Children's Research Hospital, and the National Science Foundation (MCB 2239500). The authors thank Anowarul Amin and Stefan Maas of the NCI Division of Cancer Biology for their organizing and moderating of the workshop. The authors also appreciate the support of the Division of Cancer Biology at NCI in facilitating this meeting. A special thanks to all esteemed speakers (listed below) who enriched the workshop with their profound expertise and insights:

Priya R. Banerjee, SUNY at Buffalo
Sarah Bondos, Texas A&M University
Alex Holehouse, Washington University in St. Louis
Jennifer Hurley, Rensselaer Polytechnic Institute
Hao Jiang, University of Virginia
May Khanna, New York University
Richard Kriwacki, St. Jude Children's Research Hospital
Wenbo Li, UT Health Science Center Houston
Joshua Mendell, UT Southwestern Medical Center and Howard Hughes Medical Institute
Elizabeth Rhoades, University of Pennsylvania
Paul Robustelli, Dartmouth College
Alexander Sobolevsky, Columbia University
Robert Tycko, NIDDK, NIH
Jin Zhang, University of California San Diego
Yixian Zheng, Carnegie Institution for Science

Competing Interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Alberti S, Gladfelter A and Mittag T (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell 176:419-434. doi: 10.1016/j.cell.2018.12.035
2. Boija A, Klein IA and Young RA (2021) Biomolecular Condensates and Cancer. Cancer Cell 39:174-192. doi: 10.1016/j.ccell.2020.12.003
3. Shirnekhi HK, Chandra B and Kriwacki RW (2023) The Role of Phase-Separated Condensates in Fusion Oncoprotein–Driven Cancers. Annual Review of Cancer Biology 7:73-91. doi: 10.1146/annurev-cancerbio-061421-122050
4. Forman-Kay Julie D and Mittag T (2013) From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. Structure 21:1492-1499. doi: 10.1016/j.str.2013.08.001
5. Hultqvist G, Åberg E, Camilloni C, Sundell GN, Andersson E, Dogan J, Chi CN, Vendruscolo M and Jemth P (2017) Emergence and evolution of an interaction between intrinsically disordered proteins. eLife 6:e16059. doi: 10.7554/eLife.16059
6. Mittag T and Pappu RV (2022) A conceptual framework for understanding phase separation and addressing open questions and challenges. Molecular Cell 82:2201-2214. doi: 10.1016/j.molcel.2022.05.018
7. Shin Y, Berry J, Pannucci N, Haataja MP, Toettcher JE and Brangwynne CP (2017) Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. Cell 168:159-171 e14. doi: 10.1016/j.cell.2016.11.054
8. Li W, Hu J, Shi B, Palomba F, Digman MA, Gratton E and Jiang H (2020) Biophysical properties of AKAP95 protein condensates regulate splicing and tumorigenesis. Nature Cell Biology 22:960-972. doi: 10.1038/s41556-020-0550-8
9. Davis RB, Moosa MM and Banerjee PR (2022) Ectopic biomolecular phase transitions: fusion proteins in cancer pathologies. Trends Cell Biol 32:681-695. doi: 10.1016/j.tcb.2022.03.005
10. Tripathi S, Shirnekhi HK, Gorman SD, Chandra B, Baggett DW, Park C-G, Somjee R, Lang B, Hosseini SMH, Pioso BJ, Li Y, Iacobucci I, Gao Q, Edmonson MN, Rice SV, Zhou X, Bollinger J, Mitrea DM, White MR, McGrail DJ, Jarosz DF, Yi SS, Babu MM, Mullighan CG, Zhang J, Sahni N and Kriwacki RW (2023) Defining the condensate landscape of fusion oncoproteins. Nature Communications 14:6008. doi: 10.1038/s41467-023-41655-2
11. Zhu J, Salvatella X and Robustelli P (2022) Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. Nature Communications 13:6390. doi: 10.1038/s41467-022-34077-z
12. Dunlap JC (1999) Molecular Bases for Circadian Clocks. Cell 96:271-290. doi: 10.1016/S0092-8674(00)80566-8
13. Hurley JM, Loros JJ and Dunlap JC (2016) Circadian Oscillators: Around the Transcription–Translation Feedback Loop and on to Output. Trends in Biochemical Sciences 41:834-846. doi: 10.1016/j.tibs.2016.07.009
14. Greer AM, Huang Z, Oriakhi A, Lu Y, Lou J, Matthews KS and Bondos SE (2009) The Drosophila Transcription Factor Ultrabithorax Self-Assembles into Protein-Based Biomaterials with Multiple Morphologies. Biomacromolecules 10:829-837. doi: 10.1021/bm801315v
15. Alshareedah I, Moosa MM, Pham M, Potoyan DA and Banerjee PR (2021) Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides. Nat Commun 12:6620. doi: 10.1038/s41467-021-26733-7

Box 1: IDPs in circadian rhythms and biomaterials

Intrinsic protein disorder in circadian rhythms

Humans have circadian rhythms that allow us to organize our physiology and behavior such that appropriate activities occur at biologically advantageous times within the day-night cycle. The circadian rhythms are generated by a molecular oscillator, or clock, that comprises a transcription-translation negative feedback loop (TTFL)[12, 13]. The conserved intrinsic structural disorder has been suggested to be essential for the function of this molecular clock in post-transcriptional gene regulation. Future studies dissecting how these "Fuzzy" complexes form and are timed would enable the design of therapeutics to target these complexes as well as the development of chronotherapeutics to treat cancer.

Intrinsically disordered proteins as programmable biomaterials

From a material science perspective, multivalent IDPs offer new avenues for controlled self-assembly of soft biomaterials with programmable structure, mechanics, and desired function [14, 15]. IDP-based self-assembled soft functional matters, such as phase-separated droplets, hydrogels, and glassy materials that show controlled responses to environmental cues such as light, temperature, pH, and concentration, hold great potential as artificial extracellular matrix (ECM) material in tissue engineering, regenerative medicine, cell therapy, immunomodulation, and in the creation of artificial cells and synthetic organelles.

Dissecting biophysics and biology of intrinsically disordered proteins

Resources

1. https://www.cancer.gov/about-nci/organization/dcb/news/idp-workshop
2. https://ccsm.uth.edu/FusionGDB/
3. https://www.kobic.re.kr/chimerdb/