Towards structure prediction and design of disordered proteins

Kresten Lindorff-Larsen
Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science,
University of Copenhagen, Denmark

Intrinsically disordered proteins and regions (collectively IDRs) are pervasive across proteomes in all kingdoms of life, help shape biological functions, and are involved in numerous diseases.[1] IDRs populate a diverse set of transiently formed structures yet defy commonly held sequence-structure-function relationships[1,2]. Developments in protein structure prediction have led to the ability to predict the structures of folded proteins at the proteome scale and enabled large-scale studies of structure-function relationships. In contrast, knowledge of the conformational properties of IDRs is scarce, in part because the sequences of disordered proteins are poorly conserved and because only few have been characterized experimentally. In my talk I will describe how we can use molecular simulations with coarse-grained models to study the relationship between sequence, conformational properties, and functions of IDRs.

I will describe how we have used experimental data on up to 100 proteins to learn a coarse-grained molecular energy function to predict conformational properties of IDPs.[3-6] Using our Bayesian formalism[3] we have optimized a transferable model, called CALVADOS, to study the conformational ensemble of IDPs and flexible multi-domain proteins in the absence of experimental data.[4-6] I will also briefly describe how this model enables us to study interactions within and between proteins in biomolecular condensates.[4,6]

I will then describe how CALVADOS makes it possible to perform large-scale simulations to explore the relationship between sequence, structure, and function of IDRs.[7] I will describe how we have

---

[1] Holehouse, A. S., & Kragelund, B. B. (2023). The molecular basis for cellular function of intrinsically disordered protein regions. *Nature Reviews Molecular Cell Biology*, 1-25.

[2] Lindorff-Larsen, K., & Kragelund, B. B. (2021). On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *Journal of Molecular Biology*, *433*(20), 167196.

[3] Norgaard, A. B., Ferkinghoff-Borg, J., & Lindorff-Larsen, K. (2008). Experimental parameterization of an energy function for the simulation of unfolded proteins. *Biophysical Journal*, *94*(1), 182-192.

[4] Tesei, G., Schulze, T. K., Crehuet, R., & Lindorff-Larsen, K. (2021). Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proceedings of the National Academy of Sciences*, *118*(44), e2111696118.

[5] Tesei, G., & Lindorff-Larsen, K. (2023). Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Research Europe*, *2*, 94.

[6] Cao, F., von Bülow, S., Tesei, G., & Lindorff-Larsen, K. (2024). A coarse-grained model for disordered and multi-domain proteins. *bioRxiv*, 2024-02.

[7] Tesei, G., Trolle, A. I., Jonsson, N., Betz, J., Knudsen, F.E., Pesce, F., Johansson, K. E., & Lindorff-Larsen, K. (2023). Conformational ensembles of the human intrinsically disordered proteome. *Nature, 626(8000), 897-904.*

generated conformational ensembles of all intrinsically disordered regions of the human proteome, and used these to provide insight into sequence-ensemble relationships and evolutionary conservation of IDR properties.

Finally, I will describe initial work on how we can use the information encoded in CALVADOS to design disordered proteins with desired conformational properties.[8] I will describe the basic design algorithm and experimental validation on both single-chain compaction and measurements of phase separation.

[8] Pesce, F., Bremer, A., Tesei, G., Hopkins, J. B., Grace, C. R., Mittag, T., & Lindorff-Larsen, K. (2023). Design of intrinsically disordered protein variants with diverse structural properties. *bioRxiv*, 2023.10.22.563461