



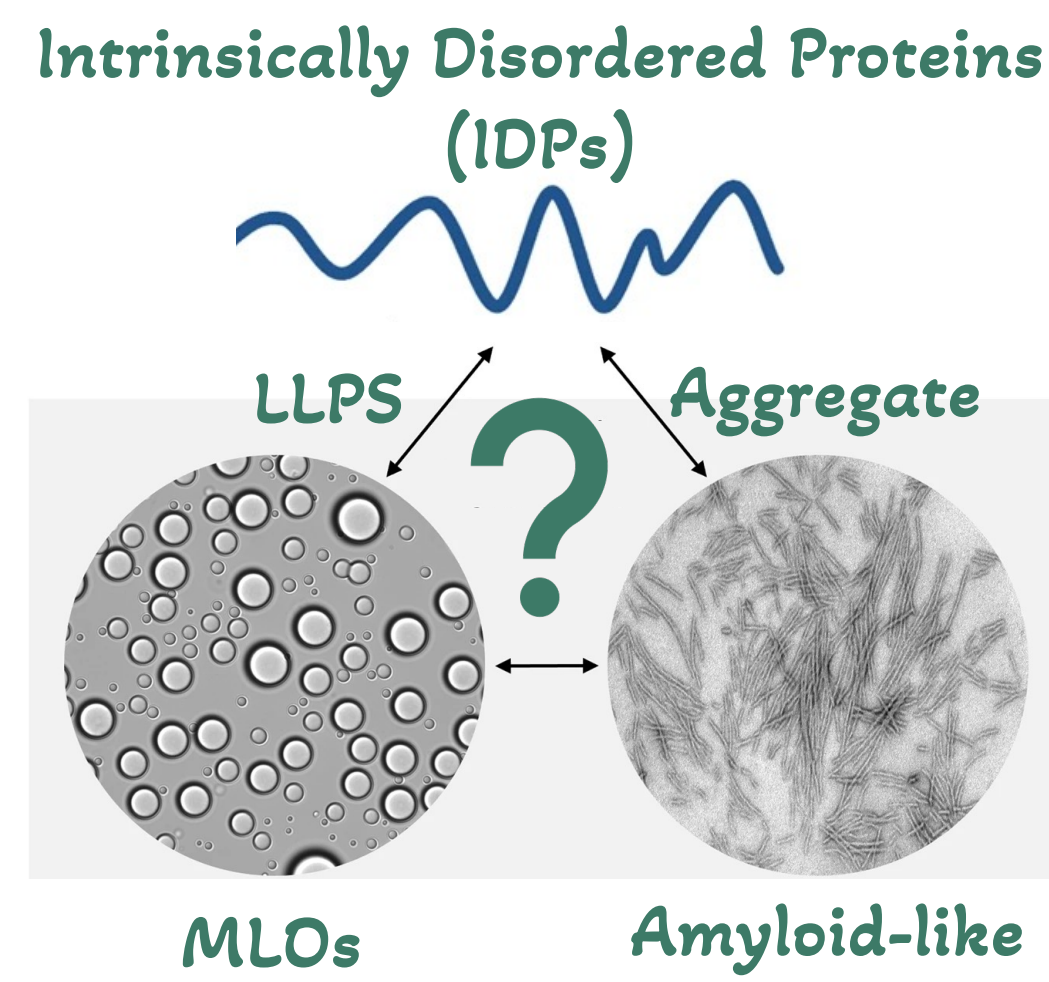
Protein Language Model Identifies Disordered, Conserved Motifs Implicated in Phase Separation

Yumeng Zhang¹, Jared Zheng², Bin Zhang^{1*}

¹MIT Department of Chemistry, ²MIT Department of Electrical Engineering and Computer Science

Motivation & Questions

Intrinsically disordered proteins (IDPs) are key players in liquid-liquid phase separation (LLPS) and frequently regulate the formation of membraneless organelles (MLOs). Mutations in IDPs can disrupt their multivalent interaction network, altering the phase behavior, and leading to various diseases.

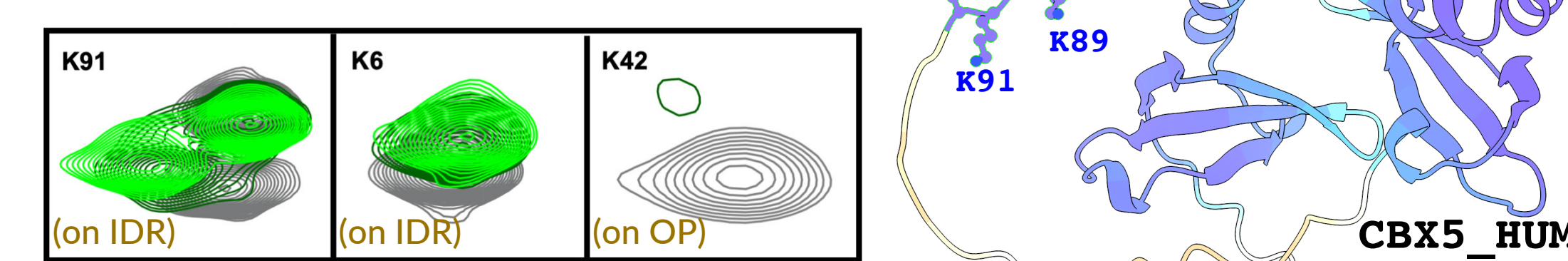


IDPs: Beautifully Unpredictable

IDPs evolve fast, fold rarely, yet function precisely

Sequence conservation for IDRs may indicate conserved functions.

HP1α (AlphaFold2) structure and sequence alignment with UniRef100 database.



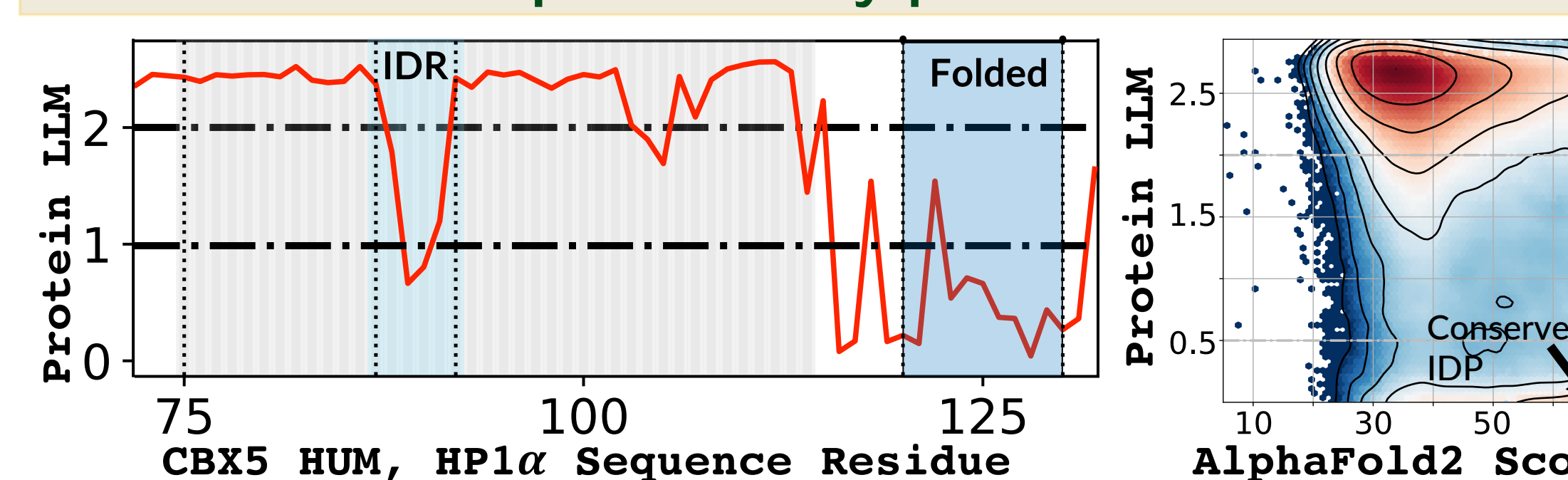
Evolutionary analysis for IDPs is challenging!

Sequence alignment of HP1α (AlphaFold2) with UniRef100 database. Conserved residues are highlighted in red.

From Words to Protein: Large Language Model

Large Language Model (LLM) ChatGPT (4o)

LLM can be particularly powerful for IDP

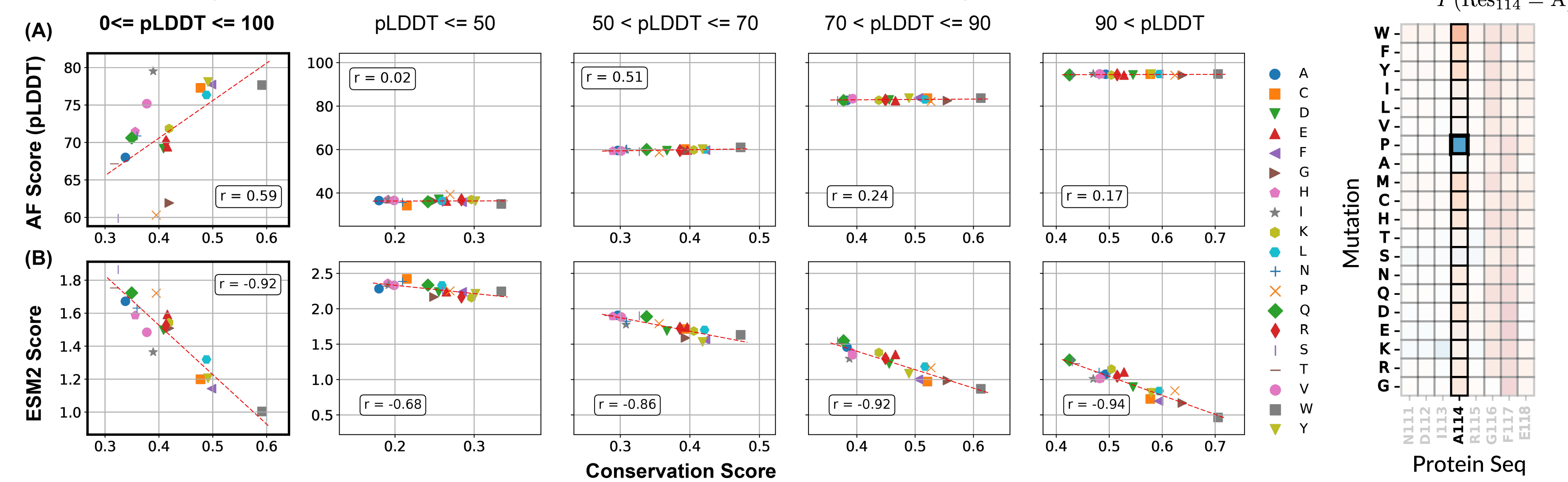


Results: Biophysical Roles behind LLM

ESM2 (Protein-LLM) scores reflect evolutionary conservation.

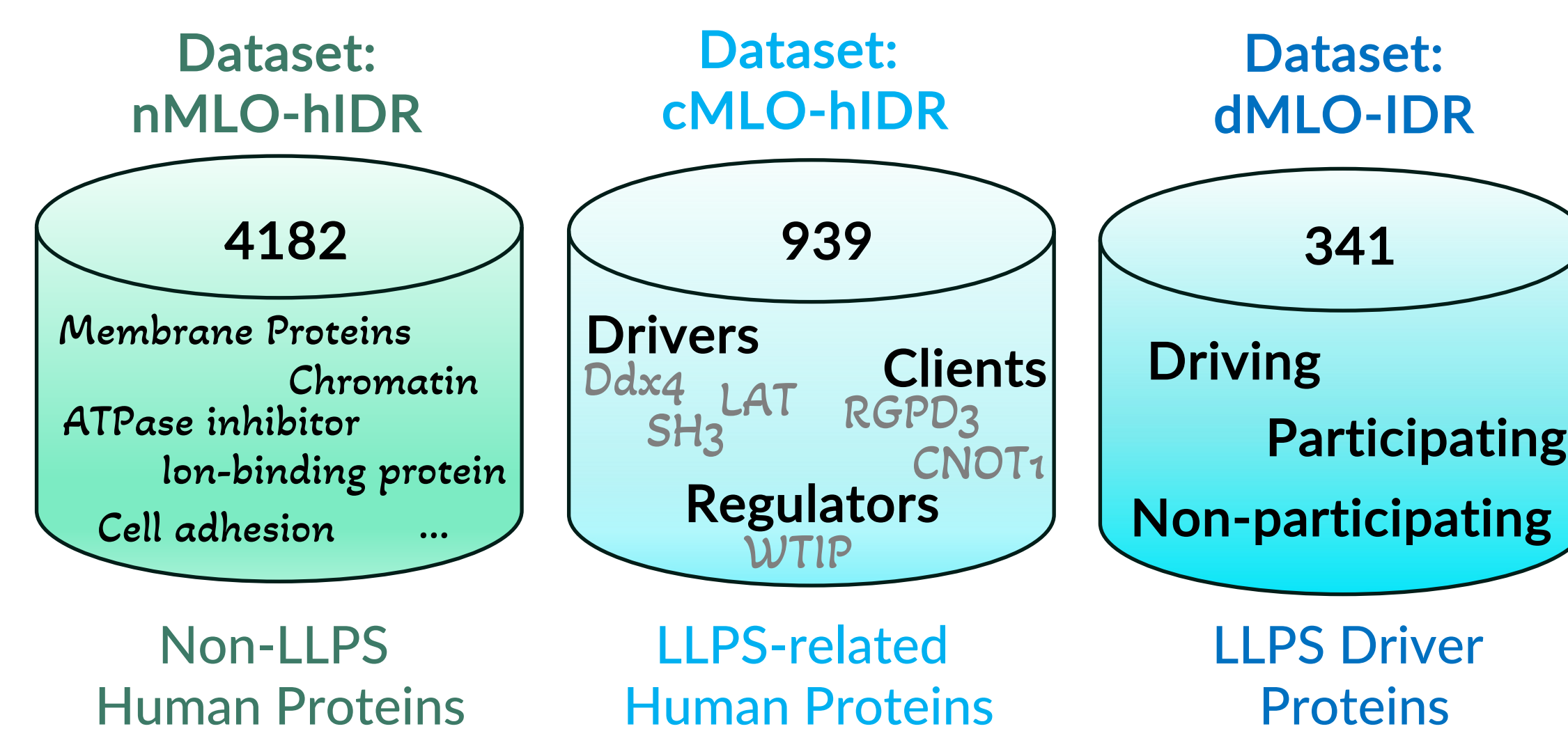
Conservation Score: from multi-sequence alignment (MSA)

ESM2 Score: Log-likelihood ratio value from Protein LLM ESM2 model in PlogP scale

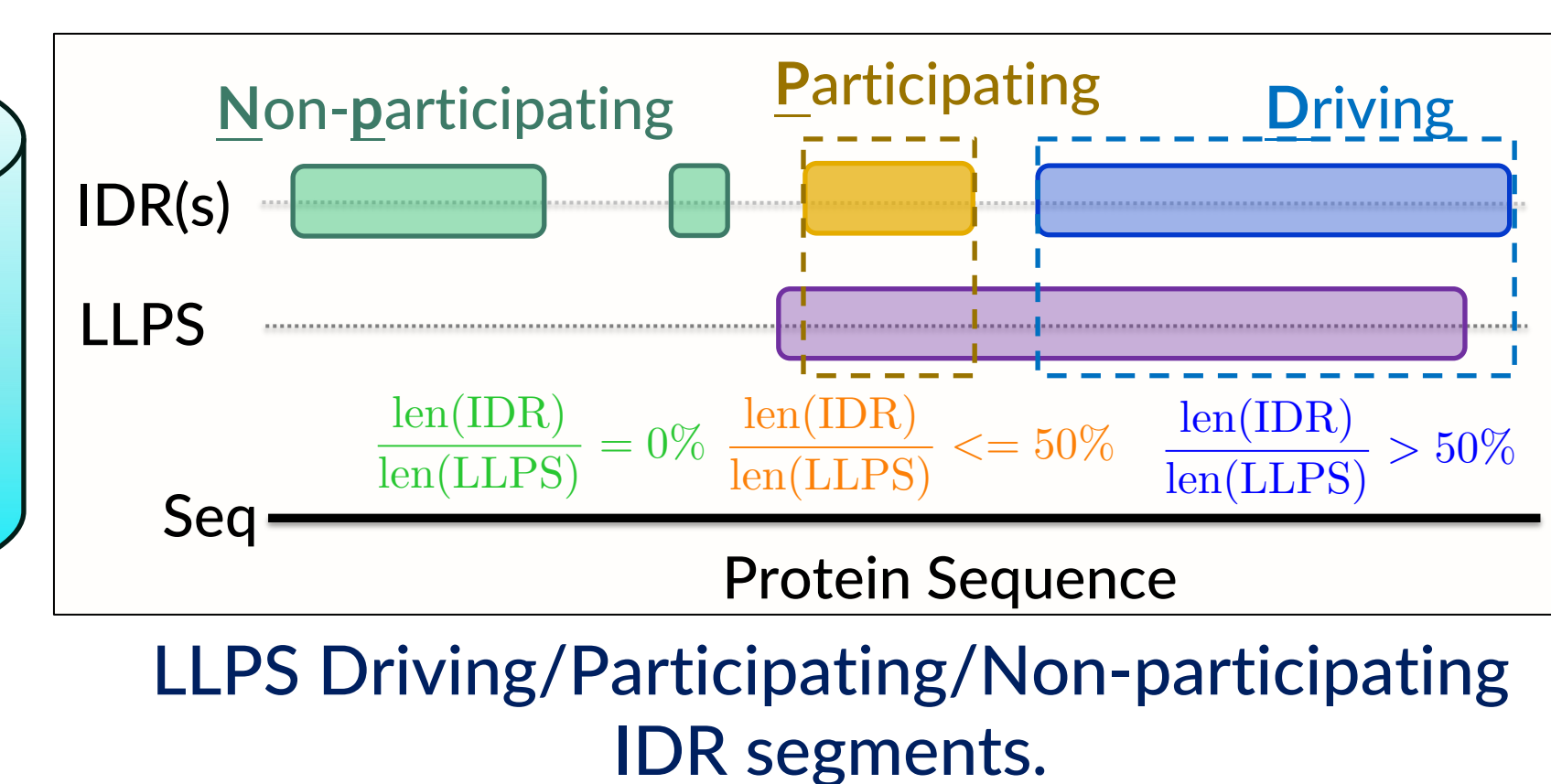


Higher population of conserved disorder residues appear in LLPS-related proteins.

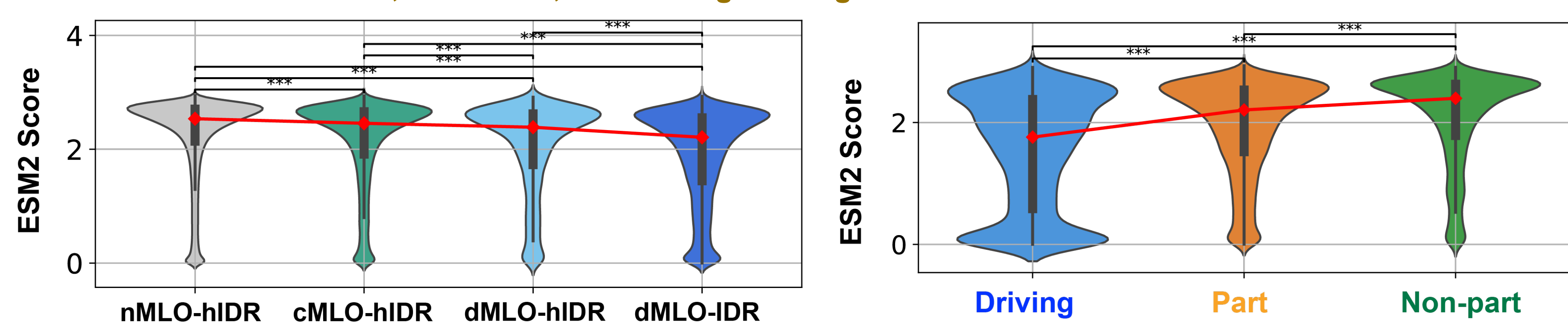
Dataset Preparation



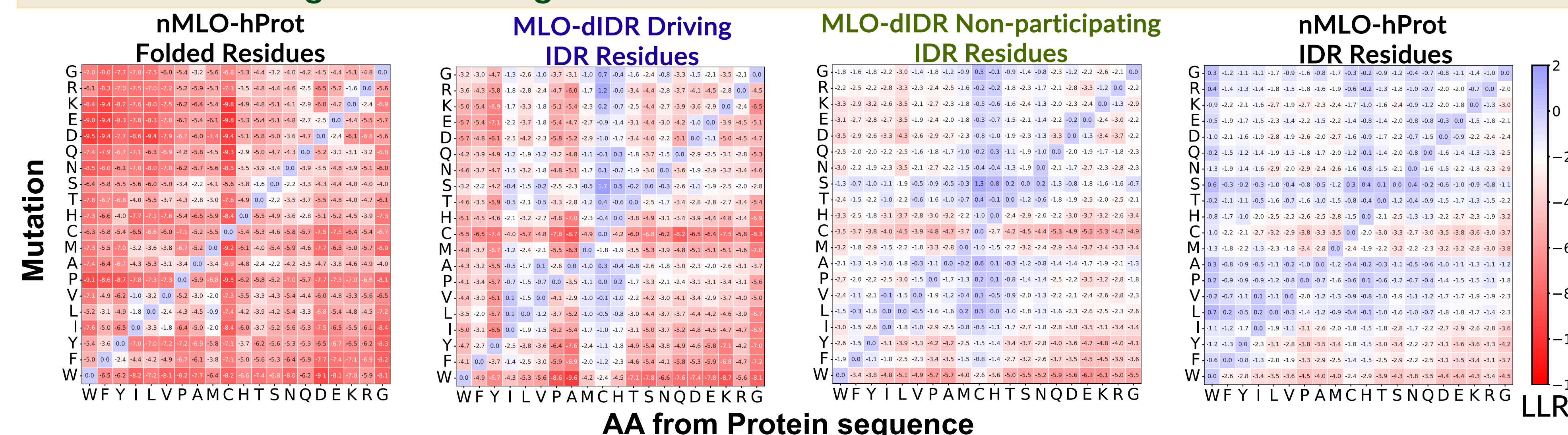
Segment dMLO-IDR Dataset Based on their LLPS functions



LLPS related, dMLO-IDR, and Driving IDR segments enrich in conserved residues

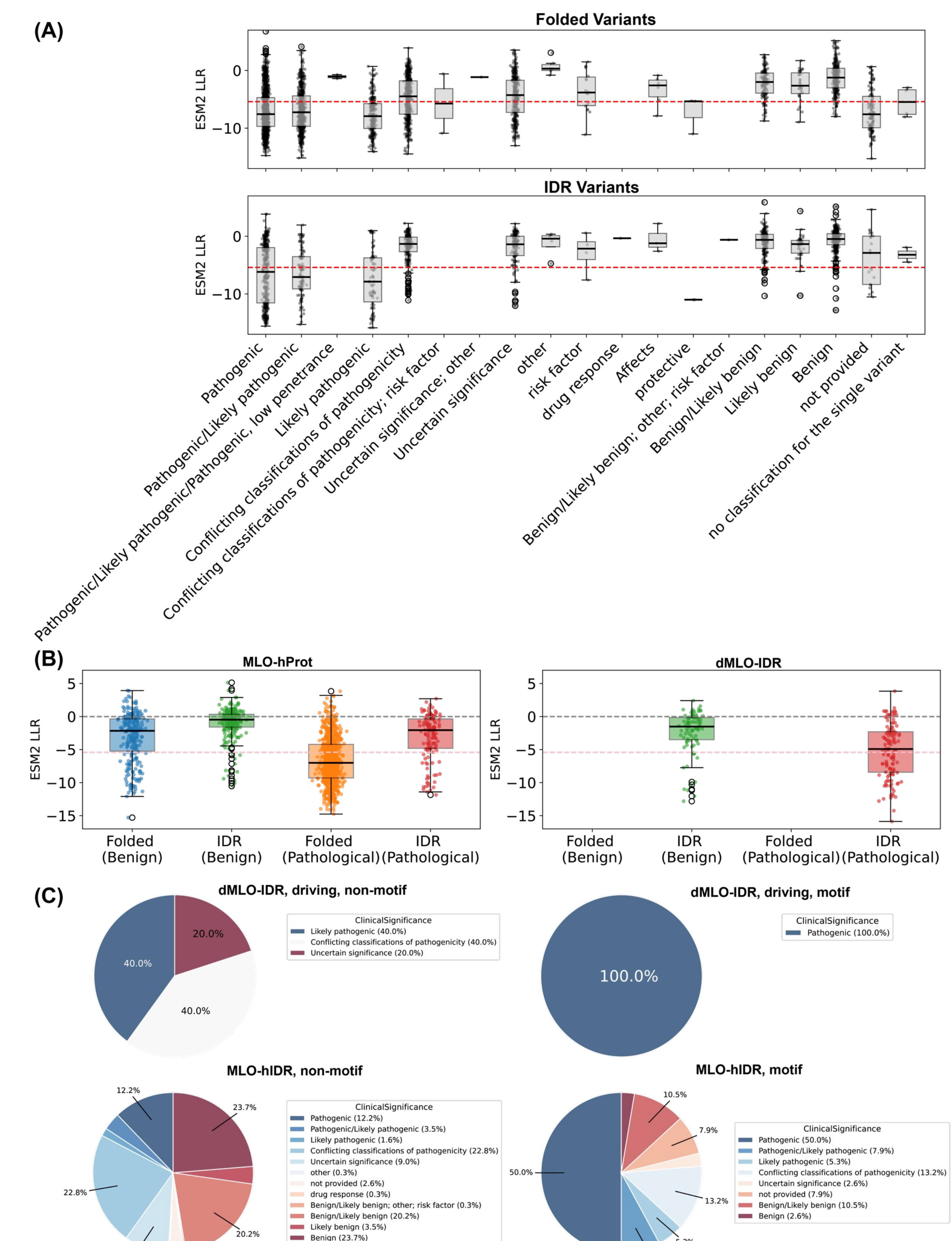


dMLO-Driving IDR show higher Resistance to Mutations.



Results: IDR Conserved Motifs and Diseases

Mutation on IDR conserved residues are likely to be pathological.



Conclusions

- 1. ESM2 Score Indicates Evolution Conservation.
- 2. ESM2 identifies disordered, conserved regions.
- 3. LLPS-driving IDPs contain more conserved residues.
- 4. Mutations appear at IDR conserved region are pathological.

References

- [1] Zhang et al, eLife (2025)
- [2] Ukmar-Godec et al, Proc. Natl. Acad. Sci (2023)
- [3] Lee et al, Exp Mol Med (2022)
- [4] Lin et al, Science (2023)
- [5] Jumper et al, Nature (2021)
- [6] Sood et al, Biophysics. J. (2024)

Acknowledgement

