PI: Professor Paul Harrison
Department of Biology

Exploring the Conservation of Intrinsically Disordered Protein Regions in Drosophila

Julien Hovan, B.Sc. Computer Science and Biology

Background:

Intrinsically disordered proteins (IDPs) have been in the spotlight for their unique properties, such as their lack of secondary structures and low sequence complexity. IDPs are a very large and functionally important class of proteins and their discovery has disproved that the three-dimensional structures of proteins must be fixed in order to accomplish their biological functions (Park et Al, 2022).

IDPs in humans have been commonly correlated with neurodegenerative diseases such as Parkinsons, Huntingtons, Alzheimers. This has been a recently interesting area of research as these disease-related IDPs with low sequence complexity have been shown to possess a spectrum of unique properties, such as heat resistance and unique aggregation properties. IDPs are commonly known to undergo different types of post-tranlsational modifications (PTMs) such as phosphorylation and ubiquitination (Park et Al, 2022).

Compositionally-biased regions (CBRs) in biological sequences are enriched for a subset of sequence residue types. These can be shorter regions with a concentrated bias ('low-complexity'), or longer regions that have a compositional skew. These regions comprise a prominent class of the uncharacterized 'dark matter' of the protein universe. This compositional bias for a subset of residues is a widespread phenomenon in protein sequences, for it has historically been linked to proteins having a structural role, or displaying some intrinsic protein disorder (Harrison PM, 2006).

Drosophila melanogaster is a small, common fly found near unripe and rotted fruit. It has been in use for over a century to study genetics and behavior. For humans and fruitflies, CBRs have been analyzed for conservation, length, functional linkages, and predicted protein disorder content. Some of the universally abundant biases are linked to nuclear localization and transcription in Human and/or Drosophila (Harrison PM, 2006). There are prevalent CBRs in the Drosophila proteome, particularly rich in residues that all have adenine and cytidine codons.

Objectives:

By exploring the dynamics of conservation for these compositionally-biased regions (CBRs) in Proteome of the Drosophila species, this research aims to explore and possibly expand on the evolutionary importance of CBRs and the role of IDPs in Drosophila. We can derive the implications to the genome of Drosophila and have a more complete understanding of the functionality of these intrinsically disordered proteins. For example, a highly conserved region may play a critical role in protein-protein interactions, substrate recognition, or regulation. Understanding the conserved regions of a protein can provide important information for drug discovery and therapeutic design. It can also aid in the identification of new biological targets and the understanding of the molecular mechanisms of disease.

Methodology (to be expanded):

To explore the conservation of CBRs in Drosophila, multiple sequence alignments (MSAs) will be the main method to identify conserved regions or motifs within the protein and to compare their sequence across Drosophila species. We will be using UniProt, the world's leading resource of protein sequences, to access 27 species of Drosophila and their sequenced proteomes.

To identify the CBRs, we will use *fLPS 2.0* package (Harrison PM, 2021) , which rapidly annotates CBRs and also includes added consideration of DNA sequences. It outputs protein CBRs labeled with bias classes reflecting the physico-chemical character of biasing residues.

Evaluation:

We will be considering a couple of metrics to gauge the success of the approach: The evolutionary significance of the conserved region can be evaluated by comparing the conservation patterns of the protein across different taxa and by inferring the evolutionary history of the protein. Also quantifying the sequence conservation can be compared to other species with prevalent CBRs to extrapolate the confidence of the MSA.

Bibliography:

Park, H., Yamanaka, T. & Nukina, N. Proteomic analysis of heat-stable proteins revealed an increased proportion of proteins with compositionally biased regions. Sci Rep 12, 4347 (2022). https://doi-org.ezproxy.lib.gla.ac.uk/10.1038/s41598-022-08044-z

Harrison, P.M. Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and Drosophila. BMC Bioinformatics 7, 441 (2006). https://doi.org/10.1186/1471-2105-7-441

Harrison PM. fLPS 2.0: rapid annotation of compositionally-biased regions in biological sequences. PeerJ. 2021 Oct 28;9:e12363. doi: 10.7717/peerj.12363. PMID: 34760378; PMCID: PMC8557692.