**AB INTIO PREDICTION OF INTRINSICALLY DISORDERED PROTEINS (IDPs) PROPERTIES USING LARGE LANGUAGE MODELS**

**ABSTRACT**

Intrinsically Disordered Proteins(IDP) are a unique class of proteins that lack a stable, well-defined three-dimensional structure. IDPs are important because they play crucial roles in a wide variety of biological functions. Their structural flexibility and dynamic behavior allow them to adapt to multiple targets and conditions. For detecting IDPs , traditional diagnostic techniques, such as Nuclear Magnetic Resonance Spectroscopy, Circular Dichroism Spectroscopy, Small-Angle X-ray Scattering, Single-Molecule Fluorescence Resonance Energy Transfer, and Mass Spectrometry are used. However, these methods have inherent limitations, such as low structural resolution, difficulty in analyzing large or complex proteins, and sensitivity to environmental conditions. Additionally, these methods are time-consuming and expensive. Therefore, Artificial Intelligence (AI) act as a pathway to overcome limitations by improving data resolution, predicting protein properties, handling large proteins, and reducing time and cost. Although, several deep learning models have been developed for predicting IDPs from their sequences, these models face difficulties in capturing 3D structures, and encounter challenges in capturing long range information from large sequences. To address these limitations, an efficient method leveraging Protein Language Models (PLMs) to map sequences directly to IDPs properties is proposed. All the experiments will be conducted on a protein dataset extracted from the DisProt database. The dataset contains 2,306 samples, categorized into four classes: sequences(seqs), radius of gyration (rog), Heat capacity (cv), Decorrelation Time(tau). The model will be evaluated using evaluation metrics such as Mean Squared Error (MSE) and Coefficient of determination (R2). The proposed approach aims to efficiently predict IDPs properties directly from amino acids, including Radius of Gyration, end-to-end Decorrelation Time, and Heat Capacity for a given arrangement of amino acids in protein sequences.

**Team Members Internal Guide**

M.Varshitha - 21WH1A1277 Dr. Mukhtar Ahmad Sofi

K.Vigna - 21WH1A1296 Assistant Professor

V.Sahithya - 21WH1A12B5

CH.Ruthvika - 21WH1A12B9