A Machine Learning Approach to Identifying Ordered Binding Regions on Order-Disorder Protein Interfaces

Cui, Jake

Intrinsically disordered proteins (IDPs) are unique proteins that lack rigidly defined three-dimensional structures. Due to their conformational flexibility, IDPs serve diverse protein functions, and make up ? of cellular signalling proteins. Unsurprisingly, many proteins important in disease, including BRCA1 for breast cancer or tau for Alzheimer’s disease, possess significant regions of protein disorder. Currently, IDPs cannot be reliably targeted in drug development as there is no way to predict specific IDP binding regions. Here, a machine learning approach was used to generate a Bayesian network to probabilistically model relationships between protein characteristics. A sliding window algorithm was created with a binary classifier that can identify the presence of IDP binding sites on 10-residue ordered protein segments at approximately 76% accuracy. These results provide the groundwork for a comprehensive IDP binding-site identifier that will facilitate the creation of modern drug treatments targeting regions of intrinsic disorder in diseases like cancer, Alzheimer's, heart disease, and more.

Awards Won:

Fourth Award of $500