A Novel Phosphosite Localization Approach Using Tandem Phosphoprotein Mass Spectra and Temporal Dilated Networks

Arthur, Margaret (School: Fairview High School)

Dysregulation of protein phosphorylation during cellular development and signal transduction is associated with a myriad of human cancers, Parkinson’s Disease, and Alzheimer’s Disease. Understanding the exact amino acid binding sites is critical for medical and pharmacological applications, but current experimental and computational methods are time-consuming or imprecise. I developed a novel approach to identifying phosphosites using peptide tandem fragmentation mass spectra, which is more applicable for drug design compared to the current method of sequence input. I created a temporal convolutional neural network (TCN) to identify the exact phosphorylated amino acid from a given spectrum. I processed the mass spectra data to create a two-vector representation for efficient analysis. My model uses a unique architecture that promotes flexibility and interpretability with considerations for long-range associations between spectral peaks critical for phosphosite identification. I accounted for neutral losses, a common phenomenon in proteomics mass spectrometry, through the independent identification of unique spectral patterns. Instead of requiring existing input features that could increase bias or decrease the identification of rare or new phosphosites, my model identifies its own patterns. My model achieved an average accuracy of 95% compared to the accuracies (76 - 78%) of sequence-based computational models. Future work includes developing my model to predict neutral losses. My novel approach provides a critical step to the biological understanding of phosphorylation events and the effective development of medical and pharmacological tools to correct biochemical pathways for targeted treatments that reduce risk and discomfort for patients.

Awards Won:
Fourth Award of $500