Society of Physics Students presents:

PHYSICS ALUMNI SERIES SEMINAR

Molecular Dialogues: Guiding Protein Design using Natural Language Prompts

Niksa Praljak '20 Biophysics PhD candidate, University of Chicago

Designing proteins with specific functional properties presents a significant challenge in molecular engineering. This work introduces a novel computational framework leveraging natural language prompts as an intuitive control mechanism for guiding the design of functional proteins. Our approach, termed the Biological Multi-Modal Model (BioM3), uniquely integrates information from protein sequences and natural language descriptions to facilitate the generation of diverse protein designs. BioM3 employs a multi-stage strategy involving the alignment of protein and text representations using foundation models and contrastive learning techniques, followed by the refinement of text-based features and the generation of protein sequences through a discrete order-agnostic generative model. This innovative method excels at capturing subtle functional relationships between proteins, even in the absence of significant sequence similarity. Notably, our joint embedding space demonstrates enhanced performance in identifying proteins with similar structural folds compared to traditional sequence-based search methods. The framework enables the functional categorization of proteins based on natural language descriptions, spanning attributes such as enzymatic activity and cellular localization. BioM3 showcases its versatility by generating a wide range of protein types without requiring specialized training for each category. Furthermore, we experimentally validated the functionality of designed proteins, demonstrating the ability to rationally control sequence diversity and maintain desired properties through intuitive manipulation of textual prompts. This research establishes the power of generative multi-modal models as a transformative tool for translating natural language intent into functional proteins, paving the way for accelerated advancements in synthetic biology, biotechnology, manufacturing, and medicine.

Tuesday, April 29th, 2025

11:30 am, SR 151

Pizza and Refreshments Provided