Physical Chemistry Seminar

Bridging Data-Driven Generative Models and Physical Interactions: A Modular Framework for Biomolecular Interfaces

Yanbin Wang

Postdoctoral Research Fellow, The Chen Group
Purdue University

In recent years, foundation generative models have been developed to predict molecular and material structures. While successful, their performance is often limited by the availability of high-quality training data, especially for protein conformations under protein–environment interactions (PEI), such as those involving organic linkers or material surfaces.

We propose a physics-guided framework that extends foundation models beyond their training domain by coupling pretrained generative models with explicit, physics-based interaction potentials describing PEI. This coupling directs sampling toward conformations consistent with physical constraints, without retraining. We demonstrate accurate, efficient predictions for (i) cyclic peptides linked with organic molecules and (ii) peptides adsorbed on gold surfaces. The generated structures provide high-quality starting points for molecular simulations, offering a systematic strategy to extend foundation models to proteins under environmental influences.

The framework naturally extends to soft-matter interfaces by representing each subsystem, such as a protein, polymer, or surface, with its own diffusion model trained on subsystem-specific data. During inference, these models jointly sample conformations consistent with their coupled interactions, significantly reducing data requirements. As a demonstration, we modeled complexes between the amyloid-beta peptide ($A\beta_{1-42}$) and a polyethylene oxide (PEO) chain. Traditional molecular dynamics struggles to generate meaningful initial configurations and requires extensive time to overcome energy barriers. In contrast, our coupled diffusion models, one for the protein and one for the polymer, efficiently produced realistic protein–polymer complexes and enabled rapid exploration along a desorption coordinate to identify stable binding configurations.

Overall, this modular, physics-guided generative strategy bridges data-driven learning and physical modeling, enabling rapid exploration of biomolecular interfaces and accelerating materials discovery.

Wednesday, October 15, 2025 **10:30am** **BRWN 4102**

PURDUE UNIVERSITY | James Tarpo Jr. and Margaret Tarpo Department of Chemistry

Wang Bio

Dr. Yanbin Wang is a Postdoctoral Research Fellow in the Department of Chemistry at Purdue University, where he develops physics-guided generative models for proteins and polymers. His interdisciplinary background spans mechanics, materials science, and chemistry, with research focusing on statistical thermodynamics, molecular simulations, and AI for Science. He earned his Ph.D. in Mechanical Engineering from the University of Maryland, College Park, where he studied interfacial water and ion dynamics under Prof. Siddhartha Das. Dr. Wang's recent work includes developing the DiffPIE framework for protein–environment interactions and advancing machine learning–aided sampling of complex biomolecular conformations. He is passionate about combining data-driven modeling with physical theory to accelerate scientific discovery.