Executive Summary

Jayant Kumar Singh

Professor and Head, Department of Chemical Engineering, IIT Kanpur

1. Title of the Project: **To develop a high-throughput platform for accelerating the discovery for materials**

2. Date of Start of the Project: October 3, 2023

3. Aims and Objectives:
   - To develop a platform for on-the-fly calculations of materials properties
   - To integrate state-of-art Monte Carlo and MD and mesoscale simulatoins methods and visualisation tools
   - To Integrate AI/ML and molecular language for accelerating the discovey of materials
   - To integrate the material platform with the PrinS3 platform

4. Significant achievements (not more than 500 words to include List of patents, publications, prototype, deployment etc)

   The project involves developing A simulation Platform for Molecular Simulation with an aim to integrate high-through discovery algorithms. In the last one year, we have mainly focused on development/utilization of algorithm and methods to expedite the discovery of new materials be it for HER or CO2RR, gas adsorption and separation or surfactant/antifreeze design. In addition, we have made some progress in our software design. The key contribution we have made some progress in two distinct areas:

   1. Integrating active learning to find optimal structure/materials for gas separation:

Harnessing digital innovation through artificial intelligence and machine learning has woven into the very fabric of computational discovery, opening new avenues for designing and discovering novel materials, particularly for gas separation applications. In the past year, we have made significant progress in integrating active learning to accelerate the discovery of optimal structures and materials for gas separation. By leveraging machine learning algorithms in conjunction with molecular simulations, we have automated the iterative process of structure-property optimization. This active learning framework refines the selection of candidate materials by continuously learning from simulation data, significantly reducing the computational time and effort required. As a result, we can efficiently explore a vast chemical space and identify high-performing materials for gas separation applications, advancing the overall platform's capabilities.

In our work[1], we specifically applied these advancements to the challenging problem of separating xylene isomers, a critical issue in the petrochemical industry due to their similar boiling points, which demand highly selective adsorbents. We employed an active learning (AL) framework integrated with molecular simulations to screen 324,426 hypothetical metal-organic frameworks (hMOFs) for optimal para-xylene (pX) adsorption. Our approach demonstrated a significant improvement over traditional methods, requiring only around 500 simulations to identify 50.5% of the top 100 candidates. A combination of a Gaussian process surrogate model with the expected improvement (EI) acquisition function proved most efficient, identifying the highest pX-selective MOF in just 86 acquisitions. Furthermore, structural analysis from our simulations revealed that pX selectivity correlates strongly with the pcu topology and a pore size of 5–6 Å. This AL-driven workflow, developed through our efforts, significantly accelerates material discovery, providing a highly efficient and data-driven approach to identifying high-performance materials for gas separation.

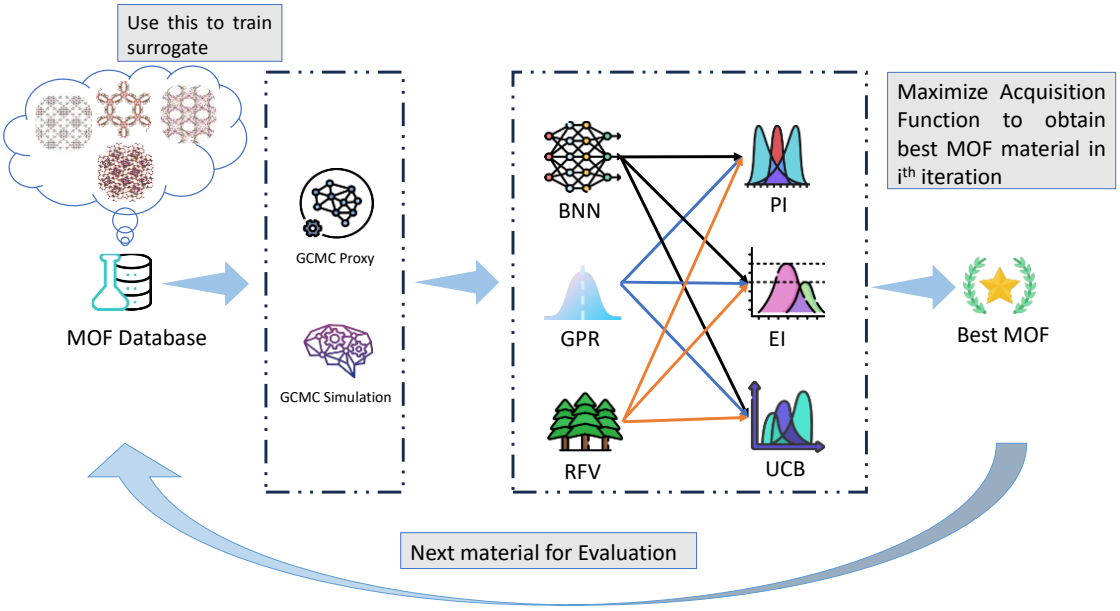

Figure 1: An overall workflow depicting active learning framework.

2. Using molecular fingerprint of anti-freeze proteins for identify MOF antifreezing activity.

Ice nucleation is a ubiquitous phenomenon when the temperature drops below 0°C, which poses significant challenges in the preservation of cells, organs, and food materials. The formation of ice, with its needle-like structure, can cause damage to cellular structures, leading to reduced cell viability. To counteract this, naturally occurring antifreeze proteins (AFPs) have been employed to inhibit ice formation. However, extracting these proteins from living organisms is time-consuming and costly. While biomimetic alternatives to antifreeze proteins have been explored, their application has been limited due to concerns about toxicity. In response to this, Metal Organic Frameworks nanoparticles (MOF NPs) could serve as an alternative for antifreezing activity in the cryopreservation of cells, organs, and food because of their highly tunable nature. To understand in detail, we formulated a model based on amino acid data available in the literature. Using Morgan fingerprints of molecules, we identified key functional groups relevant to antifreeze proteins. Our screening focused on the CoRE MOF database (2019), where we prioritized organic linkers containing -COOH functional groups, which were confirmed to be significant based on Morgan fingerprints. We then used machine learning to train a model using literature data on antifreeze molecules, targeting mean grain size as the key property. Based on this model, we predicted the Ice Recrystallization Inhibition (IRI) activity of various organic linkers in terms of their impact on mean grain size. This

approach opens up the possibility of using MOF NPs as a less toxic and more efficient alternative for cryopreservation applications.

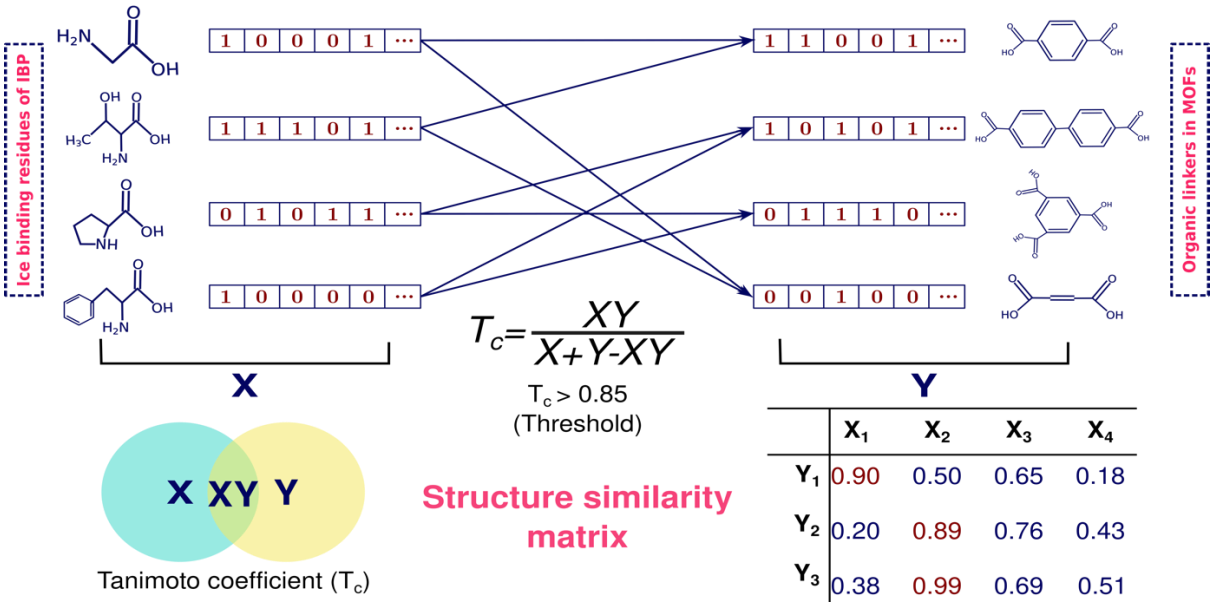

Figure 2: Schematic representation of morgan fingerprint for different chemical moieties

Publications:

(1) Aqib, M.; Daoo, V.; Singh, J. K. Efficient Xylene Isomer Separation: Accelerated Screening with Active Learning and Molecular Simulation. *Energy and Fuels*, 38, 9381, 2024

(2) Machine learning driven advancements in catalysis for predicting hydrogen evolution reaction activity, *Materials Chemistry and Physics*, 326, 129805, 2024

5. Concluding remarks

In conclusion, our work demonstrates the transformative potential of integrating machine learning and molecular simulations for material discovery. The active learning framework applied to gas separation, particularly for challenging problems like xylene isomer separation, significantly accelerates the identification of optimal structures while reducing computational costs. Similarly, our innovative approach using molecular fingerprints to predict antifreezing activity in MOF nanoparticles offers a promising, less toxic alternative for cryopreservation. These advancements pave the way for future explorations in material optimization and contribute to the broader adoption of AI-driven discovery in industrial applications.