Enabling Machine Learning-Assisted Discovery of Polyamines for Solid-State CO_2 Capture

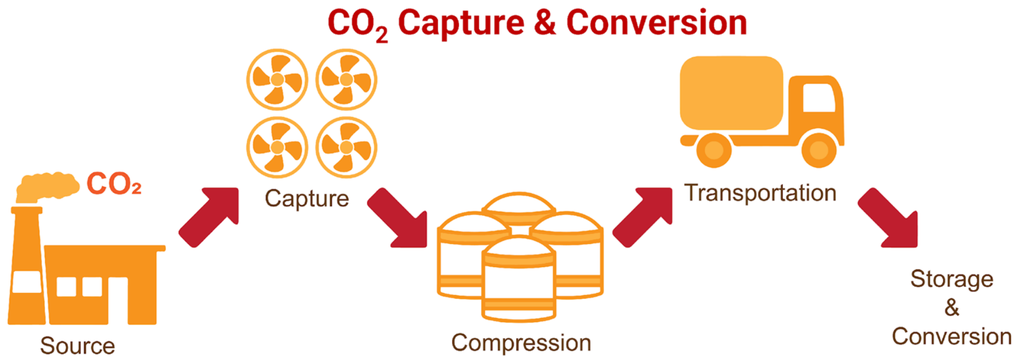
A N M Nafiz Abeer[1,*], Junhe Chen[2,*], Alif Bin Abdul Qayyum[1], Zhihao Feng[2], Hyun-Myung Woo[3], Seung Soon Jang[2], Byung-Jun Yoon[1, 4]

[1] Texas A&M University, [2] Georgia Institute of Technology, [3] Incheon National University, [4] Brookhaven National Laboratory

Polyamine Sorbent for Direct CO_2 Capture

Rise in atmospheric CO_2 and projected increase in global temperature

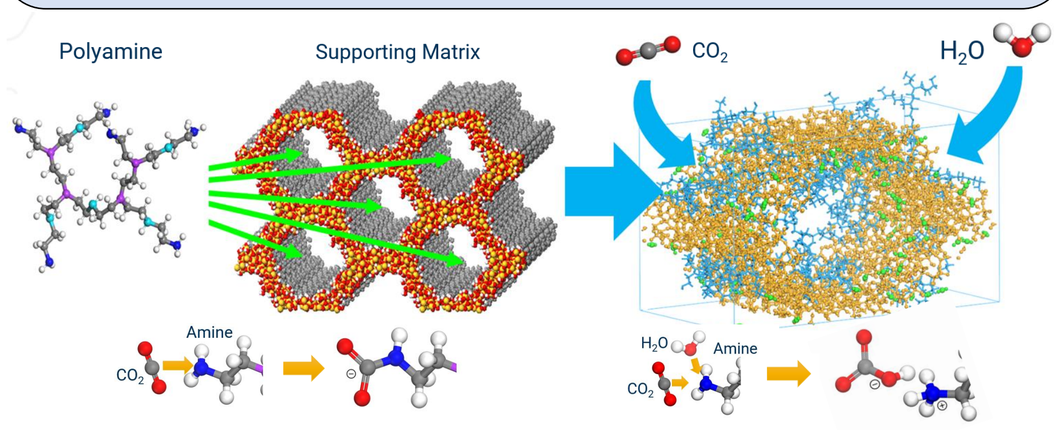
Active removal through Carbon Capture and Storage (CCS)



Need for highly selective sorbent for Direct Air Capture (DAC)

Aqueous vs. solid sorbents

- High energy requirements for solvent regeneration (aqueous)
- Performance of solid amine sorbents is governed by **accessibility of its amine groups, not by the surface area** → critical under sub-1vol% CO_2 conditions



Our Contributions

- Molecular Dynamics (MD): primarily utilized for molecular-level insights into sorption mechanism.
- Low throughput → infeasible for inverse design paradigm for polymer.

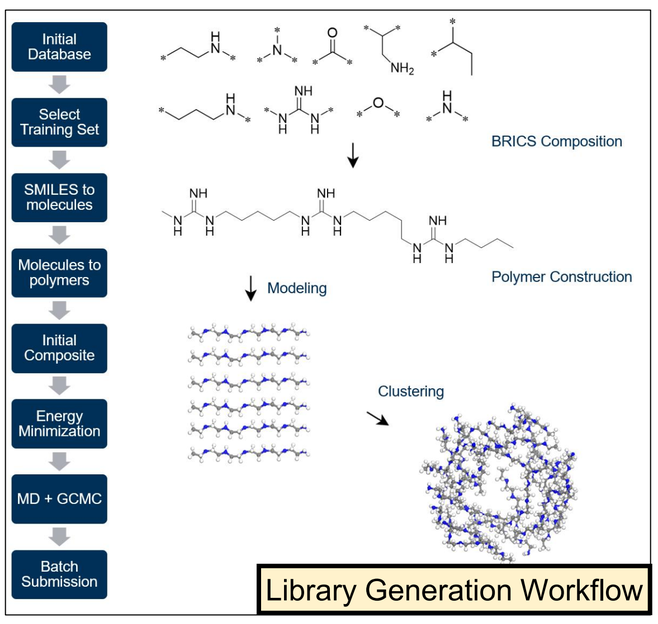
1. Developed **computationally efficient** simulation pipeline for CO_2 uptake estimates through fast MD relaxations + grand canonical Monte Carlo (GCMC)

2. Generated a **diverse library** of 1000 polymers with estimated CO_2 adsorption capacity.

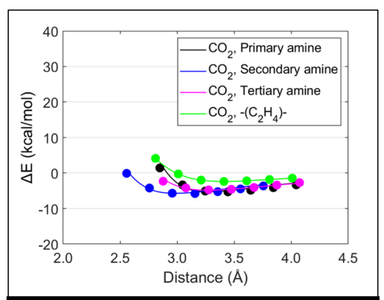
3. Investigated the utility of **Bayesian optimization (BO)** in accelerating discovery of polymer with **optimized adsorption capacity**.

Data Generation Pipeline

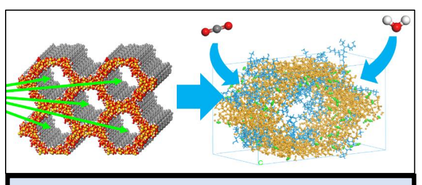
- **Refinement** of Lennard-Jones parameters → excellent **agreement** between **experimental and DFT-based** CO_2-H_2O-amines interaction. [1, 2]
- **Molecular fragments for library:** well-characterized polyamines, e.g., polyethyleneimine (PEI), polypropyleneimine (PPI) etc.
- **Fast MD:** to relax polymer conformations and approximate packing behavior, and **GCMC:** to simulate CO_2 adsorption under specific pressure and temperature.



Library Generation Workflow



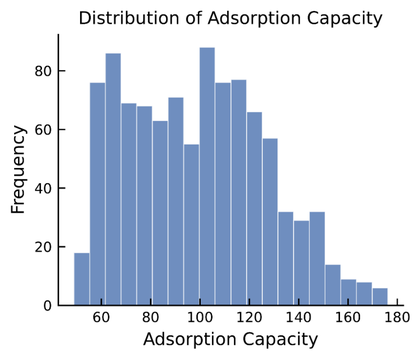
Force-field Refinement

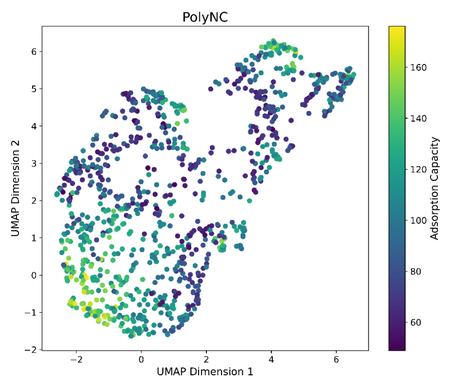


CO_2 uptake by polymers with silica-based support material

Retrospective BO Experiment

- **Initial pool of 10 labeled samples:** randomly selected from samples below $\gamma y_{\max}$.
- **Goal of BO:** to identify sample with best adsorption capacity from remaining 990 samples with **30 simulations**.





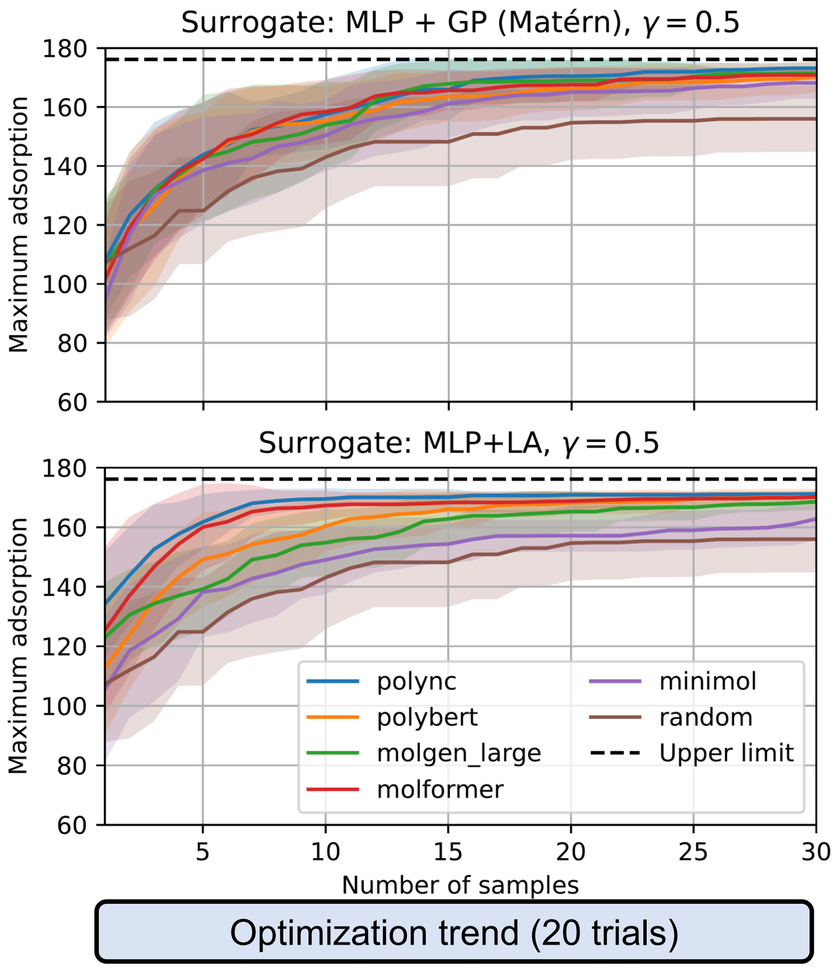
Surrogate model

- Gaussian process with deep kernel learning (GP-DKL)
- Bayesian neural network using Laplace approximation.

Polymer representation

- **Polymer-specific:** PolyNC [3], polyBERT [4]
- **General:** MoLFormer [5], MolGen-large [6], MiniMol [7]

Results and Key Takeaways



Optimization trend (20 trials)

- **Outcomes:**
  - **(1) Polymer library** with adsorption data
  - **(2) Simulation pipeline** based on fast MD + GCMC
- **BO showed promise in accelerating the high-performance polymer identification,** making it a promising approach under **small computational budget**
- **Future work:**
  - (1) Requires validation by long-timescale MD simulation.
  - (2) Need for exploration beyond linear polymers.

References

1. Chen et al. CO_2 Capture Characteristics of Hyperbranched Poly(alkylene imine): A Molecular Dynamics Simulation Approach. 2025.
2. Chen et al. Distribution and Transport of CO_2 in Hyperbranched Poly(ethylenimine)-Loaded MCM-41: A Molecular Dynamics Simulation Approach. 2023.
3. Qiu et al. Polync: natural and chemical language model for the prediction of unified polymer properties. 2024.
4. Kuenneth et al. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. 2023.
5. Ross et al. Large-scale chemical language representations capture molecular structure and properties. 2022.
6. Fang et al. Domain-agnostic molecular generation with chemical feedback, 2024.
7. Klaser et al. Minimol: A parameter-efficient foundation model for molecular learning. 2024.