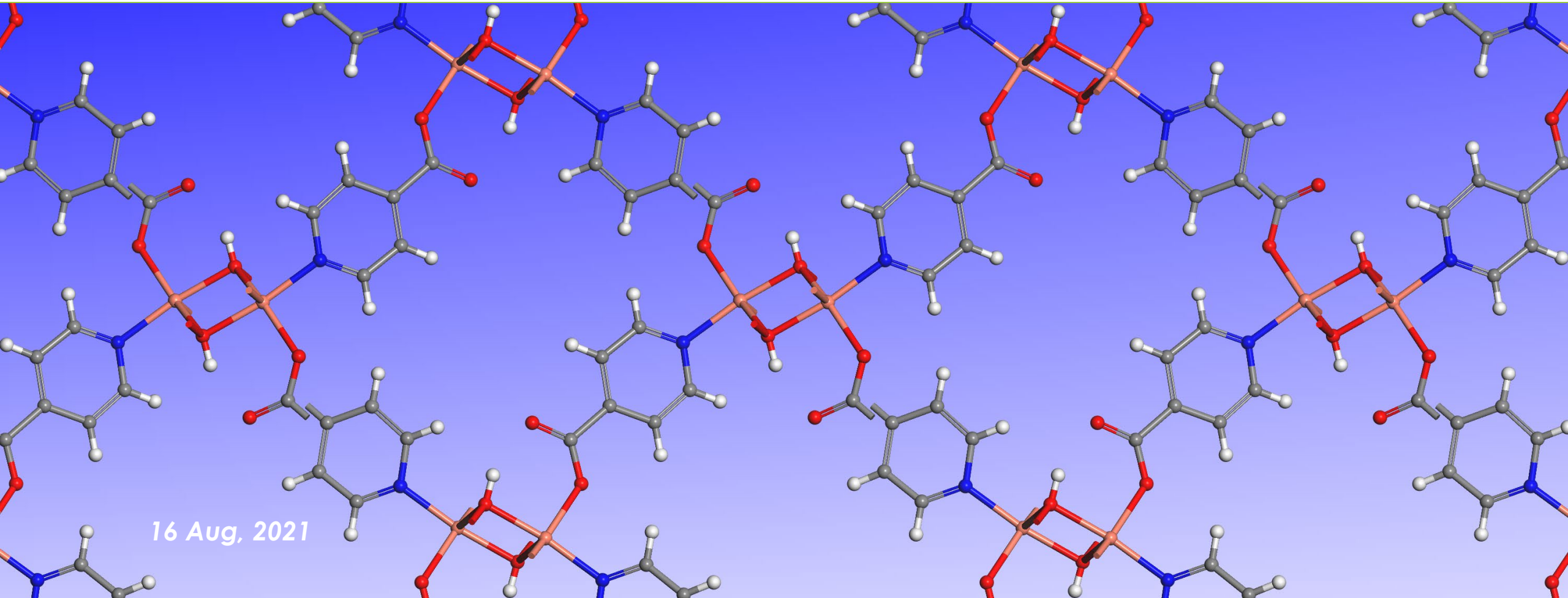
Computational Screening of Carbon Capture Materials
NETL/DOE, R&IC
Carbon Capture FWP 1022402

Surya Tiwari, Wei Shi, Samir Budhathoki and Jan Steckel

NATIONAL ENERGY TECHNOLOGY LABORATORY

Carbon Management and Natural Gas & Oil Research Project Review Mtg
Virtual Meetings August 2 through August 31, 2021



Outline

NETL R&IC Carbon Capture Field Work Proposal

Task 16: Computational Screening of Sorbents for Direct Air Capture

- Methodology Development: Machine Learning Applied to Flexible Force Field Development for Materials Screening
- Application is Sorbent for Direct Air Capture

Task 12: Computational Screening of Novel Membrane Materials

- Machine Learning applied to computational screening/design of membranes
- ML model predicts gas permeability and perm-selectivity of polymers
- Application is CO_2/N_2 membrane separation for industry (cement, steel, etc.)

Project Overview/Technology Background

Task 16: Computational Screening of Sorbents for Direct Air Capture

Screening MOFs for DAC sorbents.

MOFs can exhibit:

- enormous surface area,
- capacity for high sorption selectivity

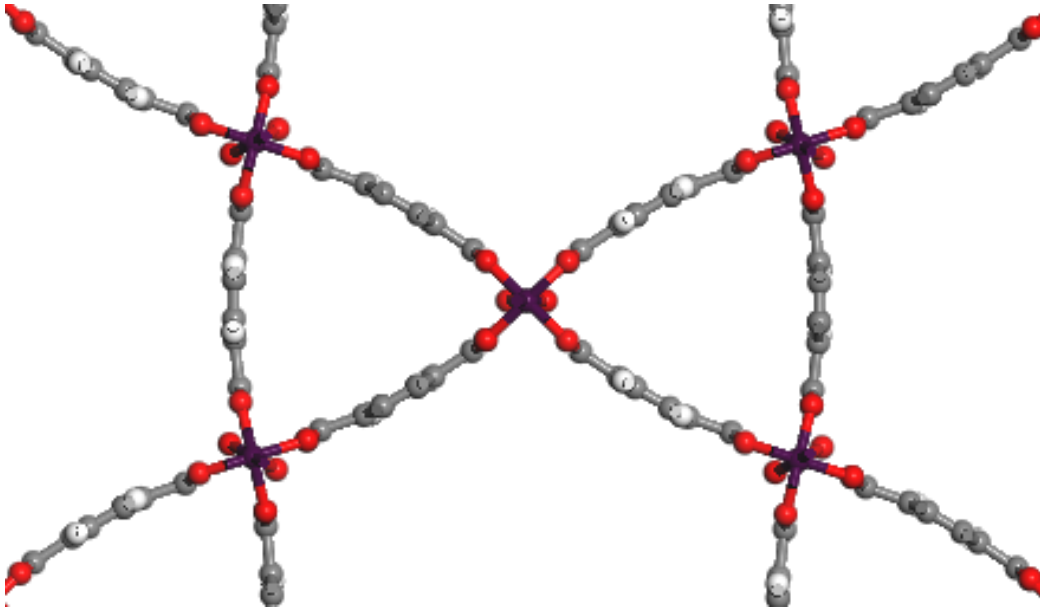
In particular, we are **developing methods to model CO_2 adsorption linked to structural changes** ← requires accurate flexible force fields

Force field = model potential, mathematical expression governing interaction of the atoms:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}}$$

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$$

$$E_{\text{nonbonded}} = E_{\text{electrostatic}} + E_{\text{van der Waals}}$$



Metal Organic Frameworks (MOFs) are organic-inorganic crystalline materials. There are many thousands of MOFs that have been synthesized. Due to the wide variety of possible constituents, the possible design space of MOFs is enormous.

Technical Approach/Project Scope

Task 16: Computational Screening of Sorbents for Direct Air Capture

Fixed-atom Force Fields

- Widely available, used for screenings
- Low accuracy
- Computationally inexpensive
- No flexibility (MOF atoms can't move)

Density Functional Theory (DFT):

- Highly accurate, based on quantum mechanics
- Can describe the flexibility of MOFs
- Computationally time-intensive, cannot get statistical averages that are needed to calculate sorption of gases

Flexible Force Field:

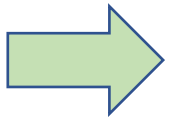
- Computationally inexpensive
- Can be used in atomistic simulations to get statistical averages – e.g. can use to obtain CO_2 sorption on MOFs
- Flexible – can describe movement of MOFs

Technical Approach/Project Scope

Task 16: Computational Screening of Sorbents for Direct Air Capture

Density Functional Theory (DFT) calculations on ~**8000** selected MOFs from the QMOF database

QuickFF to generate flexible force fields (model potential)



NETL-ML-FFF
NETL Machine Learned **Flexible Force Field**

- Obtained by using convolutional neural network on the reference data
- Can be used to calculate CO_2 sorption at low pressure on ***any MOF***
- Force field is general, transferable
- Screen **unlimited** number of MOFs

Progress and Current Status

Task 16: Computational Screening of Sorbents for Direct Air Capture

Proof of Concept Results

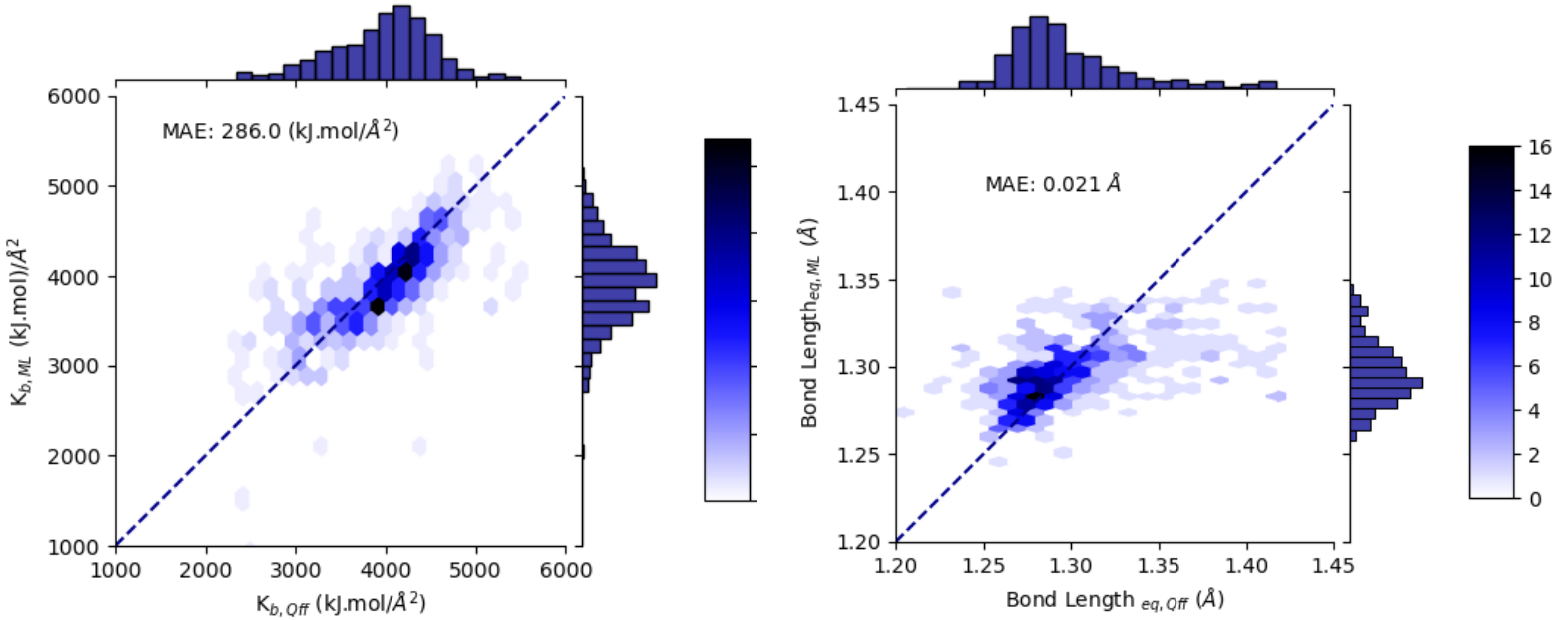
2275 data points, of which 60% were used for training, 20% were used for validation and the model was tested on 20% of the remaining data.

The results in figures are predictions for 0.2*2275 = 455 MOFs.

Convolutional Neural Network implemented in PyTorch

C-O bond (example):

- K_b (Bond Constant)
- Bond length



Project Overview

Task 12: Computational Screening of Novel Membrane Materials

The objectives of this **computational effort** are to use:

- computational database screening
- molecular simulations and machine learning tools

1. To computationally screen or design novel or previously unidentified low technology readiness level (TRL) polymer materials with properties that match or exceed the performance of the team's current best performing polymers.
2. To create a software tool useful to ourselves and other researchers that leverages ML to predict the properties of polymers based on an encoding of the repeat unit

Technology Background

Task 12: Computational Screening of Novel Membrane Materials

Year 1 (EY19):

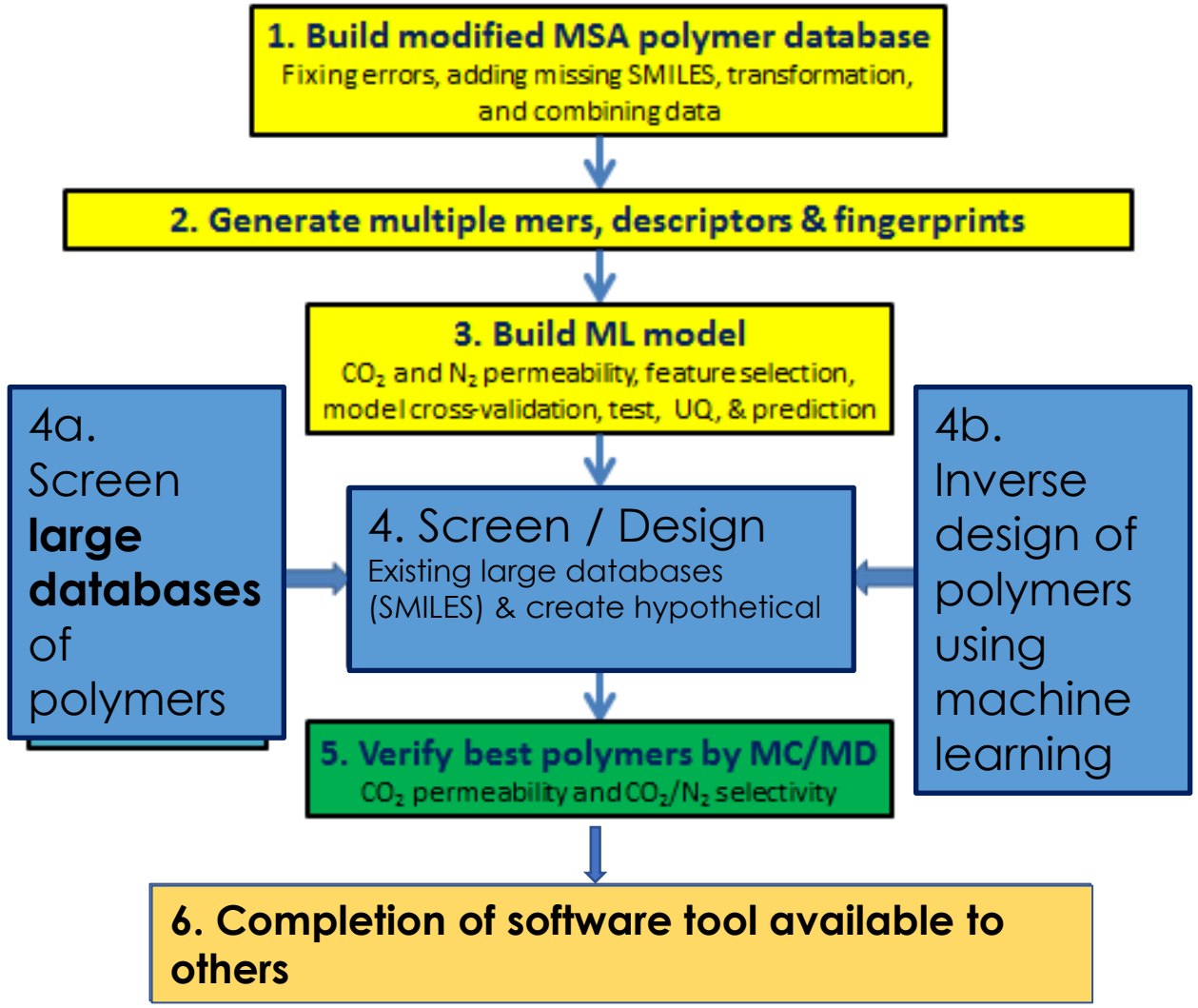
- ✓ Creation of software tools
- ✓ Group contribution methods, functional group database.
- ✓ Collection of dataset of expt. permeation data

Year 2 (EY20):

- ✓ Dataset correction and augmentation, addition of SMILES
- ✓ Creation and refinement of ML model
- ✓ Use of ML model on for screening large collections of polymers (SMILES strings)

Year 3 (EY21):

- ➢ Identification/design of promising polymers
- ➢ MD and MC-based predictions for promising polymer materials
- ➢ Completion of software tool and plan for making tool available to others.



Technical Approach/Project Scope

Task 12: Computational Screening of Novel Membrane Materials

Created a database of experimentally measured gas permeation data.

- ✓ We began this database by importing the Membrane Society of Australasia (**MSA**) database.
- ✓ Collected **original literature sources**, manually searched for data (time-consuming)
- ✓ Corrected/verified/added permeability values for **CO_2, CH_4, N_2, O_2, H_2, He, CO, Ar, D_2** and other gases
- ✓ Added new polymers, particularly highly permeable polymers
- ✓ Encoded simplified molecular-input line-entry system (**SMILES**) strings for repeat units for each polymer.
- ✓ Database is currently up to ~1670 data sets

Fit Machine-Learned models (9 models).

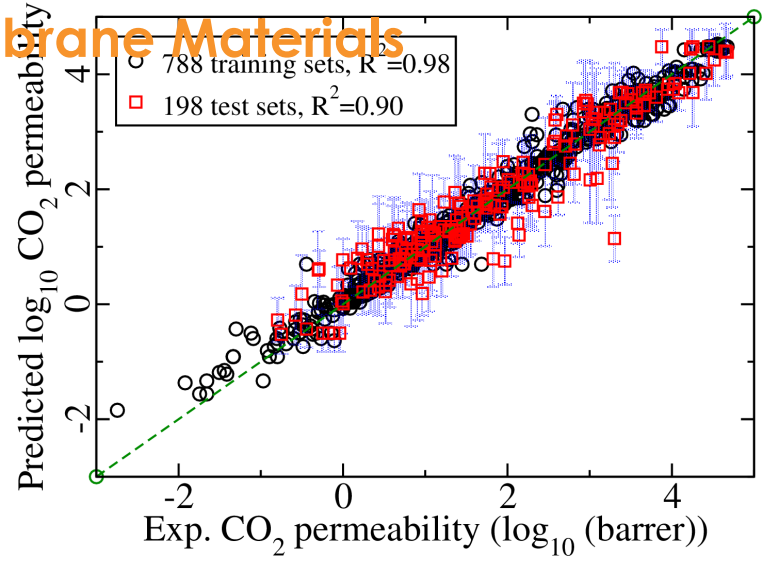
- ✓ From **SMILES**, software creates a polymer chain length as specified by the user (we are using a "10-mer")
- ✓ Software automatically calculates all descriptors
- ✓ 190 2D descriptors (depend on the graph of the molecule structure, node: atom; edge: bond).
- ✓ 11 3D descriptors (depend on xyz coordinates for the atoms).
- ✓ 5 fingerprints (RDKFingerprint / MACCS / AtomPair / Torsion / Morgan)
- ✓ ML: **Gaussian process regressor** used to **create the ML-models**
- ✓ Models were fit to **predict gas permeability** for CO_2, CH_4, N_2, O_2, H_2, He.
- ✓ Models to **predict gas perm-selectivity** for CO_2/N_2, O_2/N_2, H_2/CO_2 .

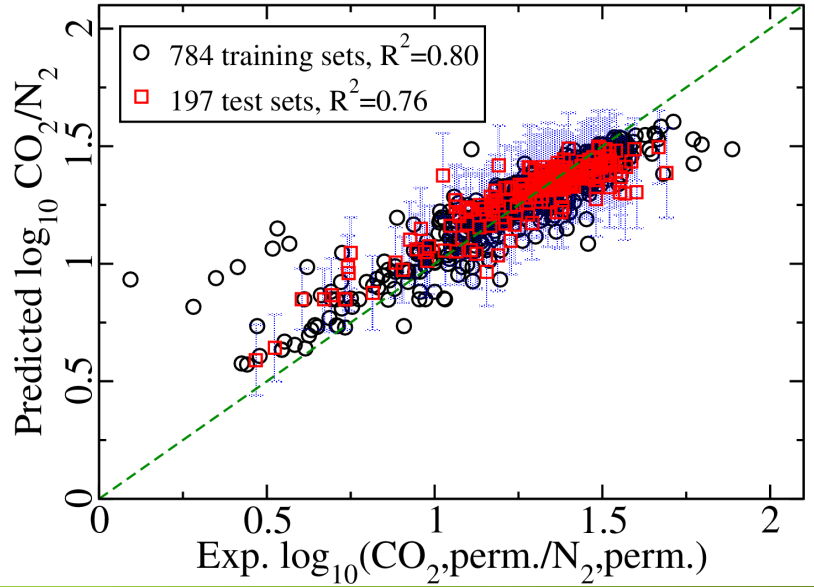
Technical Approach/Project Scope

Task 12: Computational Screening of Novel Membrane Materials

Detailed Results on the ML-models

Gas permeability or permeability selectivity	Total number of data sets	Training data size	Test data set	R^2 for training data set	R^2 for test data set
CO_2	986	788	198	0.98	0.90
N_2	1068	854	214	0.98	0.90
O_2	1065	852	213	0.98	0.91
CH_4	887	709	178	0.98	0.92
H_2	710	568	142	0.98	0.91
He	574	459	115	0.98	0.92
CO_2/N_2	981	784	197	0.80	0.76
O_2/N_2	1064	851	213	0.87	0.79
H_2/CO_2	629	503	126	0.96	0.81



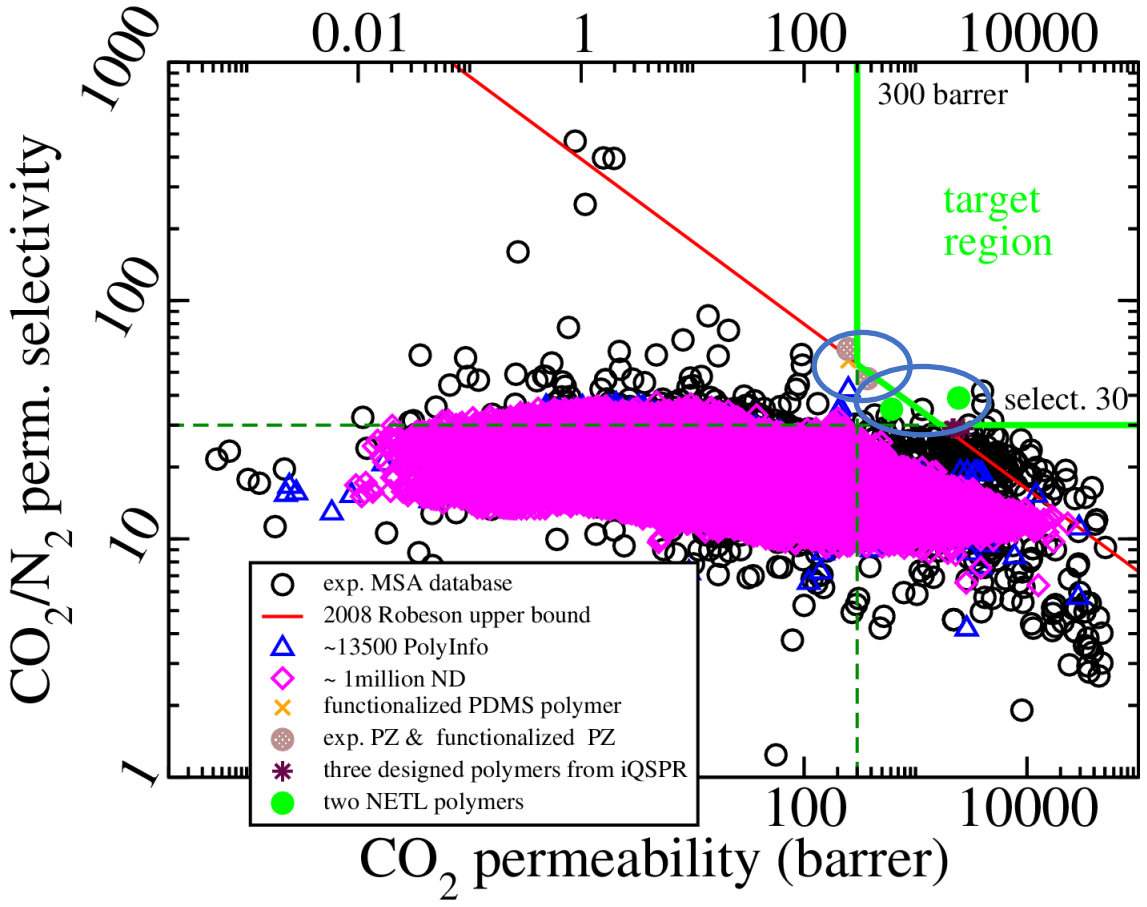


Progress and Current Status

Task 12: Computational Screening of Novel Membrane Materials

We have completed screening of several databases, ~million polymers (PolyInfo, Notre Dame).

- Our objective was to identify a selective layer for a thin film composite:
  - rubbery (not glassy)
  - permeability > 250 Barrer (~2500 GPU if 0.1 μm)
  - selectivity > 30
  - on or above the 2008 Robeson Bound
- A few promising rubbery polymers have been identified on the boundary of our target region
- For comparison, two NETL-developed polymers are labeled; the one in the target region is glassy



PIM-1/MEEP80: 2440 Barrer, α=39 (Sekizkardes J. Mater. Chem. A 2018)
XL MEEP: 610 Barrer, 35 α= (Kusuma et al. ACS App. Mat. Int. 2020)

Progress and Current Status

Task 12: Computational Screening of Novel Membrane Materials

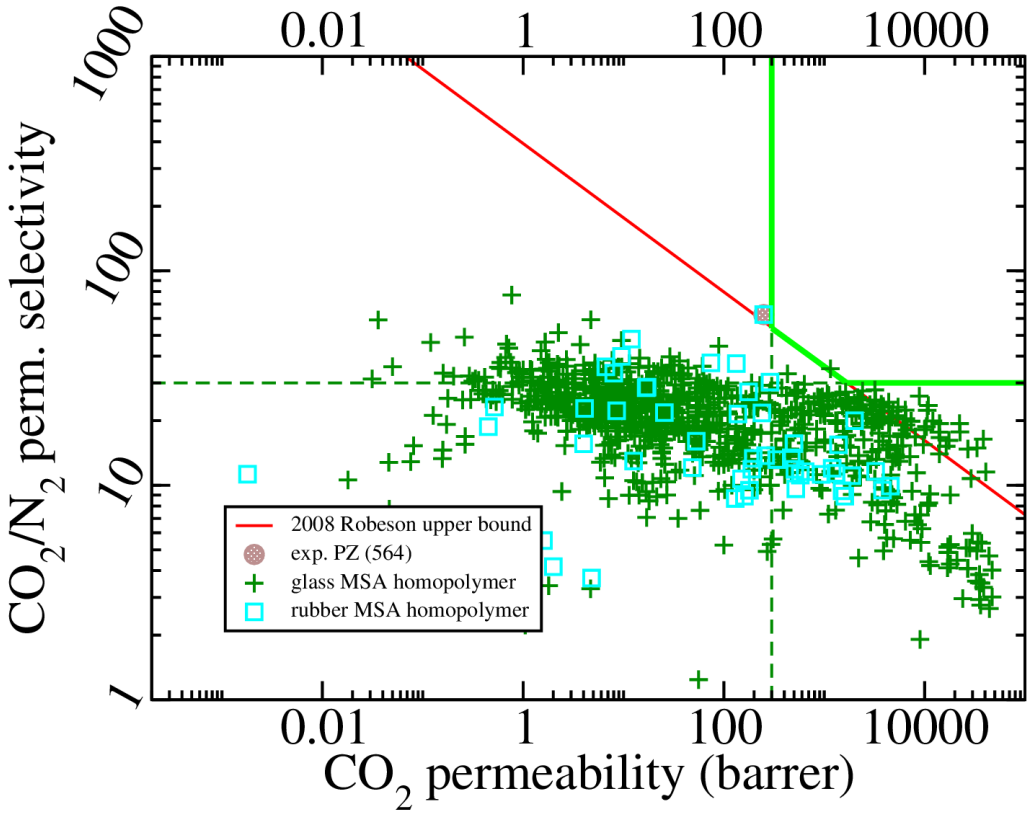
ML model for the prediction of glassy versus rubbery (T_g < 25°C = rubbery)

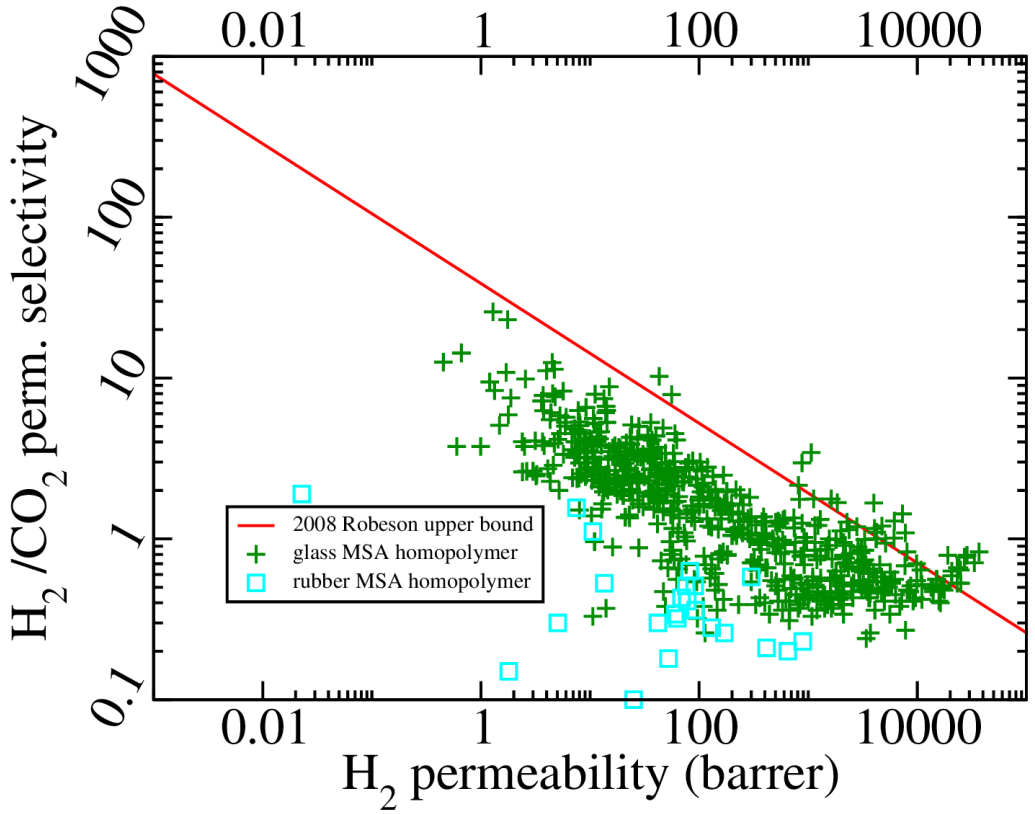
Gradient Boosting Classifier

547 sets of data, 383 training, 164 test

189 2D descriptors

R^2: 0.93 (train) 0.95 (test)



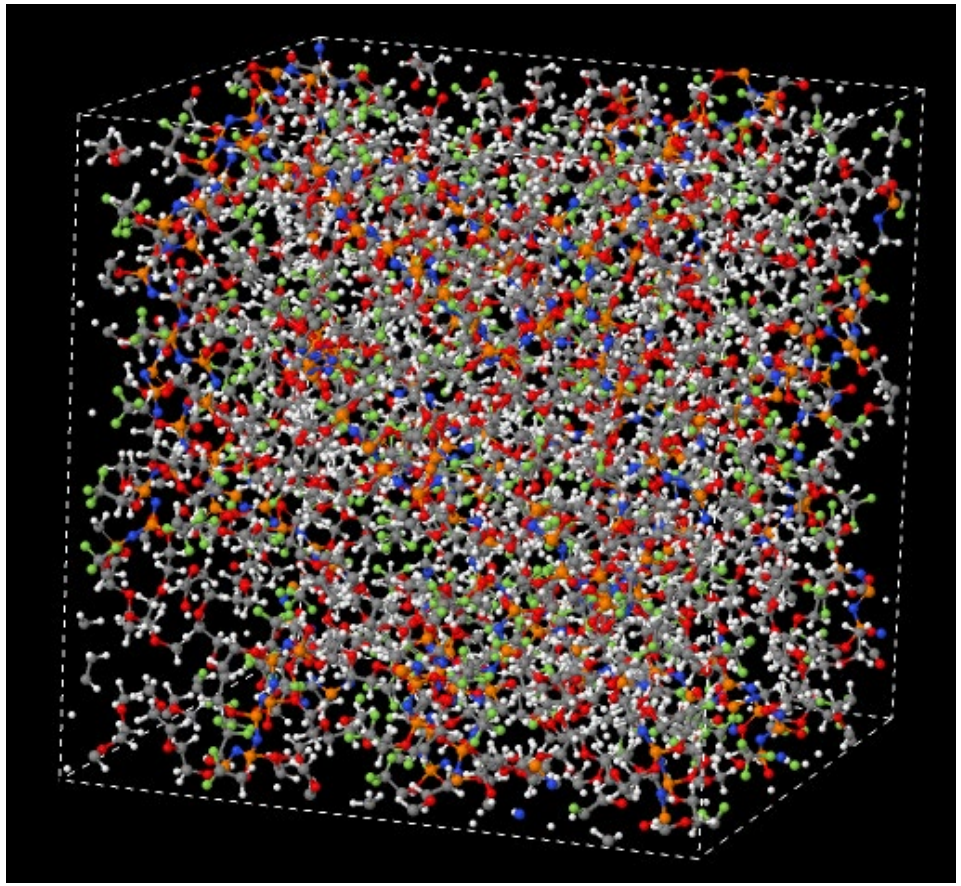


Progress and Current Status

Task 12: Computational Screening of Novel Membrane Materials

iQSPR (inverse quantitative structure–property relationships).

- **Not screening**, we are **designing** new (hypothetical) structures.
- Makes use of the previously-fit ML model
- Designed about 14 polymers with properties near the target region
- We down-selected to 3 rubbery polymers
  - 318 Barrer, $\alpha = 29$
  - 278 Barrer, $\alpha = 29$
  - 265 Barrer, $\alpha = 29$
- MD simulations (results pending) of three polymers to ***confirm***:
  - CO_2 permeability, CO_2/N_2 perm-selectivity
  - Rubbery versus glassy



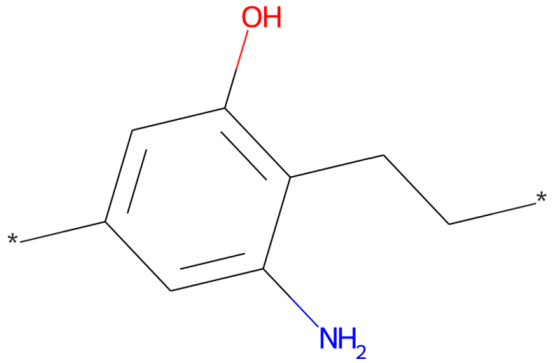
Snapshot from molecular dynamics simulation to confirm ML-predicted properties of an NETL computationally-designed polymer

Progress and Current Status

Task 12: Computational Screening of Novel Membrane Materials

Software release: "**CMLProP**"

- **C**heminformatics **M**achine **L**earning **Pro**perties of **P**olymers
- Github (planned)
  - Freeware – source code download by user
  - Software functional, distribution possible soon
- Webpage:
  - Prototype page is in testing.
  - User inputs SMILES string for their polymer repeat unit – (You can use freeware - ChemSketch)



SMILES string:
Nc1cc(*)cc(O)c1CC*

CMLProP

Welcome to CMLPROP, the Cheminformatics Machine Learning Properties of Polymers.

Currently, CMLPROP can predict gas permeability, permeability selectivity and rubbery/glassy in homo-polymers.

The software tools are developed, maintained, updated, and optimized by the computational materials and carbon capture teams at NETL and Battelle.

Instructions: Input a simplified molecular-input line-entry system (SMILES) for the repeat unit of your polymer.

Input SMILES:
_____

Permeability Results (Barrer):
CO_2: _____
CH_4: _____
N_2: _____
O_2: _____
H_2: _____
He: _____

Perm-selectivity Results:
CO_2/N_2: _____
O_2/N_2: _____
H_2/CO_2: _____

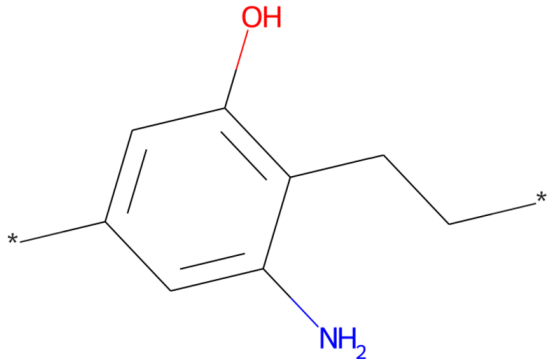
Glass Transition Results: Rubbery vs Glassy:
_____

Progress and Current Status

Task 12: Computational Screening of Novel Membrane Materials

Software release: "**CMLProP**"

- **C**heminformatics **M**achine **L**earning **Pro**perties of **P**olymers
- Github (planned)
  - Freeware – source code download by user
  - Software functional, distribution possible soon
- Webpage:
  - Prototype page is in testing.
  - User inputs SMILES string for their polymer repeat unit – (You can use freeware - ChemSketch)



SMILES string:
Nc1cc(*)cc(O)c1CC*

CMLProP

Welcome to CMLPROP, the Cheminformatics Machine Learning Properties of Polymers.

Currently, CMLPROP can predict gas permeability, permeability selectivity and rubbery/glassy in homo-polymers.

The software tools are developed, maintained, updated, and optimized by the computational materials and carbon capture teams at NETL and Battelle.

Instructions: Input a simplified molecular-input line-entry system (SMILES) for the repeat unit of your polymer.

Input SMILES:
Nc1cc(*)cc(O)c1CC*
Permeability Results (Barrer):
CO_2: 4.31
CH_4: 0.27
N_2: 0.15
O_2: 0.72
H_2: 14.35
He: 6.47

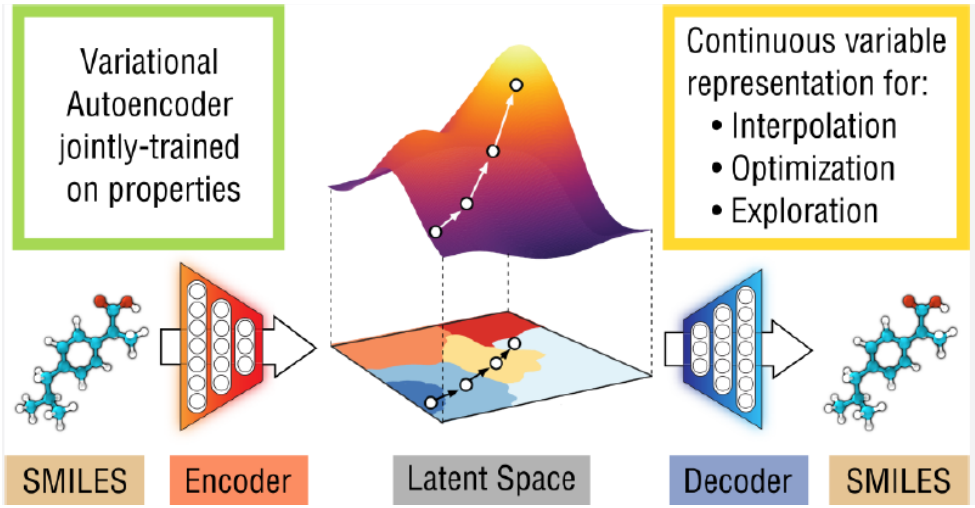
Perm-selectivity Results:
CO_2/N_2: 20.7
O_2/N_2: 4.9
H_2/CO_2: 3.0

Glass Transition Results:
Rubbery vs Glassy:
Glassy

Future Plans

Task 12: Computational Screening of Novel Membrane Materials

- Publication: Polymer Gas Permeation and Rubber/Glass Classification via Machine Learning
- Publication: Inverse Design of Polymers using Machine Learning
- Finalize CHEMLProP software, distribute via github, webpage
- While the iQSPR successfully designed novel polymers, it doesn't extrapolate well into the target region.
- Variational Autoencoder Method for the design of new polymers
  - Coverts the discreet data (of our polymer DB) into a continuous representation.
  - Continuous representation better suited for interpolation, optimization and exploration of polymer space
  - Implementation is in progress



R. Gómez-Bombarelli et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent. Sci. 2018, 4 (2), 268–276

Acknowledgements

The authors would like to warmly acknowledge significant assistance from:

- Dave Hopkinson
- Kevin Resnik
- Ali Sekizkardes
- Victor Kusuma
- Lingxiang Zhu
- Shouliang Yi
- Patrick Muldoon
- Lynn Brickett
- Dan Hancu