AI assisted Search for Atmospheric CO2 Capture

Shivshankar

Abstract

Carbon capture technologies is an important tool for mitigating climate change [42]. In recent years, polymer membrane separation methods have emerged as a promising technology for separating CO2 and other green house gases from the atmosphere. Designing new polymers for such tasks is quite difficult. In this work we look at machine learning based methods to search for new polymer designs optimized for CO2 separation. An ensemble ML models is trained on a large database of molecules to predict permeabilities of CO2/N2 and CO2/O2 pairs. We then use search based optimization to discover new polymers that surpass existing polymer designs. Simulations are then done to verify the predicted performance of the new designs. Overall result suggests that ML based search can be used to discover new polymers optimized for carbon capture.

1 Introduction

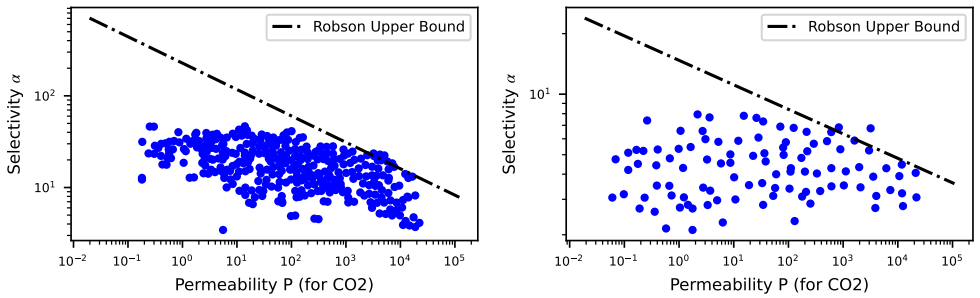
Global warming and associated climate change is one of the most pressing and challenging problems we are facing today [50]. Reduction in anthropogenic CO2 emissions is a critical in our efforts to "prevent dangerous anthropogenic interference with climate systems" [1]. One of the main strategies for mitigating CO2 emissions is carbon capture. [7, 24, 42]. Membrane based separation methods are a major contender for energy efficient and flexible carbon capture technologies [24, 8]

While polymeric membranes have been shown to be effective for seperation for small molecules [4], including CO2 [19, 46], the current efficiency of these polymers is not enough for developing truly cost-efficient and flexible carbon capture systems [9, 31]. One needs high CO2 permeable membranes with atleast moderately high CO2/gas selectivity. Increasing CO2 permeability and selectivity of mem branes are crucial for increasing the overall process throughput, reducing energy costs, and achieving removal of greater amounts of CO2. However, there exists a well-known negative correlation between these properties, which makes increasing both selectivity and permeability difficult [36, 13]. This is often expressed in terms of the Robeson upper bound [36], which represents the frontier of this trade-off.

Contribution Our objective in this work, is to develop a ML-driven approach to tackle the problem of designing polymer membranes focused on CO2 separation from air. As N2 and O2 are the dominant component of the atmosphere (comprising more than 99% of gases); high membrane selectivity for CO2/N2 and CO2/O2 pairs is important for this task; and so we focus on these metrics. From experimental data, we train an ensemble multi-task ML model to predict a polymer's permeability for all three gases - CO2, N2, and O2 as well as its CO2/N2 and CO2/O2 selectivity. Then, we employ a RL based search algorithm to design new polymers and evaluated their performance with respect to the Robeson upper bounds using these machine learning models. Using this framework we were able to identify new polymer membranes that are promising for both CO2/N2 and CO2/O2 separations.

[1] `https://unfccc.int/resource/ccsites/zimbab/conven/text/art02.htm`

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.



(a) CO2/N2 (b) CO2/O2

Figure 1: Double logarithmic plots of selectivity versus CO2 permeability for gas pairs a) CO2/N2 and b) CO2/O2 from a database of existing molecules (e.g. PolyInfo[32],PI1M[27]). The Robeson upper-bound is marked in black line

2 Background and Related Work

Robeson Bound The efficacy of gas seperation via a membrane is determined by two important factors a) permeability i.e. the rate at which a gas can pass through a membrane, and b) selectivity i.e. the ability of the membrane to selectively transport one gas while rejecting others. In the context of gas separation membranes, there is often a tradeoff between permeability and selectivity [36]. In Figure 1, we plot the selectivity and permeability of various polymers, and we can clearly see the general decreasing trend.

While there are physical reasons for existence of this tradeoff [13], the Robeson bound is more an empirical bound than an intrinsic one [37, 40]. Identifying new materials that break this upper bound has driven and continues to drive material research. Since Robeson's initial work, improvement in polymer sciences have pushed outward this 'empirical' Robeson bound [35, 11]. Indeed, for many gas pairs the effective upper bounds was updated in 2019 [11]. In Figure1, we have plotted the empirical Robson bound from Comesaña-Gándara *et al.* [11] which can be considered to be a Pareto frontier of selectivity permeability tradeoff.

ML-aided Polymer Design The accurate and efficient property prediction is essential to the design of polymers in various applications[26]. Traditionally, chemists and polymer scientists have relied on physical simulation to study and predict properties of various molecules. These simulations while accurate, are compute-intensive. On the other hand deep-ML models provide a quick inference method, which has led to them their rising popularity in the field of property prediction [14, 1, 22, 10, 29, 30]. This has been further aided by the availability of multiple large scale datasets [39, 48, 32, 49], that are needed to develop accurate deep-learning models [41, 43, 18, 28]. Recently, with the advent of large language models, large transformer inspired architectures for doing property predictions are also becoming popular [18, 45, 20].

Barnett *et al.* [5] show that machine learning models generalize well to the task of predicting polymer properties to new polymer repeat units. Yang *et al.* [48] show a method to utilize this for interpretable discovery of new polymers. However, these methods do not generate novel candidates, and instead used a large database of possible designs. This simplifies the discovery process but cannot generate either novel candidates or optimized candidates for gas separation task. Secondly, they did not focus specifically on CO2 separation, and indeed their best polymers do not surpass the robeson bound for the CO2/N2 and CO2/O2 pairs important for CO2 separation from air.

3 Method

Now we describe our approach to using ML for aiding polymer discovery for CO2 separation. The basic design is simple and intuitive, and similar versions have been proposed in different contexts such as drug discovery [47], exploring metal-organics [1], virtual screening [48]. Recently Giro *et al.* [15], have proposed a similar framework for aiding discovery of new polymers for carbon capture.

Overview We use a combination of ML-based property prediction, and RL-based search to efficiently and effectively discover new molecules with desired properties, and then validate them with molecular dynamics simulations.

The first step is to train machine learning (ML) models to predict the required properties, such as selectivity and permeability, using existing training sets like PoLyInfo[32] and PI1M [27]. We refer to this model as a Property Prediction Function (PPF) in our work. Once a well-trained and calibrated PPF is available, the next stage is to utilize it in an objective function to discover new molecules. We use a reinforcement learning (RL)-based method that proposes new molecules to be evaluated. PPF is used to estimate the desired properties for every new proposal, and the corresponding objective is computed. The PPF estimates are then provided as a reward to the RL-based search methods, which proceed iteratively to propose newer interesting molecules to evaluate. Finally, once a candidate pool of interesting molecules are found, the predicted properties of a candidate are validated via a standard molecular dynamics simulation.

Features and ML Model We train an ensemble model consisting of various machine learning algorithms, such as elastic nets, boosted random forests (Xgboost), feed-forward neural networks (FFN), graph convolutional networks (pGCN [33]) and Transformers (TransPolymer [45]) For the classical models, we used chemically relevant descriptors including bond angles, ring type and substructures, RDKit chemical descriptors [21] and Morgan fingerprints [38], among others. These features are widely used in the literature and are known to be important for predicting various properties of molecules [6]. In addition, we used learned embeddings for substructures and important atoms in the neural network models. The larger deep-learning models, such as Transformers and GCNs, were trained using both unsupervised losses and multi-task regression losses. To determine the best ensemble, we conducted a grid search to optimize the hyperparameters and chose the best ensemble based on cross-validation results.

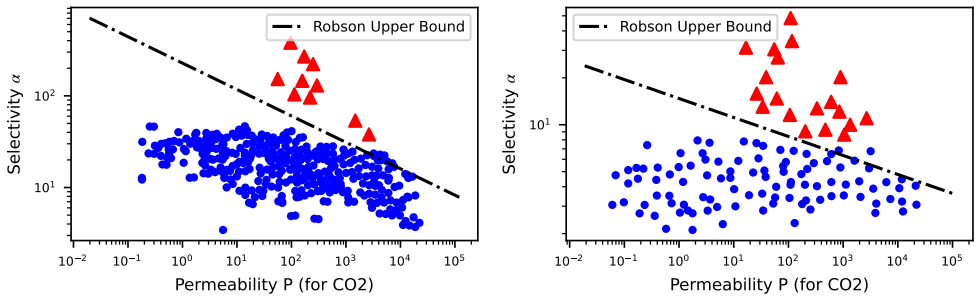
Environment and Score/Reward Function The score function is the reward provided for the agent to maximize. Ideally we want a polymer which has high CO2 permeability while also high selectivity for CO2 compared to O2 and N2. A natural metric is the product of the pair selectivities and CO2 permeability [48]. We use the estimate of selectivity and permeability produced by the PPF.

State and Policy Network Each state vector, represents a specific intermediate molecular configuration. The policy network takes the current graph/molecule as state and chooses an action which is applied to the current molecule to modify it. Each policy action is specified via three steps: I), the identification of the specific site where a new fragment is to be introduced; II, the selection of the new fragment itself; and III), the precise bonding site on the new fragment.

We use three different multi-layered perceptrons M_1, M_2, M_3 to parameterize each individual decision of the three step policy action. Following the approach of Hwang *et al.* [17], Yang *et al.* [47], we represent the current state molecule as an undirected graph, using a graph convolutional network to obtain node embeddings that are then aggregated to produce a graph embedding. In the first step, the M_1 takes as input node embeddings of each attachment site on the molecule and a GCN (H) based graph embedding of the entire molecule, and samples a site to expand the current fragment. Next, a different GCN (E) is used to compute graph embedding of the molecule with contextualized information relevant for the chosen attachment site. This embedding is combined with H based embeddings of candidate extension fragments to predict a distribution over the space of valid fragments to add. Finally, both these decision are fed into M_3 next step, i.e. the joining site on the new fragment.

4 Results

Molecular simulation on the discovered molecules suggest that these are effective units for CO2 separation. In Figure 2 we present in red few of the most promising candidates discovered by our method overlayed against the Robeson bound figures. In general, the ensemble model predictions well approximated the predicted properties given by molecular simulations, even when molecules are optimized using the predicted value. While the model uncertainty is greater as the scoring networks is used to extrapolate to unseen (and potentially out of distribution molecules), overall the results seem validated by molecular simulations. More analysis of the results is present in the Appendix.



(a) CO2/N2 (b) CO2/O2

Figure 2: Double log plots, overlaying experimental data from existing databases (blue points), the Robeson bound (black) and a sample of novel molecules found by our method (red). The figures show results for the gas pairs a) CO2/N2 and b) CO2/O2. It is clear that our method generates molecules that surpass both upper bounds

5 Conclusion

In this work we present a ML driven framework for finding novel polymers for CO2 separation. The framework involves using an ensemble of ML models to predict relevant properties like permeability and sensitivity. This ML model is then used to compute a reward function for an RL based search method to produce optimized polymer molecules. The proposed method is able to find close relatives of recent high performance polymers. We verify with molecular simulation the predicted performance of the proposed polymers, and find them to be reasonably accurate. Overall the framework shows promise as an effective means of discovering novel polymers for CO2 separation from the atmosphere.

Limitations Our framework while easily adjustable to incorporate other important metrics like synthetic accessibility, does not take them into consideration. Even ignoring questions of accessibility, there are other factors important for gas separation that are ignored. A polymer's gas filtration performance also depends heavily on factors like morphology of the polymer, operating temperature etc [34] which have been ignored here. Moreover, we have focused on separating CO2 from N2 and O2, and hence on atmospheric separation. Polymers for carbon separation at an industrial or power plant are different, as they will function under different operating conditions and dealing with post-combustion gases. Furthermore, even for molecules discovered this way, extensive simulation via different methods and experimental analysis is required for stronger conclusions.

References

[1] Cigdem Altintas, Omer Faruk Altundal, Seda Keskin, and Ramazan Yildirim. Machine learning meets with metal organic frameworks for gas storage and separation. *Journal of Chemical Information and Modeling*, 61(5):2131–2146, 2021.

[2] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016.

[3] Youn-Sang Bae, Jian Liu, Christopher E Wilmer, Hahnbi Sun, Allison N Dickey, Min Bum Kim, Annabelle I Benin, Richard R Willis, Dushyant Barpaga, M Douglas LeVan, et al. The effect of pyridine modification of ni–dobdc on co 2 capture under humid conditions. *Chemical Communications*, 50(25):3296–3298, 2014.

[4] Richard W Baker and Bee Ting Low. Gas separation membrane materials: a perspective. *Macromolecules*, 47(20):6999–7013, 2014.

[5] J Wesley Barnett, Connor R Bilchak, Yiwen Wang, Brian C Benicewicz, Laura A Murdock, Tristan Bereau, and Sanat K Kumar. Designing exceptional gas-separation polymer membranes using machine learning. *Science advances*, 6(20):eaaz4301, 2020.

[6] Jozef Bicerano. *Computational modeling of polymers*, volume 25. CRC press, 1992.

[7] S Bode and M Jung. Carbon dioxide capture and storage (ccs)-liability for nonperformance under the unfccc. *HWWA discussion paper*, (25):1, 2005.

[8] Carbon capture. https://www.c2es.org/content/carbon-capture/.

[9] Guining Chen, Tianlei Wang, Guangru Zhang, Gongping Liu, and Wanqin Jin. Membrane materials targeting carbon capture and utilization. *Advanced Membranes*, 2:100025, 2022.

[10] Kamal Choudhary, Taner Yildirim, Daniel W Siderius, A Gilad Kusne, Austin McDannald, and Diana L Ortiz-Montalvo. Graph neural network predictions of metal organic framework co2 adsorption properties. *Computational Materials Science*, 210:111388, 2022.

[11] Bibiana Comesaña-Gándara, Jie Chen, C Grazia Bezzu, Mariolino Carta, Ian Rose, Maria-Chiara Ferrari, Elisa Esposito, Alessio Fuoco, Johannes C Jansen, and Neil B McKeown. Redefining the robeson upper bounds for co 2/ch 4 and co 2/n 2 separations using a series of ultrapermeable benzotriptycene-based polymers of intrinsic microporosity. *Energy & Environmental Science*, 12(9):2733–2740, 2019.

[12] Naiying Du, Gilles P Robertson, Jingshe Song, Ingo Pinnau, Sylvie Thomas, and Michael D Guiver. Polymers of intrinsic microporosity containing trifluoromethyl and phenylsulfone groups as materials for membrane gas separation. *Macromolecules*, 41(24):9656–9662, 2008.

[13] Benny D Freeman. Basis of permeability/selectivity tradeoff relations in polymeric gas separation membranes. *Macromolecules*, 32(2):375–380, 1999.

[14] Victor Fung, Jiaxin Zhang, Eric Juarez, and Bobby G Sumpter. Benchmarking graph neural networks for materials chemistry. *npj Computational Materials*, 7(1):1–8, 2021.

[15] Ronaldo Giro, Hsianghan Hsu, Akihiro Kishimoto, Toshiyuki Hama, Rodrigo F Neumann, Binquan Luan, Seiji Takeda, Lisa Hamada, and Mathias B Steiner. Ai powered, automated discovery of polymer membranes for carbon capture. *npj Computational Materials*, 9(1):133, 2023.

[16] Yuan He, Francesco M Benedetti, Sharon Lin, Chao Liu, Yanchuan Zhao, Hong-Zhou Ye, Troy Van Voorhis, M Grazia De Angelis, Timothy M Swager, and Zachary P Smith. Polymers with side chain porosity for ultrapermeable and plasticization resistant materials for gas separations. *Advanced Materials*, 31(21):1807871, 2019.

[17] Doyeong Hwang, Soojung Yang, Yongchan Kwon, Kyung Hoon Lee, Grace Lee, Hanseok Jo, Seyeol Yoon, and Seongok Ryu. Comprehensive study on molecular supervised learning with graph neural networks. *Journal of Chemical Information and Modeling*, 60(12):5936–5945, 2020.

[18] Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3(1):015022, 2022.

[19] M Kawakami, H Iwanaga, Y Hara, M Iwamoto, and S Kagawa. Gas permeabilities of cellulose nitrate/poly (ethylene glycol) blend membranes. *Journal of Applied Polymer Science*, 27(7):2387–2393, 1982.

[20] Christopher Kuenneth and Rampi Ramprasad. polybert: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nature Communications*, 14(1):4099, 2023.

[21] Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8, 2013.

[22] Sangwon Lee, Baekjun Kim, Hyun Cho, Hooseung Lee, Sarah Yunmi Lee, Eun Seon Cho, and Jihan Kim. Computational screening of trillions of metal–organic frameworks for high-performance methane storage. *ACS Applied Materials & Interfaces*, 13(20):23647–23654, 2021.

[23] Jan Leike, Miljan Martic, Victoria Krakovna, Pedro A Ortega, Tom Everitt, Andrew Lefrancq, Laurent Orseau, and Shane Legg. Ai safety gridworlds. *arXiv preprint arXiv:1711.09883*, 2017.

[24] Dennis YC Leung, Giorgio Caramanna, and M Mercedes Maroto-Valer. An overview of current status of carbon dioxide capture and storage technologies. *Renewable and sustainable energy reviews*, 39:426–443, 2014.

[25] Xiaoyan Luo, Yan Guo, Fang Ding, Hongqing Zhao, Guokai Cui, Haoran Li, and Congmin Wang. Significant improvements in co2 capture by pyridine-containing anion-functionalized ionic liquids through multiple-site cooperative interactions. *Angewandte Chemie*, 126(27):7173–7177, 2014.

[26] Hang Luo, Sheng Chen, Lihong Liu, Xuefan Zhou, Chao Ma, Weiwei Liu, and Dou Zhang. Core–shell nanostructure design in polymer nanocomposite capacitors for energy storage applications. *ACS Sustainable Chemistry & Engineering*, 7(3):3145–3153, 2018.

[27] Ruimin Ma and Tengfei Luo. Pi1m: a benchmark database for polymer informatics. *Journal of Chemical Information and Modeling*, 60(10):4684–4690, 2020.

[28] Rishikesh Magar, Yuyang Wang, and Amir Barati Farimani. Crystal twins: Self-supervised learning for crystalline material property prediction. *arXiv preprint arXiv:2205.01893*, 2022.

[29] Peyman Z Moghadam, Sven MJ Rogge, Aurelia Li, Chun-Man Chow, Jelle Wieme, Noushin Moharrami, Marta Aragones-Anglada, Gareth Conduit, Diego A Gomez-Gualdron, Veronique Van Speybroeck, et al. Structure-mechanical stability relations of metal-organic frameworks via machine learning. *Matter*, 1(1):219–234, 2019.

[30] Aditya Nandy, Chenru Duan, and Heather J Kulik. Using machine learning and data mining to leverage community knowledge for the engineering of stable metal–organic frameworks. *Journal of the American Chemical Society*, 143(42):17535–17547, 2021.

[31] Nadia Norahim, Pacharaporn Yaisanga, Kajornsak Faungnawakij, Tawatchai Charinpanitkul, and Chalida Klaysom. Recent membrane developments for co2 separation and capture. *Chemical Engineering & Technology*, 41(2):211–223, 2018.

[32] Shingo Otsuka, Isao Kuwajima, Junko Hosoya, Yibin Xu, and Masayoshi Yamazaki. Polyinfo: Polymer database for polymeric materials design. In *2011 International Conference on Emerging Intelligent Data and Web Technologies*, pages 22–29. IEEE, 2011.

[33] Jaehong Park, Youngseon Shim, Franklin Lee, Aravind Rammohan, Sushmit Goyal, Munbo Shim, Changwook Jeong, and Dae Sin Kim. Prediction and interpretation of polymer properties using the graph convolutional network. *ACS Polymers Au*, 2022.

[34] Clem E Powell and Greg G Qiao. Polymeric co2/n2 gas separation membranes for the capture of carbon dioxide from power plant flue gases. *Journal of Membrane Science*, 279(1-2):1–49, 2006.

[35] Lloyd M Robeson, Qiang Liu, Benny D Freeman, and Donald R Paul. Comparison of transport properties of rubbery and glassy polymers and the relevance to the upper bound relationship. *Journal of Membrane Science*, 476:421–431, 2015.

[36] Lloyd M Robeson. Correlation of separation factor versus permeability for polymeric membranes. *Journal of membrane science*, 62(2):165–185, 1991.

[37] Lloyd M Robeson. The upper bound revisited. *Journal of membrane science*, 320(1-2):390–400, 2008.

[38] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

[39] Andrew S Rosen, Shaelyn M Iyer, Debmalya Ray, Zhenpeng Yao, Alan Aspuru-Guzik, Laura Gagliardi, Justin M Notestein, and Randall Q Snurr. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, 2021.

[40] Marius Sandru, Eugenia M Sandru, Wade F Ingram, Jing Deng, Per M Stenstad, Liyuan Deng, and Richard J Spontak. An integrated materials approach to ultrapermeable and ultraselective co2 polymer membranes. *Science*, 376(6588):90–94, 2022.

[41] Philippe Schwaller, Theophile Gaudin, David Lanyi, Costas Bekas, and Teodoro Laino. "found in translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chemical science*, 9(28):6091–6098, 2018.

[42] Jim Skea, Priyadarshi Shukla, and Şiir Kılkış. Climate change 2022: mitigation of climate change, 2022.

[43] Sun-Ting Tsai, En-Jui Kuo, and Pratyush Tiwary. Learning molecular dynamics with simple language model built upon long short-term memory neural network. *Nature communications*, 11(1):1–11, 2020.

[44] Aristofanis Vollas, Thanasis Chouliaras, Valadoula Deimede, Theophilos Ioannides, and Joannis Kallitsis. New pyridinium type poly (ionic liquids) as membranes for co2 separation. *Polymers*, 10(8):912, 2018.

[45] Changwen Xu, Yuyang Wang, and Amir Barati Farimani. Transpolymer: a transformer-based language model for polymer property predictions. *arXiv preprint arXiv:2209.01307*, 2022.

[46] Baolong Xue, Xianwu Li, Longcheng Gao, Min Gao, Yao Wang, and Lei Jiang. Co 2-selective free-standing membrane by self-assembly of a uv-crosslinkable diblock copolymer. *Journal of Materials Chemistry*, 22(21):10918–10923, 2012.

[47] Soojung Yang, Doyeong Hwang, Seul Lee, Seongok Ryu, and Sung Ju Hwang. Hit and lead discovery with explorative rl and fragment-based molecule generation, 2021.

[48] Jason Yang, Lei Tao, Jinlong He, Jeffrey R McCutcheon, and Ying Li. Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Science Advances*, 8(29):eabn9545, 2022.

[49] Qi Yuan, Mariagiulia Longo, Aaron W Thornton, Neil B McKeown, Bibiana Comesana-Gandara, Johannes C Jansen, and Kim E Jelfs. Imputation of missing gas permeability data for polymer membranes using machine learning. *Journal of membrane science*, 627:119207, 2021.

[50] Zhu Zhongming, Lu Linong, Yao Xiaona, Liu Wei, et al. Ar6 synthesis report: Climate change 2022. 2022.

A Analysis of Results

We first look at common substructures observed in polymers discovered by our methods. Some of these are reported in Figure 3. Additionally we highlight the fragments which are also common in the 100 most high scoring existing polymers from our dataset. We find that combinations of arenes and azaarenes, are common in both these sets of molecules. Additionally, a less common but promising candidate discovered by our method is pyridine-3,5-diyl (depicted in the top-middle of Figure 3). Studies have shown that pyridine-based poly-ethers are effective for developing CO2-separating polymers [25, 3]. In fact, our procedure has "discovered" a molecule that is related to one presented in Vollas *et al.* [44], but was not present in the training or validation data used in the PPF.

Furthermore, our procedure has identified other useful and common substructures, such as polycyclic backbones and two fluoromethyl chains. Trifluoromethyl side chains are known to cause steric frustration and restricted mobility in polymer chains [12]. This leads to high fractional free volume, which favors increased gas permeability [16]. In addition, our discovered polymers contain a larger number of aliphatic rings, (which also increase steric hindrance), and aromatic oxygen and nitrogen (polar atoms that tend to increase solubility and selectivity). Similar findings were reported in Yang *et al.* [48]. Overall, these results suggest that our method has successfully discovered polymers with promising structural features that are known to improve gas separation properties.

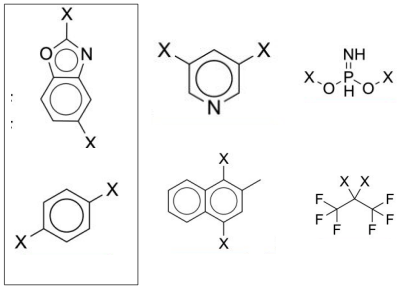
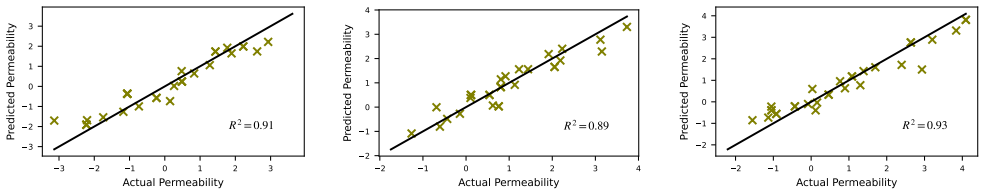


Figure 3: Most common fragments for the best performing discovered polymers. The first two fragments also occurs with high frequency in best known polymers. This suggests our procedure is picking on these fragments to build better polymers

RL and search procedures have the potential to exploit the biases and gaps in the reward function [23, 2]. Therefore, solely relying on the test data performance of the trained models is not sufficient, and it is critical to validate whether our reward function generalizes to newer polymers discovered by our procedure. To this end, we conducted a computer simulation on a sample of 100 molecules discovered by our search algorithm and calculated their permeability for N2, O2, and CO2 gases, respectively. In Figure 4, we present the predicted and simulation-based permeabilities for these polymers. The plot shows a generally good agreement between the two values with a high R^2 value. However, at the very high and low values of permeability, the model tends to be more optimistic than what simulations suggest, indicating the presence of some overfitting in the ML model and/or reward-hacking by the RL search.



(a) N2 (b) O2 (c) CO2

Figure 4: Comparison of the prediction performance of the ensemble ML model trained to predict permeabilities for a) N2 , b) O2 and c) CO2 on generated polymers where we take molecular simulation results as the actual permeability