A Self-Improvable Polymer Discovery Framework Based on Conditional Generative Model

Xiangyun Lei[†1], Weike Ye[†1], Zhenze Yang[1,2], Daniel Schweigert[1], Ha-Kyung Kwon[1], Arash Khajeh[1†*]
[1]Toyota Research Institute,
4440 El Camino Real, Los Altos, California 94022, United States of America.
[2]Department of Materials Science and Engineering,
Massachusetts Institute of Technology,
77 Massachusetts Ave., Cambridge, Massachusetts 02139, United States of America
arash.khajeh@tri.global

Abstract

In this work, we introduce a polymer discovery platform designed to identify polymers with tailored properties efficiently, exemplified through the discovery of high-performance polymer electrolytes. The platform integrates three core components: a conditioned generative model, validation modules, and a feedback mechanism, creating a self-improving system for material innovation. To demonstrate the efficacy of this platform, it is used to identify polymer electrolyte materials with high ionic conductivity. A simple conditional generative model, based on the minGPT architecture, can effectively generate candidate polymers that exhibit a mean ionic conductivity that is significantly greater than those in the original training set. This approach, coupled with molecular dynamics simulations for validation and a specifically designed acquisition mechanism, allows the platform to refine its output iteratively. Notably, after the first iteration, we observed an increase in both the mean and the lower bound of the ionic conductivity of the new polymer candidates. The platform's effectiveness is underscored by the identification of 19 polymer repeating units, each displaying a computed ionic conductivity surpassing that of Polyethylene Oxide (PEO). The discovery of these polymers validates the platform's efficacy in identifying potential polymer materials. Acknowledging current limitations, future work will focus on enhancing modeling techniques, validation processes, and acquisition strategies, aiming for broader applicability in polymer science and machine learning.

***K*eywords** Polymer Electrolyte · Generative Model · Conditioned Generation

1 Introduction

The domain of polymer science is pivotal in shaping advancements across diverse technological arenas. Polymers, with their extraordinary adaptability and customizability, cater to a wide range of applications, spanning from biodegradable materials and high-performance aerospace composites to conducting elements in electronic devices and smart materials in sensor technologies. Notably, polymer electrolytes also play a crucial role, particularly in the field of energy storage [1, 2, 3, 4], exemplifying the versatility of polymers. These applications highlight the immense potential of polymers to be engineered for specific functionalities, such as biocompatibility in medical devices or environmental resilience in various coatings.

Identifying polymers with the optimal blend of properties for specific applications, including polymer electrolytes for energy storage, is a significant scientific challenge. The complexity of polymer structures, combined with the necessity to balance multiple properties like mechanical strength, electrical conductivity, and thermal stability, makes the discovery process highly intricate. Traditional methods, though foundational, are often limited in their ability to rapidly identify innovative materials, including advanced polymer electrolytes for next-generation energy storage solutions. This limitation points to the need for more efficient and comprehensive approaches to polymer discovery.

Machine learning, particularly in the realm of generative modeling, presents a transformative approach to this challenge. Generative models in machine learning have shown promise in various domains, including material science, by enabling the exploration of vast chemical spaces with unprecedented efficiency. These models can quickly navigate the intricacies of polymer chemistry, suggesting novel and plausible compositions and structures for investigation, thereby streamlining the discovery process.

Within this realm, conditioned generative modeling presents a particularly relevant technique. By training models on specific conditions or properties, it becomes possible to generate content that meets predetermined criteria. In the current landscape, while conditioned generative models specifically for polymers are still emerging, the concept of integrating machine learning with material science to tailor polymer properties is gaining traction. Our work contributes to this field by introducing a comprehensive polymer discovery platform that leverages the principles of conditioned generative modeling. This platform is not limited to merely suggesting potential polymer candidates but is designed to iteratively improve and refine its suggestions based on continuous feedback and validation. Such a self-improving system embodies a significant leap from traditional methods, offering a more holistic and efficient pathway to polymer material innovation [5, 6, 7].

Specifically, this paper demonstrates the application of our discovery platform in the realm of polymer electrolytes for energy storage technologies. We focus on polymer electrolytes due to their crucial role in the efficiency and safety of devices such as lithium-ion batteries and supercapacitors. Our platform successfully identified polymer electrolytes with ion conductivities superior to the current benchmark, Polyethylene Oxide (PEO), as validated by molecular dynamics (MD) simulations. This achievement not only highlights the potential of our platform in accelerating the discovery of high-performance polymer electrolytes but also sets a precedent for its application across various polymer-based technologies.

2 Platform

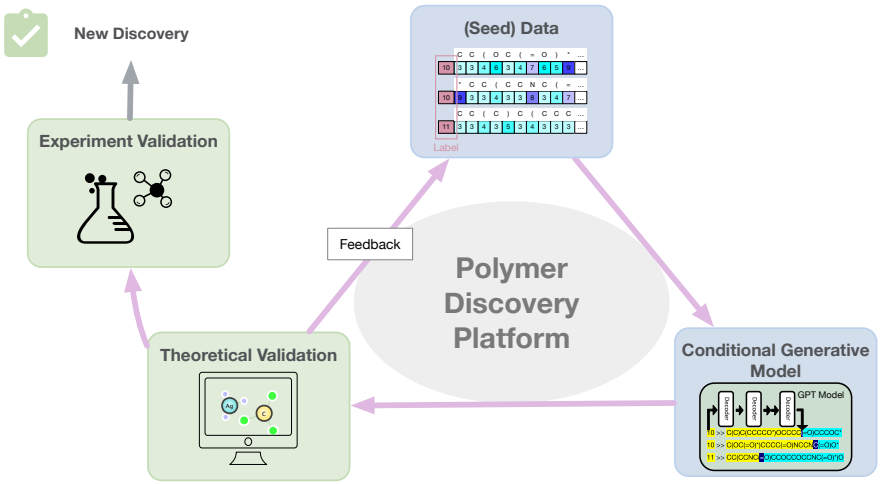

Figure 1: Schematic illustration of the platform.

Here, we present a platform structured around an iterative, self-sustaining, and self-improving workflow, comprising three essential components: a conditional generative model, a validation module, and a feedback mechanism. This integrated system allows for continuous refinement and evolution of the discovery process, which we term a "discovery campaign."

The conditioned generative model, the heart of this platform, is tasked with proposing potential polymer candidates. Tailored to incorporate specific target properties, this module is responsible for generating polymers, either by constructing repeating units, oligomers, polymer chains, or 3D structures. For this proof of concept, our focus is on generating the 2D representation of repeat units of polymers. Several aspects heavily influence the performance of the generative model: the (seed) data, the model architecture and hyperparameters, and the training strategies. Different from traditional regression models where a numerical loss is clearly defined, the generative tasks are more ambiguous to evaluate and often require domain knowledge. The process of formulating a comprehensive and domain-relevant evaluation schema and performing the benchmarking across a series of model architectures and training strategies presents its own set of challenges. Our separate research paper [8] addresses these challenges, offering a systematic approach for

benchmarking different generative models in different tasks and proposing a comprehensive set of evaluation metrics for the task of generating novel polymers.

Once a batch of polymers is proposed by the generative model, the validation module takes over. This component is responsible for assessing the target properties of the proposed polymers, employing both simulation and experimental validations. For theoretical validation, in the current study, we rely on molecular dynamics (MD) simulations (see Experiment Setup for details). Experimental validation serves as the definitive confirmation for our newly discovered polymers. While this crucial phase falls beyond the scope of the current study, we intend to conduct experimental tests on these candidates in a subsequent phase of our research.

Establishing a feedback mechanism is pivotal to allowing active learning and continuous self-improvement of the model. At the end of each campaign iteration, all validated results are recorded in a database, and strategically sampled for enriching the training data (please see the details of sampling in 3). The model is then retrained to the new data to become increasingly adept at targeting the desired polymers. For future improvements, the uncertainty of the generative model ought to be quantified, and sophisticated acquisition strategies to balance exploration and exploitation are to be tested and implemented.

The initialization and deployment of the platform are crucial steps of the discovery campaign. The quality and distribution of the seed data are paramount to the campaign's success, and they represent the main responsibility of the user. With the appropriate seed data and clearly defined target properties, the platform is designed to be capable of operating autonomously, minimizing the need for further human intervention. This autonomous nature highlights the platform's potential for high-throughput, efficient discovery campaigns and opening new horizons in the field of polymer science.

3 Experiment Setup

3.1 Scientific task

To demonstrate the usage of the platform, we deployed it to identify polymer electrolytes with high ionic conductivity, as advancements in energy storage technologies are increasingly contingent on the development of such materials. Despite significant progress, finding materials that surpass the performance benchmarks set by current standards, such as Polyethylene Oxide (PEO), remains a challenge. PEO, while widely used, falls short of meeting the growing demands for efficiency and stability under varied operational conditions. This limitation is particularly pronounced in applications requiring high charge-discharge cycles and under extreme temperature variations. Our endeavor aims to address this gap by identifying polymer electrolytes with superior ionic conductivities. This quest is not only about surpassing PEO's performance but also about unlocking new possibilities for safer, more durable, and higher-capacity energy storage solutions.

3.2 Conditional generation

The core of this demonstration revolves around a conditional generative model based on the minGPT [9] architecture. When trained, this model architecture is capable of completing strings given prompts, just like the popular large language models. The specific model used utilizes the gpt-nano architecture with approximately 120,000 trainable parameters. It is trained on and generates the Simplified Molecular Input Line Entry System (SMILES) [10] code of polymers' repeating units. Although relatively simple in its design, it is powerful in its application. To direct the model towards polymers with high ionic conductivity, we implemented a method to incorporate the properties of the input during training. This involved the modification of tokenized SMILES strings of known polymer electrolytes, prefixed by their ionic conductivity classes. Specifically, given the range (0.007-0.506 mS/cm) and distribution (mean= 0.062 mS/c, std= 0.036 mS/c) of ionic conductivity in our dataset, we assigned different class labels to high-conductive (top 5%) and low-conductive (lower 95%) polymers and used this class as the leading digit in the input to the model. For example, a polymer with an ionic conductivity of 0.18 (mS/cm) is labeled with class 1 (high conductivity), while another polymer with an ionic conductivity of 0.06 (mS/cm) is labeled with 0 (low conductivity). As the data set evolves over time, the allocation of polymers to the respective high- and low-conductive groups will change accordingly. Additionally, to maintain the importance of the property class in comparison to the lengthy SMILES, and to ensure the model can effectively guide us toward desirable structures, we replicate the property class five times, converting the desired class tokens to "11111". Inspired by the simple design of PEO with a very short repeat unit (OCC) and high ionic conductivity (1.15 mS/cm, at 353K, $Li^{+}.TFSI^{-}$ molality= 1.5 mol/kg), during the iterative polymer generation loops, the model is also biased towards generating small repeating units with SMILES strings containing 10 or fewer tokens. This approach, albeit unrefined, proved crucial for the model's ability to generate high-conductivity polymers. The reason for the effectiveness of conditioning on short repeat units can be the short distance between negatively

charged atoms, such as oxygen atoms, in the polymer backbone that coordinates with Li ions. An effective coordination environment can help with salt dissociation to individual ions and the easier hopping of cations from one coordination site to another. [11, 12]

3.3 Data

The dataset used in this study is a subset of polymers from High-Throughput Polymer Design - Molecular Dynamics (HTP-MD) database [13, 14], consisting of 6024 linear chain homopolymers. Selected polymers were all unique and composed of H, C, F, S, P, O, N, elements, previously filtered from 53362 structures in the Zinc database [15] to ensure both synthesizability and potential application as electrolytes [16]. The ionic conductivity values for polymer-Li^+.$TFSI^-$ systems computed from MD simulations performed in the large atomic molecular massively parallel simulator (LAMMPS) [17] with the interaction parameters from Polymer Consistent Force Field ($PCFF^+$) [18, 19]. The simulations carried out for dataset generation have been previously performed in another study[14], at 353 K and the salt molality of 1.5 mol/kg. More details on the details of simulations and calculation of ionic conductivity can be found in the original work [14, 20].

To skew the model towards high-conductivity polymers, we randomly oversampled the top 5% of polymers in terms of ionic conductivity. This provides a train set that includes the same number of polymers from low-conductive and high-conductive classes. Additionally, PEO was added to the train set and oversampled 4000 times in the seed data. This method of selective oversampling was shown to be instrumental in guiding the model towards generating more promising polymer candidates.

3.4 Validation and feedback

For validation purposes, each iteration of the discovery process involved the generation of 50 polymer candidates, with their ionic conductivities evaluated through molecular dynamics simulations. These simulations adhered to the same protocol used in creating a previous dataset (HTP-MD: [13]). Details of MD simulations, dataset composition, and computing ionic conductivity have been included in 3.3 section, as well as previous studies [16, 14, 20, 13]. To ensure robustness, each candidate underwent five independent simulation replicas to determine its conductivity. Given the randomness in MD simulation results originating from different conformation sampling, this rigorous validation step was crucial for ascertaining the potential of each proposed polymer.

The feedback mechanism of our platform plays a vital role in its iterative learning process. After validation, we add both PEO and newly discovered polymers showing conductivity higher than PEO to the train set and oversample 4000 in total from all newly added polymers. This is to ensure the model can still explore polymers different than PEO. This enriched dataset is then used for retraining the generative model, thereby enhancing its ability to propose increasingly relevant and high-performance polymers in subsequent iterations. For demonstration purpose, the discovery campaign is conducted for two iterations.

4 Results

4.1 Conditional Generation

Generative models have gained substantial traction within the scientific community, showcasing their effectiveness and widespread adoption. However, within the realm of materials science, the primary challenge often lies not in simply generating physically reasonable candidates but rather in producing materials with specifically targeted properties, a process commonly referred to as inverse design. To illustrate this point, consider the case of polymer electrolytes where researchers aspire to discover novel polymers exhibiting enhanced ion conductivity. Historically, experimental development of new polymer electrolytes that can compete with PEO has been mainly focused on different trial errors, including altering the O/C ratio in the polymer backbone. However, this approach is inefficient, and the previous attempts were not completely successful.

In this study, we present an approach that leverages the transformer-based generative model, with a modified encoding approach to achieve conditional generation based on the desired property (see Experiment Setup for details). The encoding contains two layers of information: the SMILES representation of the repeating unit, and the class of the ion conductivity. When trained with more than 6000 data, the model not only learned how to generate the valid SMILES string as polymer repeating units but also the relationship between the repeating unit and the ion conductivity (please also see [8]). The outcome of the model is a generative batch of novel polymer repeating units, which exhibits a shifted distribution of ion conductivity with a notably higher mean value (0.89 mS/cm) when compared to the training set (0.06 mS/cm, excluding the added PEO polymers) (Figure 2). It is a remarkable improvement by a factor of 15.

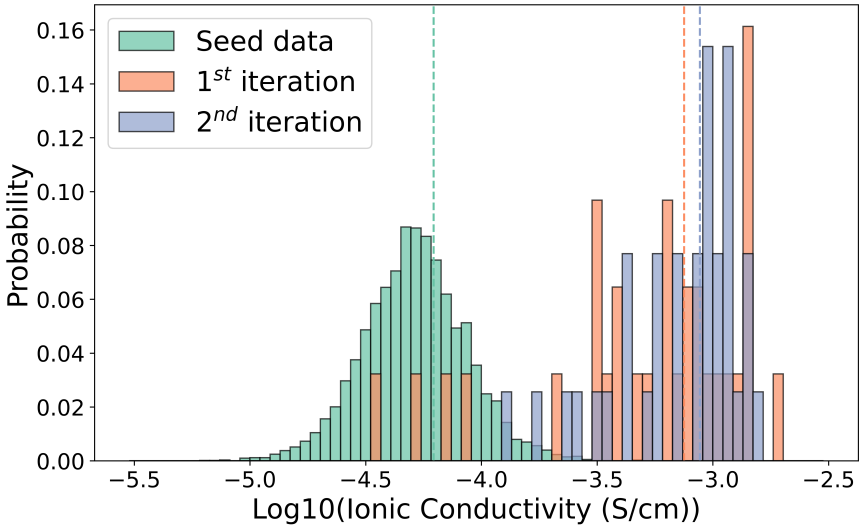

Figure 2: Comparison between the distribution of ionic conductivity in the train set (green) generated set in the first iteration (orange), and second iterations (purple). The dotted vertical lines show the mean ionic conductivity in each distribution. The values in the histograms are computed from MD simulations.

4.2 Active learning

While the initial success in conditional generation is commendable, we recognize that the current data on polymers remains sparse. To foster creativity and generate a wider variety of candidates, it's crucial to expose our model to new samples continuously. This involves integrating a feedback loop from the verification module in our platform.

Presently, our method involves using Molecular Dynamics (MD) simulations to validate ionic conductivity in newly generated polymers. Polymers demonstrating higher ion conductivity are then added to our training dataset. We examined how the distribution of polymers evolved across various iterations of active learning. The immediate finding is that the average ion conductivity of polymers generated by a model after one retraining cycle exceeded that of the initial model by 13%, as shown in Figure 3. Further, the lowest ion conductivity reported from the 2nd iteration batch is 3.6 times that of the 1st iteration. Furthermore, in Figure 4, we highlight new repeating units discovered through this process, which show higher ion conductivity compared to those in the original training set. After the second iteration of this process, the number of new discoveries increased from 8 to 11. This progression indicates the model's potential for ongoing enhancement, leading to increasingly effective outputs via a systematic feedback approach.

4.3 Discovered polymers

In Figure 4, we introduce 19 novel polymer repeating units whose ion conductivities, as confirmed by MD simulations, surpass that of PEO. It's important to reiterate that PEO currently holds the record for the highest ion conductivity for dry polymers (around 1 mS/cm at 353K, and $Li^{+}.TFSI^{-}$ molality of 1.5 mol/kg) in this field, and despite extensive research, no superior polymer electrolytes have been found—until now.

Among these repeating units, multiple polyacetals stand out. Polyacetals are polymers with a high oxygen-to-carbon ratio, similar to PEO, which facilitates efficient lithium salt solvation and creates effective pathways for lithium ion transport. Notably, poly(1,3-dioxolane) (P(EO-MO))—a polyacetal with a repeating unit of 1,3-dioxolane included in our list—demonstrates considerable promise. Its MD-calculated ion conductivity is 1.515 (± 0.199) mS/cm. Although experimental measurements (referenced in [21]) show slightly lower conductivity for P(EO-MO) at 0.4 mS/cm, its potential as a polymer electrolyte candidate remains significant due to its improved ion transport efficacy.

Our findings also reveal that many candidate polymers incorporate new nitrogen and sulfur elements diverging from the traditional focus on polycarbonates which consist only of carbon and oxygen. Particularly, the polymer with the highest conductivity identified during our research has the repeating unit of *ONCCOC*. In molecular dynamics simulations, it demonstrated an average ionic conductivity approximately double that of polyethylene oxide (PEO). This evidence underscores the creativity of our models and significantly broadens the scope for future research in this area.

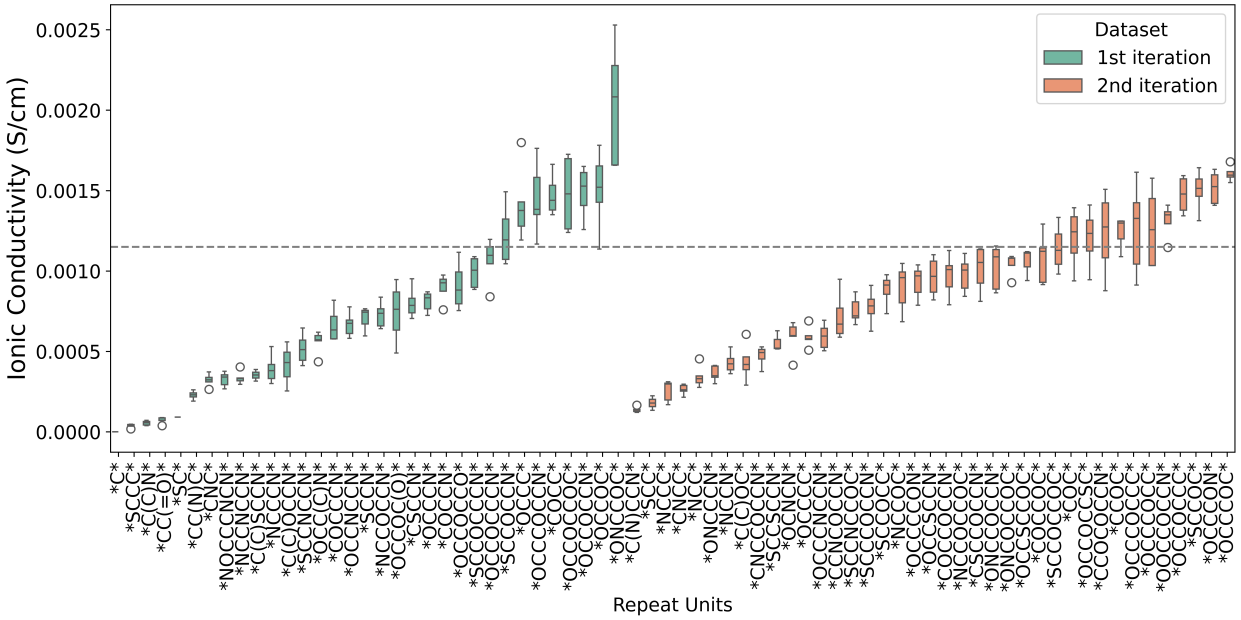

Figure 3: The MD simulation validation of polymers generated from two iterative candidate generations. The box plots show the mean and standard deviation in 5 MD simulations performed for each listed polymer.

The calculated ionic conductivity, derived from the cluster Nernst-Einstein equation, arises from both ion diffusivity and clustering. To elucidate the mechanisms underpinning the superior ionic conductivity observed in generated polymers, we conducted a comparative analysis of conductivity, ion diffusivity, and concentrations of free ions between PEO and the polymers exhibiting enhanced performance, as illustrated in Fig. 5. In this context, "free ions" denote those not incorporated into any clusters and moving freely, with their concentration determined as an average across simulation durations. The analysis reveals that the augmented conductivity in the most effective polymers from the first two iterations is attributed to both an increase in ion diffusivity and a higher prevalence of free ion clusters. Interestingly, it was also noted that several of the developed polymers exhibit a more efficient dissociation of the $Li^+.TFSI^-$ salt compared to PEO, indicating a potential for improved ion transport properties.

5 Discussion and Conclusion

As we reflect on the undeniable capabilities of our discovery platform, it is equally critical to acknowledge its current limitations and the vast potential for future enhancements. This recognition not only grounds our work in a realistic context but also opens avenues for exciting and impactful future research.

As mentioned earlier, the conditioned generative model, in its present state, is somewhat crude and unrefined. The model's efficacy as the primary driver of polymer discovery underscores the need for further development and refinement. Data oversampling strategies, that have proven to be instrumental in the experiment, also need to be thoroughly and systematically investigated. Future efforts could focus on improving the current model or exploring alternative architectures, potentially enhancing the platform's efficiency and accuracy in generating high-quality polymer candidates. Notably, developing generative algorithms for such specialized tasks presents its own set of challenges, particularly in evaluating their effectiveness. Unlike more straightforward metrics such as Mean Absolute Error (MAE) used in regression tasks, assessing the performance of generative models is less direct. A sister publication [8], which is released concurrently, delves deep into the benchmark of different generative model architectures in various generative tasks. It addresses the issue of evaluation by proposing a comprehensive and systematic methodology to benchmark different models for polymer SMILES code generation. This approach will significantly aid in comparing and optimizing generative models for polymer discovery.

In this study, we selected ionic conductivity as the primary metric to identify new polymer electrolytes. However, as highlighted in previous studies [21, 22], a more holistic measure is efficacy, defined as the product of conductivity and cationic current fraction. This measure considers both conductivity and the mobility of the preferred charge (cation here) in the system. In our study, ion conductivity calculations were standardized at a salt concentration of 1.5 mol/kg, which

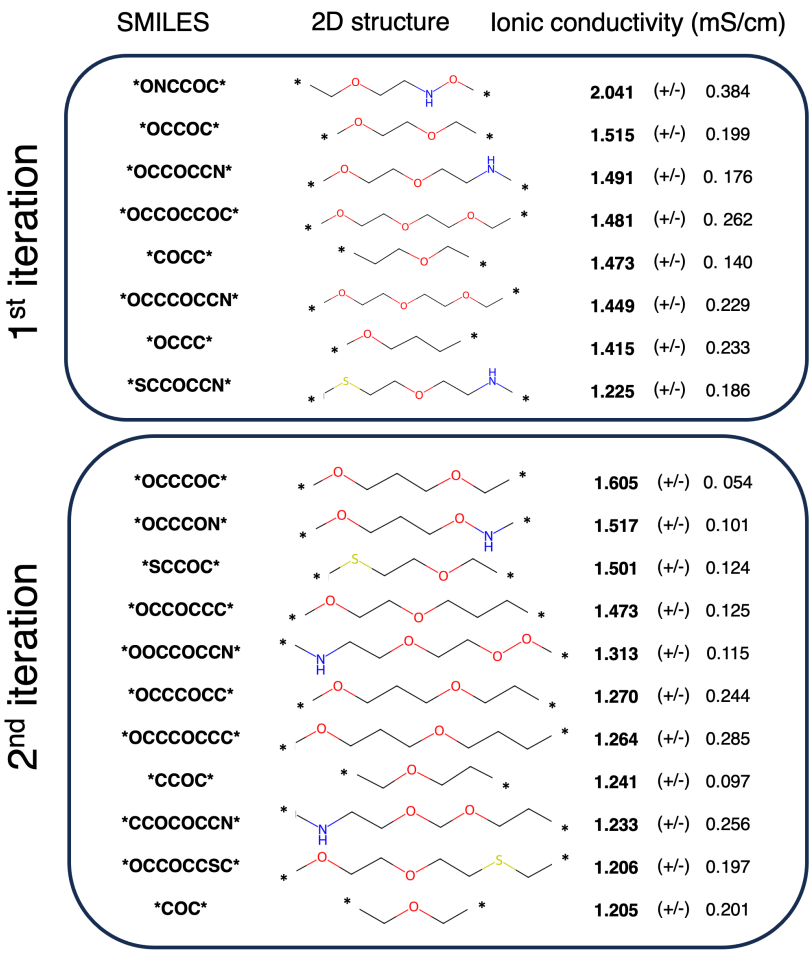

Figure 4: Discovered polymers from two iterative generation cycles. The polymer listed for each iteration exhibited an ionic conductivity superior to that of PEO.

is around the optimum for ion mobility for PEO. Moving forward, we plan to modify our evaluation criteria to include ion transport efficacy based on cation mobility and to establish a database encompassing various salt concentrations. Also, the synthesizability of the generated polymers needs to be considered as well, potentially by using it as an additional constraint for generation.

Another vital aspect of our platform is the validation process. Currently, we employ classical molecular dynamics simulations, which are fast and effective but may lack robustness. The ultimate goal is to corroborate our findings through experimental validation. However, given the current limitations in high-throughput polymer synthesis and characterization, simulation-based validation remains the most feasible approach. Our future endeavors will likely involve validating candidates through simulations before proceeding with experimental trials. Furthermore, the reliance on classical force fields, which are typically tailored to specific polymer classes, raises concerns about their extrapolative power to other materials. There is a noticeable disagreement between the reported ionic conductivity for the poly(EO-MO) (OCCOC) from ref[21, 22] and our computed one, which could be a result of the artifact of the force field. To address this, one promising direction is the use of ab initio molecular dynamics (AIMD) simulations with methods like Density Functional Theory (DFT). While these simulations offer greater robustness, their computational expense is a significant barrier. A potential solution could be the development and use of machine-learned force fields that combine the accuracy of DFT with the efficiency of classical force fields. Additionally, experimental verification would be critical for the promising candidates.

The workflow of our platform, which currently lacks full modularity and automation, also presents an opportunity for further improvement. Throughout the discovery campaign, the process has been devoid of human intervention, which

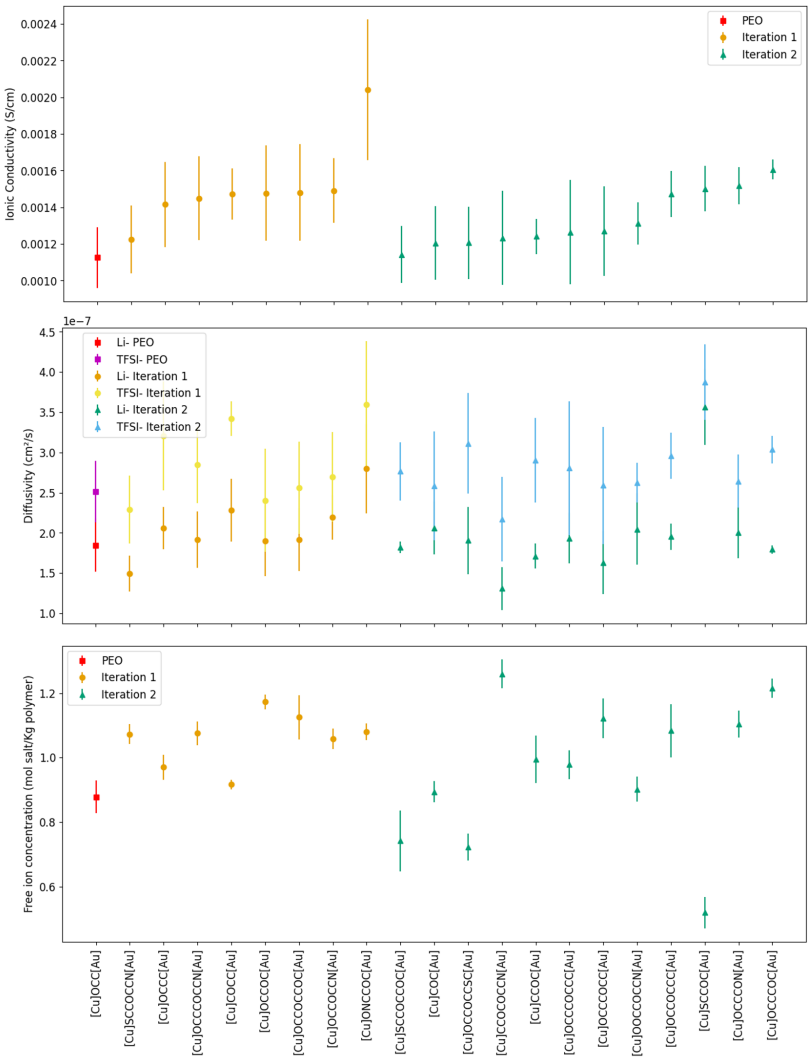

Figure 5: Comparison of the ionic conductivity, ions' diffusivity, and free ion concentration between PEO and the generated polymers with higher conductivity

lays the groundwork for developing a fully automated system. Future efforts will be directed toward enhancing the platform's modularity, enabling compatibility with various generative models and validation methods. This improvement would not only streamline the discovery process but also expand the platform's utility across different domains of polymer research.

6 Code and Data Availability

The dataset used to train the generative models can be accessed on the HTP-MD website: https://www.htpmd.matr.io/. The code for training the generative models can be found here: https://github.com/TRI-AMDD/PolyGen. The code for running the verification MD simulations can be found here: https://github.com/TRI-AMDD/htp_md

7 Disclosures

The authors wish to acknowledge that the discovery framework and materials discovered using our generative model framework, as described within this manuscript, are subject to a provisional patent application. This application has been submitted with the following details: U.S. Patent Application No. 63/582,871 titled "Methods of Designing Polymers and Polymers Designed Therefrom" with TEMA Reference No. IP-A-6823PROV and Darrow Reference No. TRI-1107-PR. The authors confirm that this does not alter our adherence to ArXiv policies on sharing data and materials.

Acknowledgments

We thank Professors Grossman, Yang Shao-Horn, Jeremiah Johnson, and Rafael Gomez Bombarelli, and Drs. Tian Xie, Sheng Gong, and Arthur France-Lanord at the Massachusetts Institute of Technology for their critical support in our molecular dynamics simulations and polymer electrolyte research. Their guidance and insightful discussions have greatly enhanced our study's development and robustness.

References

[1] Kai Wu, Jianhang Huang, Jin Yi, Xiaoyu Liu, Yuyu Liu, Yonggang Wang, Jiujun Zhang, and Yongyao Xia. Recent advances in polymer electrolytes for zinc ion batteries: Mechanisms, properties, and perspectives. *Advanced Energy Materials*, 10(12):1903977, 2020.

[2] Jeffrey Lopez, David G. Mackanic, Yi Cui, and Zhenan Bao. Designing polymers for advanced battery chemistries. *Nature Reviews. Materials*, 4(5), 4 2019.

[3] Yilin Hu, Xiaoxin Xie, Wei Li, Qiu Huang, Hao Huang, Shu-Meng Hao, Li-Zhen Fan, and Weidong Zhou. Recent progress of polymer electrolytes for solid-state lithium batteries. *ACS Sustainable Chemistry & Engineering*, 11(4):1253–1277, 2023.

[4] Qing Zhao, Sanjuna Stalin, Chen-Zi Zhao, and Lynden A. Archer. Designing solid-state electrolytes for safe, energy-dense batteries. *Nature Reviews. Materials*, 5(3), 2 2020.

[5] Rishi Gurnani, Deepak Kamal, Huan Tran, Harikrishna Sahu, Kenny Scharm, Usman Ashraf, and Rampi Ramprasad. polyg2g: A novel machine learning algorithm applied to the generative design of polymer dielectrics. *Chemistry of Materials*, 33(17):7008–7016, 2021.

[6] Ruimin Ma and Tengfei Luo. Pi1m: A benchmark database for polymer informatics. *Journal of Chemical Information and Modeling*, 60(10):4684–4690, 2020. PMID: 32986418.

[7] Chiho Kim, Rohit Batra, Lihua Chen, Huan Tran, and Rampi Ramprasad. Polymer design using genetic algorithm and machine learning. *Computational Materials Science*, 186:110067, 2021.

[8] Zhenze Yang, Weike Ye, Xiangyun Lei, Ha-Kyung Kwon, Daniel Schweigert, and Arash Khajeh. De novo design of polymer electrolytes with high conductivity using gpt-based and diffusion-based generative models. Unpublished manuscript, 2023.

[9] Mingpt, https://github.com/karpathy/mingpt.

[10] Andrey A Toropov, Alla P Toropova, Dilya V Mukhamedzhanoval, and Ivan Gutman. Simplified molecular input line entry system (smiles) as an alternative for constructing quantitative structure-property relationships (qspr). 2005.

[11] Anirban Roy, Bula Dutta, and Subhratanu Bhattacharya. Correlation of the average hopping length to the ion conductivity and ion diffusivity obtained from the space charge polarization in solid polymer electrolytes. *RSC advances*, 6(70):65434–65442, 2016.

[12] Yan Zhang, Jiande Wang, Petru Apostol, Darsi Rambabu, Alae Eddine Lakraychi, Xiaolong Guo, Xiaozhe Zhang, Xiaodong Lin, Shubhadeep Pal, Vasudeva Rao Bakuru, et al. Bimetallic anionic organic frameworks with solid-state cation conduction for charge storage applications. *Angewandte Chemie International Edition*, 62(42):e202310033, 2023.

[13] htpmd web app, https://www.htpmd.matr.io.

[14] Tian Xie, Ha-Kyung Kwon, Daniel Schweigert, Sheng Gong, Arthur France-Lanord, Arash Khajeh, Emily Crabb, Michael Puzon, Chris Fajardo, Will Powelson, Yang Shao-Horn, and Jeffrey C. Grossman. A cloud platform for automating and sharing analysis of raw simulation data from high throughput polymer molecular dynamics simulations, 2022.

[15] John J. Irwin and Brian K. Shoichet. Zinc - a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1):177–182, 2005. PMID: 15667143.

[16] Tian Xie, Arthur France-Lanord, Yanming Wang, Jeffrey Lopez, Michael A Stolberg, Megan Hill, Graham Michael Leverick, Rafael Gomez-Bombarelli, Jeremiah A Johnson, Yang Shao-Horn, et al. Accelerating amorphous polymer electrolyte screening by learning to reduce errors in molecular dynamics simulated properties. *Nature communications*, 13(1):1–10, 2022.

[17] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1–19, 1995.

[18] H. Sun. Force field for computation of conformational energies, structures, and vibrational frequencies of aromatic polyesters. *Journal of Computational Chemistry*, 15(7):752–768, 1994.

[19] David Rigby, Huai Sun, and B. E. Eichinger. Computer simulations of poly(ethylene oxide): force field, pvt diagram and cyclization behaviour. *Polymer International*, 44(3):311–330, 1997.

[20] htpmd source code, https://github.com/tri-amdd/htp_md.

[21] Rachel L Snyder, Youngwoo Choo, Kevin W Gao, David M Halat, Brooks A Abel, Siddharth Sundararaman, David Prendergast, Jeffrey A Reimer, Nitash P Balsara, and Geoffrey W Coates. Improved li+ transport in polyacetal electrolytes: Conductivity and current fraction in a series of polymers. *ACS Energy Letters*, 6(5):1886–1891, 2021.

[22] David M Halat, Rachel L Snyder, Siddharth Sundararaman, Youngwoo Choo, Kevin W Gao, Zach J Hoffman, Brooks A Abel, Lorena S Grundy, Michael D Galluzzo, Madeleine P Gordon, et al. Modifying li+ and anion diffusivities in polyacetal electrolytes: a pulsed-field-gradient nmr study of ion self-diffusion. *Chemistry of Materials*, 33(13):4915–4926, 2021.