A Robust and Versatile Generative Model for Inverse Design of Polymers

Haoke Qiu,[†,‡,#] Haozhe Huang,[¶,§,#] Hongli Yang,[†,‡] Alán Aspuru-Guzik,[*,¶,§,‖,⊥] and Zhao-Yan Sun[*,†,‡]

†*State Key Laboratory of Polymer Physics and Chemistry & Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, China*

‡*School of Applied Chemistry and Engineering, University of Science and Technology of China, Hefei 230026, China*

¶*Department of Computer Science, University of Toronto, Toronto, Ontario M5S 2E4, Canada*

§*Vector Institute for Artificial Intelligence, Toronto, Ontario M5G 1M1, Canada*

‖*Chemical Physics Theory Group, Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada*

⊥*Lebovic Fellow, Canadian Institute for Advanced Research, Toronto, Ontario M5G 1M1, Canada*

#*These authors contributed equally.*

E-mail: aspuru@utoronto.ca; zysun@ciac.ac.cn

Abstract

Efficiently designing polymers to meet specific requirements can expedite their translation into practical applications and lower development costs. Although generative inverse design is more efficient than trial-and-error or forward prediction–screening

strategies, the imperfect validity of current polymer generative models prevents their seamless integration into scientific discovery workflows. Moreover, their limited controllability—such as the inability to reliably generate polymers with specific functional groups or classes—further constrains their practical utility. In this work, we integrate the robust Group SELFIES method with the state-of-the-art polymer generator PolyTAO to achieve generating 100% chemically valid polymer structures, removing a long-standing bottleneck in polymer design. Compared with previous polymer generation models, this work can generate—on demand—polymers that match specified chemical motifs, polymer classes, and target properties across an effectively unbounded chemical space. We further introduce a task-agnostic, continuous pretraining strategy that combines physics-informed heuristics with reinforcement learning. This approach preserves strong generative performance on user-defined tasks, even in low-data regimes. As a proof of concept, we rigorously validated the dielectric constants of 30 polyimides generated via controlled, on-demand design using first-principles calculations, finding deviations of less than 10% from their target values. Designed as a powerful backend engine for polymer inverse design, our model is deployment-ready, and integrates seamlessly with high-throughput, self-driving laboratories and industrial synthesis pipelines.

Introduction

Polymers span an exceptionally large property space due to their hierarchical architecture and diverse aggregate states, enabling a wide range of technological applications.[1–4] Within the virtually limitless design space of polymer composites, homopolymers alone represent an enormous candidate pool—exceeding 10^{18}—owing to the sheer number of available monomers.[5,6] This breadth represents a trade-off: it offers a rich design space for targeted screening, but also raises the cost of trial and error. Since Staudinger established the foundation of synthetic polymer science in the 1920s-1930s, scientists have developed numerous polymeric materials;[7] however, the discovery of new polymers with significantly enhanced

performance or lower production costs has remained remarkably slow.[8]

Computational methods based on density functional theory (DFT) and molecular dynamics (MD) have significantly accelerated the polymer discovery process over the past half century.[9–15] Together with experiments, these methods have produced large, high-quality datasets, laying the foundation for recent advances in polymer machine learning (ML).[16–18] Building on this foundation, data-driven models can efficiently and accurately predict the thermal,[19–21] electrical,[22,23] optical,[24] medical,[25,26] rheological[27] and mechanical properties[28] of previously unexplored polymers, thereby facilitating high-throughput screening.[29–31] However, this paradigm still relies on researcher-defined candidate pools, which seldom encompass the full chemical space of polymers, offering no guarantee that the optimal structures for a given target will be identified.[32]

Consequently, the inverse-design paradigm based on on-demand generation has recently attracted considerable attention from polymer scientists.[32–37] In inverse design, one starts from target property values and uses ML models to generate the corresponding polymer structures, without being restricted to a predefined pool of candidates. Typically, such models rely on pretraining tasks that reconstruct SMILES strings[32,34] or molecular graphs[35–37] across large polymer datasets-an approach analogous to masked pretraining in general-purpose large language models (LLM).[38,39] Similar to how natural language generation must obey grammatical and orthographic rules, polymer generative models must ensure chemical validity (e.g., a carbon atom must not exceed its valency of four single bonds). This requirement is critical yet challenging: standard SMILES-based variational autoencoder (VAE) models often generate molecules with validity rates below 30%.[32] Graph-based generative models improve the validity but typically remains below 90%.[35–37] Recently, our prior work PolyTAO, having incorporated a physics-supervised LLM pretraining strategy with nearly one million polymer samples and their SMILES representations, achieved a generation validity of over 99%.[34] While this was a substantial step forward, model-generated structures still require post-hoc chemical validity checks before practical use, which significantly reduces the

throughput of self-driving laboratories (SDLs) and SDL structure.[40,41] Furthermore, existing generative models tend to produce structural features in a largely random manner, making it difficult to condition models to generate polymers with user-specified substructures. This remains a crucial consideration for experimental realization, as the availability of monomers and feasible polymerization reactions in the laboratory is inherently limited.[42]

To integrate polymer generators more seamlessly into laboratory workflows, we aim to address both the issue of validity and randomness. Our previous system, which we will refer to as PolyTAO_v1, has already demonstrated the strengths of its model architecture and supervised pretraining based on P-SMILES.[43] Although P-SMILES is widely used to encode polymer structures, it captures only the underlying connectivity as tokens, and its sensitivity to syntax can limit generation validity.[44] Moreover, directly using P-SMILES also prevents the model from explicitly modeling polymer-type information (e.g., polyesters, polyimides, polystyrenes, etc.).

Motivated by the above needs, we adopt Group SELFIES[45] as our base representation. Among recently proposed advanced molecular representations,[45–49] Group SELFIES—an extension of SELFIES that guarantees 100% chemical validity—embeds chemical group information directly into its string representation. We build on this scheme and embed polymer functional group information directly into the data representation as "groups". This eliminates the need for models to relearn well-established polymer-class distinctions from token frequencies, enabling the model to learn the correspondence between polymer-type and their structural features more effectively. This integration yields chemically valid, unique, and novel (V.U.N.) polymers that are guided by polymer class and target properties. For class-conditioned polymer generation, we focus on eight representative polymer classes—polyester, polystyrene, polyureas, polysulfone, polyimide, polyketone, polycarbonate, and an "other" category—since these classes are relatively well represented in our training dataset, PI1M. It is worth noting that this approach is readily generalizable to any polymer type by extending the list of groups accordingly. We show that with standard pretraining, our model

consistently achieves 100% chemical validity for class-conditioned generation. For target-conditioned polymer generation, we develop PRECISE, a Physics-informed REinforcement learning-based ContInuouS prEtraining strategy, which enables property-conditioned design. In particular, we showcase the capability of PRECISE by generating polyimides with specified target dielectric constants and demonstrating its ability to perform reliable polymer generation in low-data regimes. In our experiments, PRECISE achieves an impressive success rate in generating entirely V.U.N. polyimides absent from both pretraining and downstream datasets. Together, these results highlight the practical effectiveness of our model for real-world on-demand polymer design.

Results and discussion

Embedding Polymer Chemistry using Group SELFIES

SELFIES[44] is a 100% robust molecular string representation that guarantees a mapping from any SELFIES string to a syntactically valid chemical graph. Group-SELFIES[50] extends the grammar of SELFIES such that functional fragments with predefined attachment points can be directly added as group-tokens into the vocabulary.

In this work, we leverage Group SELFIES as a class-aware molecular representation whereby polymer-defining substructures are directly encoded as group tokens. Effectively, this abstracts away the need for our model to (1) learn the atomic connectivity of functional fragments and (2) handle low-level syntactic validation. We achieve (1) by curating polymer fragments of interest into group-tokens, and (2) by taking advantage of the 100% validity that comes with using context-free grammar such as SELFIES. In Figure 1a, we show an example of the decoding (and encoding) process for a generated polymer. In the Section 1 of Supplementary Information (S1), we provide a more detailed summary of the syntax rules of Group SELFIES.

Notably, Group SELFIES uses a large-substructure-first strategy along with a user-

customizable priority level when encoding molecular graphs into groups. This resolves the ambiguity of polymers containing multiple motifs that map to multiple polymer classes without sacrificing SELFIES' 100% validity. For example, polyesters contain both ester functional groups ([O]C=O) and carbonyl groups (C=O). Naively, the Group SELFIES encoder may suggest a polyketone label, but by applying the large-substructure-first strategy of Group SELFIES, such polymers would be correctly categorized as polyesters rather than polyketones. We distinguish eight classes—polyesters, polystyrenes, polyureas, polysulfones, polyimides, polyketones, polycarbonates, and an "other" class for polymers outside these categories. Representative group features for each class are shown in Figure 1b. Class conditioning is implemented by inserting class-specific group tokens to the encoder for the first seven classes; sequences lacking any of these tokens are assigned to "other."

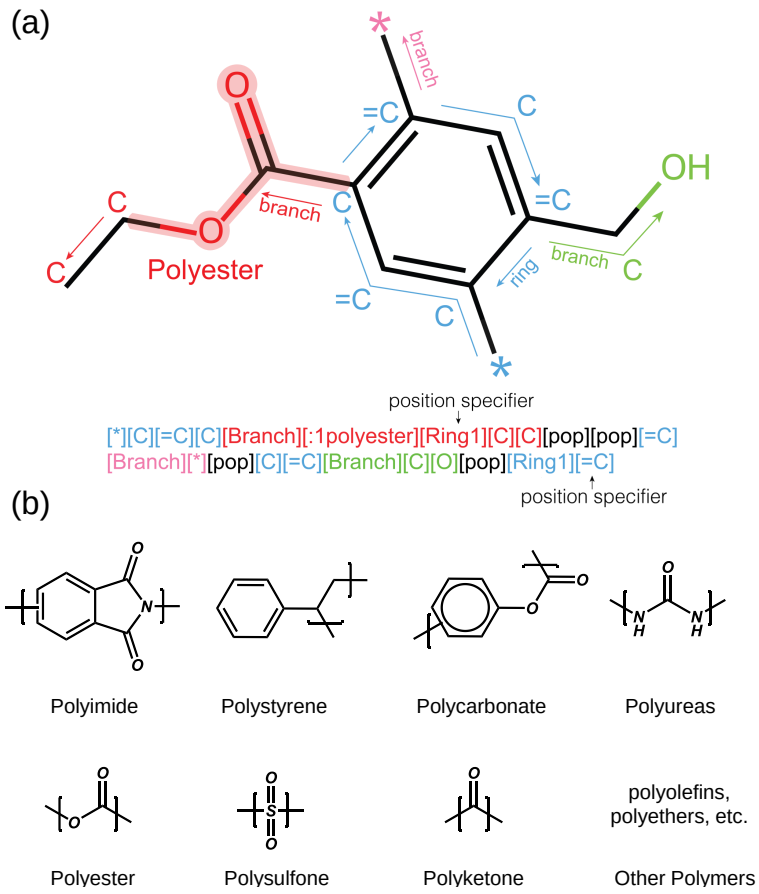

Figure 1: (a) The decoder follows the blue arrow and branches off a group or branch token until a corresponding pop token is read. The group token itself specifies the entry point, and the following token specifies the attachment position of the next token. A ring is formed with a ring token and a number token specifying the position of the atom to close the ring. Validity is ensured by (1) keeping track of the available valency at each step in the decoding process, and (2) the unique feature of SELFIES that allows every token to also be interpreted as numbers. We see that in this formulation, the Polyester fragment is encoded as a single group token. (b) Predefined polymer classes and their groups.

Generating V.U.N., Class-Conditioned Polymers

We use the PI1M dataset[43] as our basis for modelling polymer-space. To our knowledge, PI1M is the largest publicly available polymer dataset for model training and evaluation, containing nearly one million samples, which is 1.6 times the number of training samples used by the state-of-the-art generative model for inorganic materials, MatterGen.[51] We define our generation for polymers as a two-step process, in which we (1) pretrain a general base

model based on Group SELFIES for generating V.U.N. polymers, and (2) finetune the model on downstream tasks. In this section, we will focus on (1), describing the main training procedure of our framework and introducing the model architecture. We defer the discussion on finetuning and downstream applications to the next section.

To initiate the pre-training process on a warm start, we build upon our previous work, PolyTAO_v1,[34] which demonstrated the effectiveness of supervised pretraining for the task of inverse polymer generation. Building on this foundation, we adopt it as the backbone and initialize our model with its pretrained weights. The framework follows a symmetric encoder–decoder architecture, where carefully designed polymer prompt–answer pairs accelerate the acquisition of polymer chemistry knowledge and enhance inverse generation capabilities (Figure 2a). As inputs to our model, and thus the encoder, the prompts incorporate two modalities of polymer information: (i) Polymer Group features (G), which specify the polymer type, and (ii) Chemical features (F), consisting of 15 predefined physicochemical properties; the result of the encoder is fed through the decoder for which a polymer is generated and the answer/label for each prompt is the corresponding polymer string that exhibits the prompt properties (see PolyTAO_v1 for details). The pretraining objective facilitates a well-structured embedding space through large-scale learning of the mapping between the structural features of specified polymer types and their corresponding Group SELFIES representations.

The backbone's hierarchical attention layers are well-suited to capture different aspects of the task (Figure 2b). Specifically, the encoder employs a standard full self-attention mechanism, allowing the model to learn relationships among G, F, and all tokens within the prompt, thereby producing accurate polymer embeddings. The decoder first adopts a causal self-attention mechanism, where each token can only attend to itself and preceding tokens, ensuring the stepwise generation of Group SELFIES. Finally, the decoder's cross-attention mechanism enables the model to jointly attend to the input prompt and the partially generated Group SELFIES, thereby supporting controlled and context-aware inverse generation.

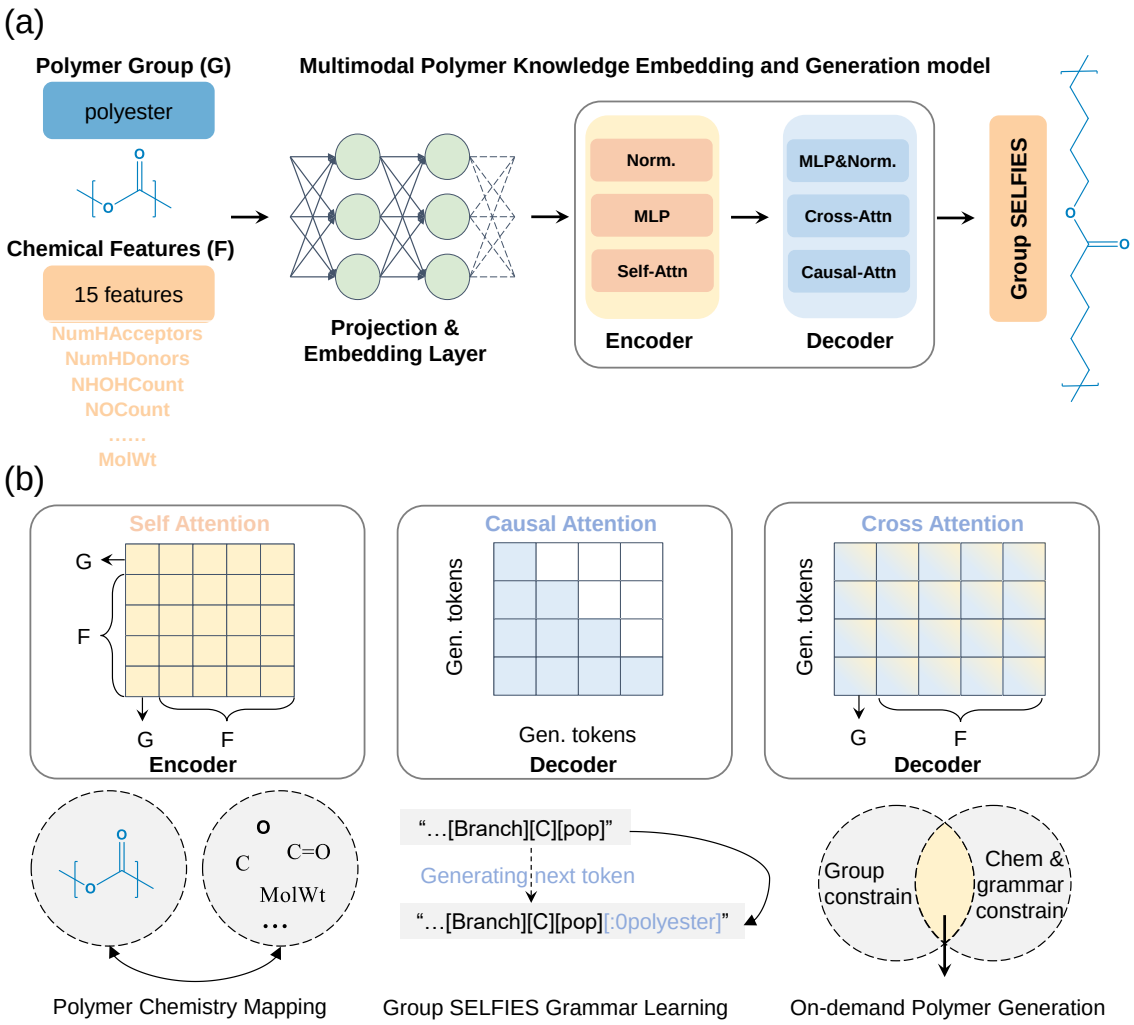

Figure 2: (a) Our model architecture, which utilizes a symmetric encoder–decoder architecture with total 24 layers. (b) The hierarchical attention mechanisms within the model architecture facilitate task learning. Colored blocks indicate activated attention, while white blocks denote positions not attended to.

Here we list the main configurations employed for our pretraining process. The learning rate was set at 1×10^{-6} and increased to its peak value over the first 20% of warm-up steps, followed by a cosine-decay schedule. The model was trained for a total of 50 epochs on an H800 GPU cluster. The model's loss converged without signs of overfitting (Figure 3a), indicating effective acquisition of Group SELFIES syntax and accurate decoding to polymer structures. In LLMs, performance is generally assumed to scale positively with both model size and the amount of training data. To probe this relationship in polymer generation,

we constructed a series of PolyTAO variants with 100M, 150M, 200M, and the full 220M trainable parameters by selectively freezing model weights. These variants were trained on datasets ranging from 0.1k to the full 800k samples and evaluated on a hold-out set of 1,000 randomly selected entries (Figure 3b). We observed that the loss on the hold-out set decreased exponentially as the training data increased. Furthermore, the optimal performance correlated positively with model scale, with the full-parameter PolyTAO achieving the lowest loss. These findings suggest that future advances of this model may be enabled by jointly scaling both the training corpus and the model capacity.

To reflect practical usage, we report performance across different sample sizes (averaged over five runs) in 3c-d. For $\leq 10^3$ samples—a common early-stage discovery setting—Uniqueness and Novelty both exceed 80%, yielding diverse, novel candidates. Around 10^4 samples—relevant to high-throughput screening—the V.U.N. fraction remains $> 60\%$, demonstrating scalability. Even at 10^5 samples, $\sim 40\%$ of outputs are V.U.N. Raising the sampling temperature (e.g., to 1.5) further boosts novelty and the V.U.N. rate (Figure 3d). Overall, the model supplies diverse, chemically meaningful candidates across a wide range of generation scales, supporting both experimental and computational workflows.

During the generation of V.U.N. polymers (See Methods for details), a particularly critical metric for polymer generative models is the chemical validity of the generated polymer structures. Therefore, in a high-throughput setting, even 99% chemical validity[34] may incur a non-negligible cost in both computation and post-generation curation. Remarkably, by leveraging the inherent robustness of Group SELFIES, we have, for the first time, achieved 100% validity in polymer generation (Figure 3e), eliminating the need for validity filtering entirely.

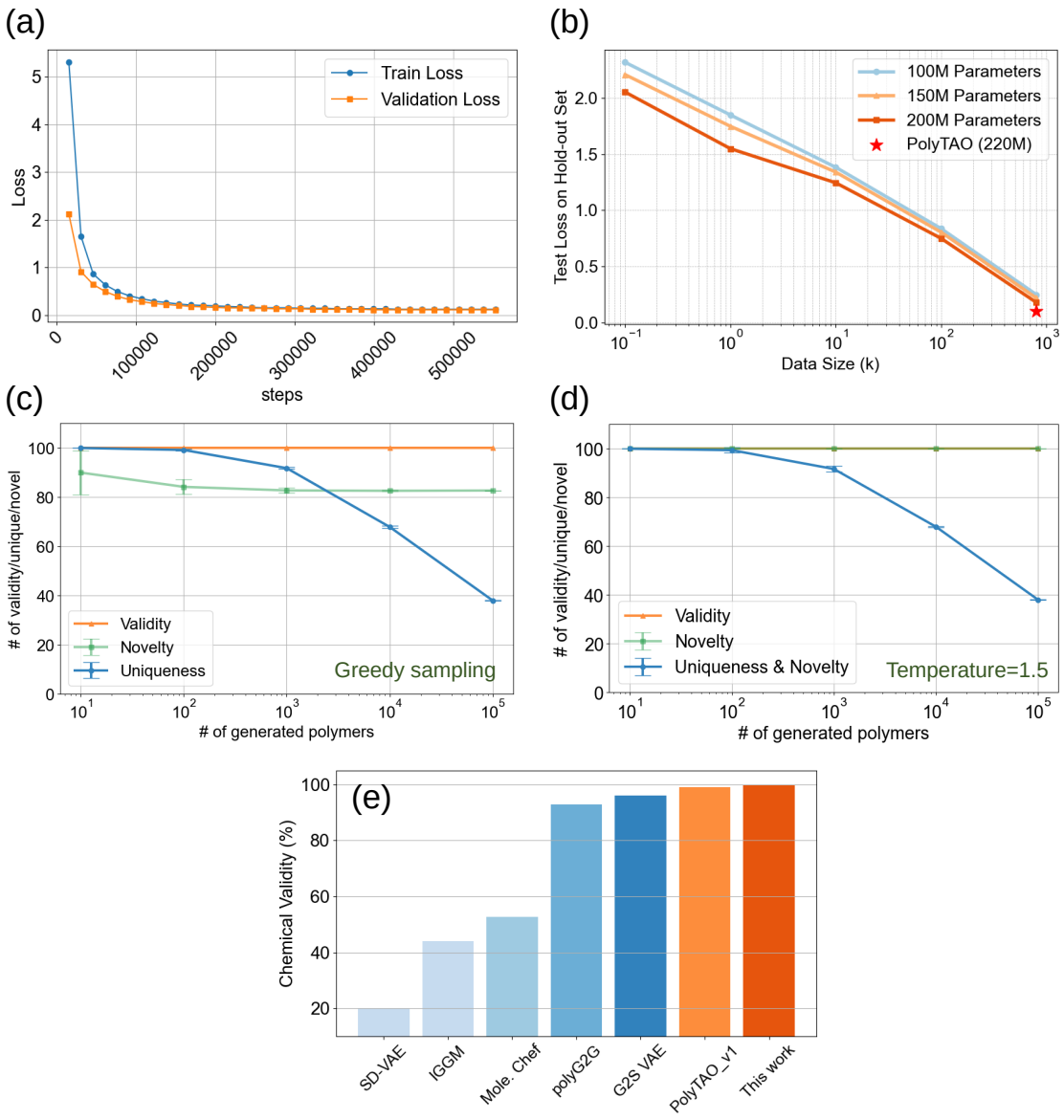

Figure 3: (a) Training dynamics of our model. (b) Test loss on the hold-out set under different trainable parameter sizes (100M, 150M, 200M and 220M) and data size (0.1k, 1k, 10k, 100k and 800k). Results of large-scale generating V.U.N. polymers when greedy sampling is enabled (c) and the temperature is to 1.5 (d). Greedy sampling is the default model setting, which enables the model to generate the next token with the highest probability. In contrast, when the generation mode is set to random sampling, for example, the temperature parameter exceeds 1.0, our model will generate polymers in a freer manner. (e) The generation capability on chemical validity of polymer generation models.[32,34–37,52]

On-demand Design with Specified Polymer Class

Beyond this milestone in generating V.U.N. polymers, our framework addresses several longstanding limitations of prior generative approaches (Table 1). Similarity-based generation (methods that decode latent features to SMILES or graphs[32,35,36]) struggles with property-specific design; for example, targeting a glass transition temperature (T_g) of 350 °C. Combination-based generation (assembling predefined functional fragments[53,54]) supports property targeting but confines exploration to a fixed motif library. Our previous PolyTAO_v1[34] generated property-conditioned polymers but could not guarantee the presence of required functional groups, which limits practical utility; for instance, applications requiring resistance to atomic oxygen in low Earth orbit expect polyimide structures.

Here, we encode polymer-type information (and, in practice, other user-specified chemical features) directly into the Group SELFIES representation. Coupled with PolyTAO's ability to navigate polymer space toward target properties, this yields a single framework that simultaneously delivers V.U.N. generation, property control, and functional-group specificity, facilitating integration of polymer generative models into manufacturing and materials design workflows.

Table 1: Comparison of different polymer generation methods in terms of their design capabilities.

Methods	Property	Functional Group	Full Space	All
Similarity-based[32,35,36]	✗	✓	✓	✗
Combination-based[52–54]	✓	✓	✗	✗
PolyTAO_v1[34]	✓	✗	✓	✗
This work	✓	✓	✓	✓

As an important proof of concept, we first evaluate this work on the previously unaddressed task of generating polymers with specified functional groups using PolyTAO, demonstrating that our approach extends the capabilities of PolyTAO in this important aspect. This proof of concept is aimed to evaluate whether the generated polymer structures corre-

sponded to the intended polymer classes (i.e. whether they contained the specified functional groups). Specifically, we quantified the match between the generated polymer classes and the target classes within the test set (∼200k samples). Impressively, when prompted to generate 'other polymers', the model achieved a 100% success rate in avoiding generating the other seven predefined polymer classes (Figure 4a), demonstrating its robust basic understanding of polymer class distinctions and highlighting its potential for class-specific polymer generation. In addition, the model attained on-demand generation success rates exceeding 70% for polymers in the polyketone, polycarbonate, and polyester categories, with representative examples shown in Figure 4b. However, the limited amount of data for certain polymer types in PI1M leads to suboptimal performance in property-targeted generation for these types. Notably, we can find a general positive correlation between the on-demand generation success rate and the number of training samples available for each class, which explains the relatively low success rate for categories such as polyimides (Figure 4a). We will present our PRECISE approach to address this issue in the following section.

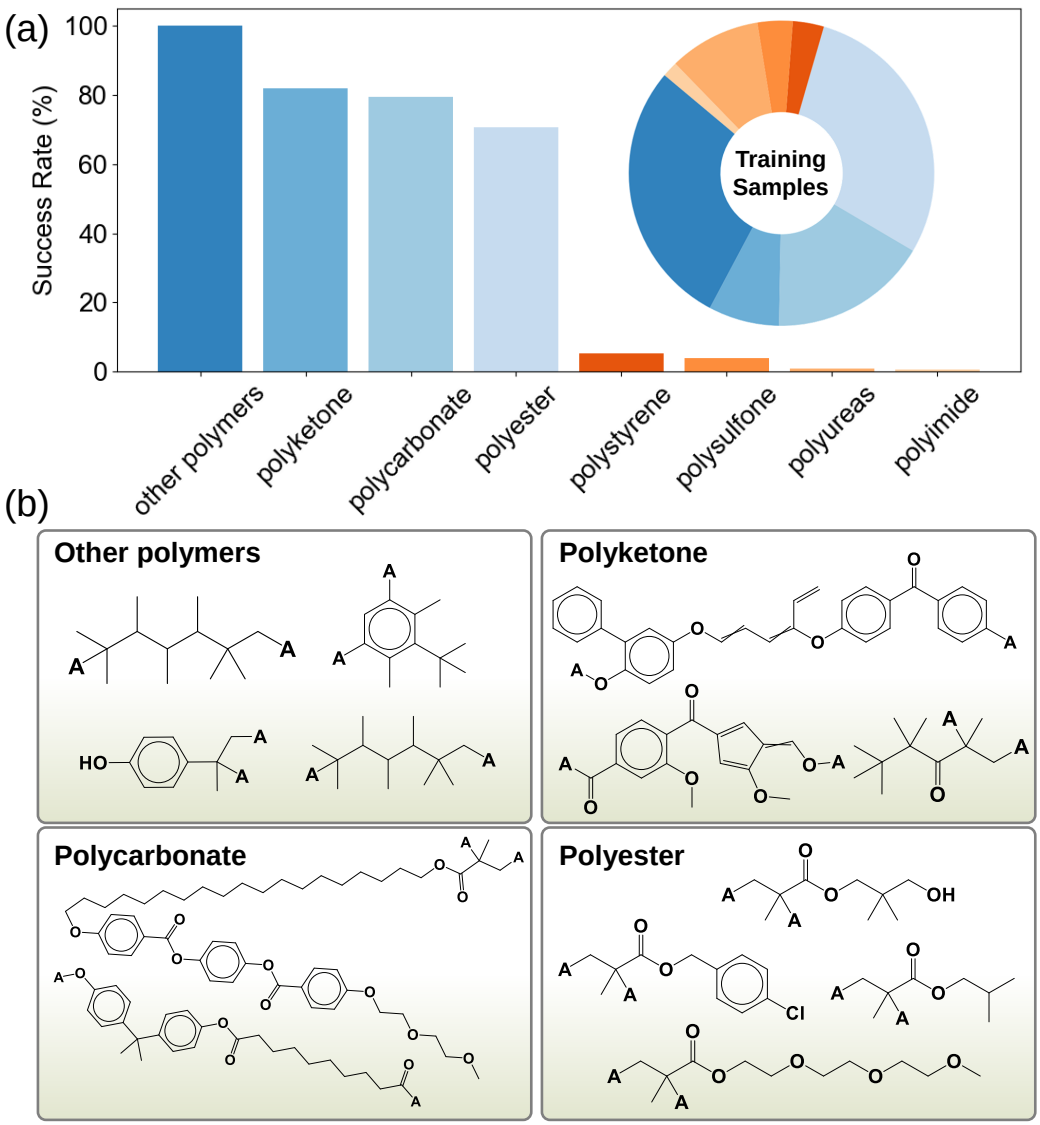

Figure 4: (a) Model performance on generating polymers of specific classes and their training samples distribution. (b) Visualization of randomly selected polymers generated by our model.

On-demand Design with Specified Polymer Class and Specified Property

PolyTAO_v1 has already demonstrated a strong capacity to generate polymers with specified macroscopic properties.[34] Here, we sought to further investigate whether the Group SELFIES-enhanced PolyTAO could achieve the same level (or better) of on-demand gen-

eration performance. Given that the current model exhibits the lowest success rate when generating polyimide-type polymers, we deliberately chose this category as a stringent test for the model's capabilities. Specifically, the objective of this section is to generate structurally diverse polyimides that meet predefined property criteria. In our previous work, we established a high-quality dataset of polyimide dielectric constants through high-throughput DFT calculations (Figure 5a), and the computational results were shown to be in good agreement with experimental values. We therefore used this dataset to continue pretraining the backbone model, aiming to strengthen its capacity for on-demand generation of polyimides (with specified dielectric constants). At the same time, this effort serves as a proof of concept for enhancing the model's on-demand generation performance in low-data regimes.

Our in-house polyimide dataset contains dielectric constant data for over 1,000 structurally distinct polyimides. Although this represents a considerable dataset within the polymer field, it remains relatively small for training LLMs; further expanding this dataset (potentially by an order of magnitude or more) would be unsustainable and computationally expensive. To address this challenge, we present our PRECISE framework(Figure 5b; see Methods for details). PRECISE is a physics-informed RL strategy designed to enhance the model's ability to generate polymers of specified types. By incorporating physical intuition, the strategy guides the model to explore and generate in a targeted manner. For example, if a dielectric constant value is commonly observed among polymers, the model is encouraged to fully leverage its exploratory capacity, potentially yielding more novel and complex structures. Conversely, in less common regimes, the model is guided to adopt a more conservative approach, favoring the generation of familiar chemical structures. This is a property-agnostic strategy that can be readily applied to other polymer property domains.

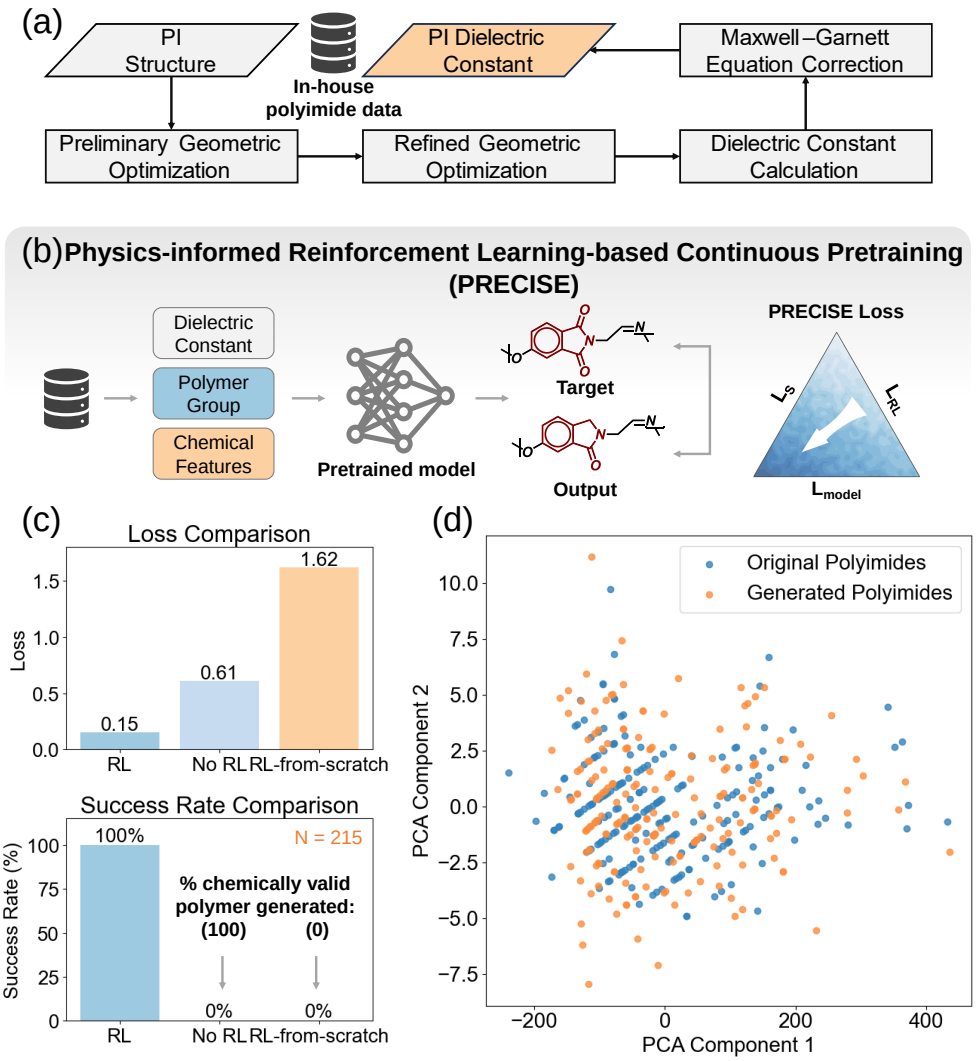

Figure 5: (a) DFT calculation workflow. (b) Through the PRECISE framework, we incorporated polymer group token (i.e., 'polyimide') and polymer property token (i.e., the value of dielectric constant) into the model inputs, enabling the generation of polymers tailored not only to specific types but also to target properties. (c) After PRECISE continuous pretraining, our model showcased the lowest loss, compared to continuous pretraining without PRECISE and pretraining from scratch with PRECISE. Notably, our PRECISE strategy enabled the model to generate structures exclusively of the desired polyimide type, whereas both of the other strategies failed in this task. Furthermore, the comparison between training without PRECISE and training from scratch with PRECISE also demonstrates the strong capability of our pretrained model to generate chemically valid polymers. (d) Visualization of original polyimides from the test set and generated polyimides. All generated polyimides were entirely novel with respect to both training and test sets.

Building on the semi-templated generation paradigm of PolyTAO, we introduced an additional token to specify the target polymer type—'polyimide', in this case (Figure 5b). Using

the same model architecture as in the pre-training stage, we set the peak learning rate to 5×10^{-5}. Given the limited size of the dataset for this task, the model was trained for 100 epochs. Using PRECISE, we successfully achieved the targeted generation of polyimides within a limited-data regime (Figures 5c–d). In contrast, without RL, the same challenge persists as during initial pretraining: the scarcity of learnable samples renders the model unable to generate the specified polymer class. Thus, our framework offers a robust solution for scientists seeking to perform on-demand generation under small-data constraints. Notably, even with RL, training this model entirely from scratch (without pretraining) under identical settings failed to generate any valid polyimide structures—and, in many cases, failed to yield chemically valid polymers at all. This underscores the strong capability of our pretrained model, when coupled with Group SELFIES, to enable high-fidelity polymer generation.

Another notable result is that, within the test set comprising 215 cases, our model achieved 100% generation of chemically valid polymers, with 100% classified as polyimides, and 100% representing novel structures (Figure 5d). This performance holds not only for the in-house computational dataset but also for the PI1M dataset. These findings underscore the model's strong capability to explore the polymer chemical space effectively.

Discovery of Polyimdes with Different Dielectric Constants

In the preceding section, we designed 215 V.U.N. polymers with target dielectric constants spanning 2.51–4.09 (Figure 6a). As a proof of concept, we performed first-principles DFT calculations on model-generated polyimides to assess the capability for on-demand design. Here, 30 structures were randomly chosen and subjected to rigorous DFT calculations. Their dielectric constants nearly span the entire range observed in both the training and test sets (Figure 6a). The computed dielectric constants closely matched the model predictions for all cases, with deviations within 10% (Figure 6b). This demonstrates the superior capability of our model in the on-demand design of polymers with both specified properties and targeted structural types.

Of note, the development of polyimides with such relatively low dielectric constants (below 3.5[55–58]) holds great promise as low-k dielectrics, where the reduced permittivity directly alleviates capacitive coupling and RC delay in integrated circuits, thereby enabling faster signal propagation and lower power efficiency.[59] A considerable fraction of the generated and validated structures fall within the low-dielectric-constant regime, with representative examples shown in Figure 6c. This advantage becomes particularly critical in the current era of AI-driven computation, where unprecedented processing power demands call for advanced dielectric materials that can sustain high-speed, energy-efficient information transfer.

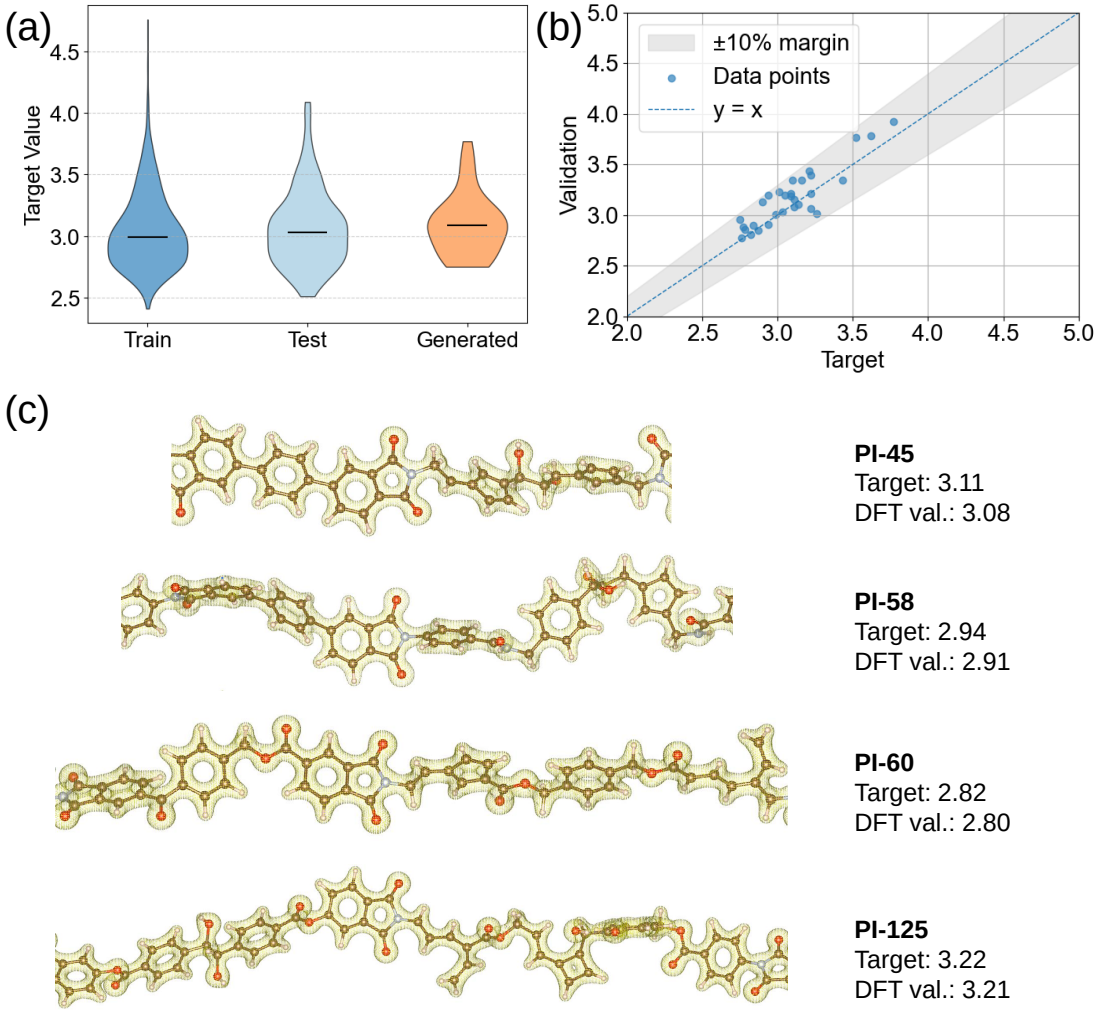

Figure 6: (a) Distributions of dielectric constant of polyimides from train set, test set and the generated polyimides. (b) Scatter plot of 30 random selected polyimdes to be calculated. (c) Illustration of four generated polyimides with relatively low dielectric constants.

Discussion and Conclusion

Inverse design of molecules and materials is a promising yet challenging task, requiring models to learn an accurate mapping between structure and property spaces. For polymers, the complexity is further amplified by their aggregated states and intrinsic multi-scale nature. PolyTAO_v1 demonstrated that LLMs can learn polymer-specific features from p-SMILES and generate polymers with target properties. However, the generation process lacked sufficient controllability, and some outputs, even if theoretically valid, could not be synthesised or polymerised due to the unavailability of required monomers in the laboratory. To address these limitations, we integrated PolyTAO with the robust Group SELFIES representation, achieving a breakthrough in generation validity—improving it to 100%—while enabling controllable polymer generation. We illustrate this capability with the task of generating polymers containing a specified functional group (imide group) and meeting a target property. Compared with conventional computation-driven discovery or ML-assisted forward-screening strategies, PolyTAO can propose a scientifically desirable polymer structure in 20 ms on a consumer-grade GPU (RTX 3090, tested for generating 1,000 samples), without the need for pre-constructed candidate libraries or additional sampling.

Despite these advances, there remain opportunities for improvement. Due to the limited diversity of available polymer data types, PolyTAO currently generates only homopolymers, with insufficient capability for co-polymer or composite design. Fine-tuning PolyTAO for co-polymer generation is a promising direction, but integrating multi-scale and multi-modal polymer datasets is an urgent task to enable PolyTAO to address a broader range of polymer design problems. PRECISE provides a proof of concept for tackling complex polymer design scenarios. With appropriate prompt engineering, PolyTAO can be further extended to multi-component polymer generation tasks, including copolymer design, monomer stoichiometry specification, and chain-structure control.

We envision that the polymer community could accelerate experimental workflows by fine-tuning PolyTAO with domain-specific datasets. Furthermore, our supervised pretraining

paradigm could be extended to other generative tasks in polymer science, such as molecular dynamics trajectory generation or infrared spectrum generation. We believe that increasingly precise generative AI will empower polymer scientists and engineers to address diverse design challenges with unprecedented efficiency.

Methods

Using V.U.N. to Benchmark Polymer Generation Models

We aim for polymer generative models to efficiently produce structurally interesting candidates—namely, V.U.N.. polymers—to explore the polymer chemical space as broadly as possible. Below, we define each of the three metrics that constitute V.U.N..:

Validity: Validity is defined as the ability of a generated polymer to be successfully parsed and converted into a chemically meaningful molecular graph. Specifically, a valid polymer must (1) conform to the syntactic rules of chemistry—including correct usage of atomic symbols, bonding descriptors, balanced branches, and ring indices; (2) pass valence checks, ensuring that no atom exceeds its allowable number of bonds; and (3) satisfy additional sanitization procedures, which verify aromaticity, bond type assignments, charge balance, ring closure, and stereochemistry. Polymers that fail at any of these stages—due to parsing errors, valence violations, or sanitization exceptions—are considered invalid. In our study, Validity is quantified as the proportion of valid polymers among all generated samples.

Uniqueness: A polymer is considered unique if it does not match any other structure within the set of generated samples. To avoid overestimating this metric due to differences caused by structural rearrangements, we apply canonicalization to all generated and reference polymer structures prior to comparison.

Novelty: A polymer is considered novel if it does not match any structure present in the PI1M dataset, which contains 995,799 distinct polymer structures.

Physics-informed Reinforcement Learning-based Continuous Pretraining

In this task, our training framework consists of three stages: (1) learning to design polymers with target dielectric constants on demand; (2) learning to generate polyimides rather than other polymers; and (3) balancing the synthesizability of the generated polyimides with adequate exploration of the relevant chemical space. The overall loss function comprises three components: the backbone model loss (L_{backbone}), a RL (RL)-based loss (L_{RL}), and a synthesizability loss (L_{S}).

To address the challenge of limited training data, we introduce a RL strategy designed to guide the model towards generating polymers of a specified target type—in this case, polyimides—while allowing researchers to adapt this approach to any desired polymer category relevant to their specific applications.

$$L_{\text{RL}} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\Big(s^{(i)} \not\supseteq k\Big) \tag{1}$$

Here, k denotes the target polymer category specified for generation—in this case, polyimide. The term $s^{(i)}$ represents the i-th Group SELFIES string generated by the model. $\mathbb{I}(\cdot)$ is an indicator function that returns 0 if the generated structure belongs to the polyimide category and 1 otherwise. N is the batch size.

The synthesizability loss is specifically introduced to balance the trade-off between generating polyimides with high synthetic accessibility and adequately exploring the chemical space of polyimide structures. For any given polyimide repeating unit, its synthesizability can be quantified by the SAscore, which ranges from 1 to 10, with lower scores indicating greater ease of synthesis. However, simply driving the model to generate polyimides with the lowest possible SAscores would undesirably limit structural diversity. To mitigate this, we integrate constraints related to the target dielectric constant of the polyimide to ensure a balanced exploration–exploitation trade-off.

Specifically, we first analyze the distribution of dielectric constants obtained from our DFT calculations, computing both the mean and standard deviation. For target dielectric constant values that fall within the most populated regions of this distribution, the model is encouraged to explore broader chemical spaces and generate structurally novel polyimides, as these regions already contain numerous viable candidates. Conversely, for target dielectric constants corresponding to sparsely populated regions, the model is incentivized to conservatively produce polyimides with higher synthesizability to maximize practical feasibility. Accordingly, we introduce two constraints: a synthesizability term s and a dielectric-interval term d. For each generated polyimide i, we define:

$$L_{\mathrm{S}} = \frac{1}{N} \sum_{i=1}^{N} s_i^{d_i} \tag{2}$$

Here, s_i denotes the synthesizability score of polyimide i. The term d_i is defined as a piecewise function that modulates the trade-off between synthesizability and dielectric constant according to its deviation from the mean. Specifically, when the target dielectric constant lies within ± 1 standard deviation from the mean, d_i is set to $-1/3$, encouraging the model to generate polyimides with higher synthesizability scores (i.e., more challenging to synthesize) to promote structural diversity. For target values between one and two standard deviations, d_i is set to $-2/3$, and for values within two to three standard deviations, d_i equals -1. When the target dielectric constant falls beyond three standard deviations from the mean, d_i is set to 1, driving the model to generate polyimides with lower synthesizability scores (i.e., easier-to-synthesize structures) to maximize feasibility in underexplored regions.

The final loss function is:

$$L_{\mathrm{final}} = L_{\mathrm{backbone}} + w_{\mathrm{RL}} \cdot L_{\mathrm{RL}} + w_{\mathrm{S}} \cdot L_{\mathrm{S}} \tag{3}$$

where w_{RL} and w_{S} are weights of L_{RL} and L_{S}, respectively.

DFT details

Dielectric constant calculations were performed based on the Vienna Ab initio Simulation Package (VASP)[60–63]software package, employing the Perdew–Burke–Ernzerhof (PBE)[64,65] functional to compute exchange-correlation energy and potential. The cutoff energy for plane-wave expansion was set to 400 eV. First, the polymer was subjected to geometry optimization. To reduce computational cost, the optimization was carried out in two steps: a coarse optimization and a fine optimization.[66] In the coarse optimization, the convergence criteria for ionic relaxation and electronic energy were set to less than 0.1 eV/Å and 0.2 eV, respectively; in the fine optimization, they were set to less than 0.05 eV/Å and 1×10^{-4} eV, respectively. Based on the optimized molecular structures, the high-frequency dielectric constant was calculated using the DFPT method. The high-frequency static dielectric constant of a material consists of two contributions: the electronic dielectric constant and the ionic dielectric constant:

$$\varepsilon_{\text{static}} = \varepsilon_{\text{elec}}(\infty) + \varepsilon_{\text{ion}}(\omega) \tag{4}$$

In order to obtain the macroscopic dielectric constant, we need to weight the dielectric constants in three directions according to the length of the side of the simulation box in each direction, and obtain the weighted average dielectric constant as the macroscopic dielectric constant:

$$\varepsilon_{\text{avg}} = \frac{\varepsilon_{xx}a + \varepsilon_{yy}b + \varepsilon_{zz}c}{a + b + c} \tag{5}$$

where a, b, and c are the edge lengths of the simulation cell, and ε_{xx}, ε_{yy}, and ε_{zz} are the dielectric constant components in the three directions.

In this work, a single-chain model was employed to construct polymer structures for computation. Ramprasad et al.[67] proposed a method to correct the dielectric constant of polymers obtained from single-chain models using the Maxwell–Garnett equation. In this study, this method was applied to correct the calculated dielectric constants, yielding the

final dielectric constant data. According to the Maxwell-Garnett equation, the dielectric constant of the polymer portion within this composite material can be expressed in the following form:

$$\frac{\varepsilon_\alpha - 1}{1 + (\varepsilon_\alpha - 1) P_\alpha} = \delta \frac{\varepsilon_\alpha^{\text{polymer}} - 1}{1 + \left(\varepsilon_\alpha^{\text{polymer}} - 1\right) P_\alpha} \tag{6}$$

α represents the Cartesian coordinate direction; ε_α represents the dielectric constant of the overall vacuum-composite material; δ denotes the volume fraction of the polymer; $\varepsilon_\alpha^{\text{polymer}}$ denotes the dielectric constant of the polymer portion. The polymer volume was estimated using the electron density cutoff method proposed by Ramprasad et al.,[67] where any spatial region with an electron density above the threshold is considered occupied by the polymer. In this work, a cutoff electron density of $0.001\text{e}/\text{Å}^3$ was used. P_α denotes the geometry-dependent depolarizing factor along the α-direction, which typically depends on the shape of the unit cell and can be approximated using the following formula,[68] taking the depolarization factor along the x-direction as an example:

$$P_x = \frac{abc}{2} \int_0^\infty \frac{ds}{(s + a^2) \sqrt{(s + a^2)(s + b^2)(s + c^2)}} \tag{7}$$

a, b, and c are the lengths of the rectangular prism along the x, y, and z directions.

Data and code availability

The PI1M dataset uesd for training the polymer generation models is publicly available at `https://github.com/RUIMINMA1996/PI1M`.

Acknowledgement

We thank the support from the National Key R&D Program of China (No. **2022YFB3707303**), and the National Natural Science Foundation of China (No. **52293471**). The work is also

supported by the hardware in the Network and Computing Center in Changchun Institute of Applied Chemistry, Chinese Academy of Sciences. A. A.-G. thanks Anders G. Frøseth for his generous support. A. A.-G. acknowledges funding by Natural Resources Canada and the Canada 150 Research Chairs program. A. A.-G. and H. Z. acknowledge funding by the Acceleration Consortium at the University of Toronto. A. A.-G. and H. Z. acknowledges funding by the US Office of Naval Research (Award No.#N000142112137)

Supporting Information Available

The following files are available free of charge.

- Details of Group SELFIES
- Generated polyimides and their dielectric constants

References

(1) Ma, P.; Dai, C.; Wang, H.; Li, Z.; Liu, H.; Li, W.; Yang, C. A Review on High Temperature Resistant Polyimide Films: Heterocyclic Structures and Nanocomposites. *Compos. Commun.* **2019**, *16*, 84–93.

(2) Lyu, Q.; Li, M.; Zhang, L.; Zhu, J. Structurally-Colored Adhesives for Sensitive, High-Resolution, and Non-Invasive Adhesion Self-Monitoring. *Nat. Commun.* **2024**, *15*, 8419.

(3) Toland, A.; Tran, H.; Chen, L.; Li, Y.; Zhang, C.; Gutekunst, W.; Ramprasad, R. Accelerated Scheme to Predict Ring-Opening Polymerization Enthalpy: Simulation-Experimental Data Fusion and Multitask Machine Learning. *J. Phys. Chem. A* **2023**, *127*, 10709–10716.

(4) McDonald, S. M.; Augustine, E. K.; Lanners, Q.; Rudin, C.; Catherine Brinson, L.;

Becker, M. L. Applied Machine Learning as a Driver for Polymeric Biomaterials Design. *Nat. Commun.* **2023**, *14*, 4838.

(5) Yue, T.; He, J.; Li, Y. Polyuniverse: Generation of a Large-Scale Polymer Library Using Rule-Based Polymerization Reactions for Polymer Informatics. *Digital Discovery* **2024**, *3*, 2465–2478.

(6) Gormley, A. J.; Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nature Reviews Materials* **2021**, *6*, 642–644.

(7) Staudinger, H. *Source Book in Chemistry, 1900–1950*; Harvard University Press, 1968; pp 259–264.

(8) Tran, H.; Gurnani, R.; Kim, C.; Pilania, G.; Kwon, H.-K.; Lively, R. P.; Ramprasad, R. Design of Functional and Sustainable Polymers Assisted by Artificial Intelligence. *Nature Reviews Materials* **2024**, *9*, 866–886.

(9) Yan, J.; Zhang, X.; Wang, J.; Hu, W. Effects of Trefoil Knots on the Initiation of Polymer Crystallization. *Macromolecules* **2024**, *57*, 3914–3920.

(10) Chen, T.; Xu, Z.; Chu, X.; Peng, L.; Huang, X.; Li, W. A Simulation Study on the Effect of Tailoring Molecular Architectures on the Mechanical Behaviors of Lamella-Forming BABCB Linear Multiblock Copolymers. *Macromolecules* **2025**, *58*, 6387–6398.

(11) Li, H.; Jin, Y.; Jiang, Y.; Chen, J. Z. Y. Determining the nonequilibrium criticality of a Gardner transition via a hybrid study of molecular simulations and machine learning. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2017392118.

(12) Qing, L.; Wang, X.; Li, S.; Zhang, J.; Jiang, J. Thermodynamic Perturbation Theory for Charged Branched Polymers. *Journal of Chemical Theory and Computation* **2025**, *21*, 333–346, PMID: 39693232.

(13) Xu, X.; Douglas, J. F.; Xu, W.-S. Generalized entropy theory investigation of the relatively high segmental fragility of many glass-forming polymers. *Soft Matter* **2025**, *21*, 2664–2685.

(14) Geng, X.-J.; Li, H.; Yang, X.; An, L.-J.; Müller, M.; Sun, D.-W. Process-Directed Self-Assembly of the Frank-Kasper A15 Structure in Linear, Conformationally Symmetric Block Copolymers. *Phys. Rev. Lett.* **2025**, *134*, 118102.

(15) Yu, X.-K.; Guo, S.-H.; Zhu, Y.-L.; Lu, Z.-Y. Insight into the Regulatory Role of Defect Evolution on the Growth of Single-Crystalline Covalent Organic Frameworks. *ACS nano* **2025**, *19*, 29141–29148.

(16) Xie, C.; Qiu, H.; Liu, L.; You, Y.; Li, H.; Li, Y.; Sun, Z.; Lin, J.; An, L. Machine Learning Approaches in Polymer Science: Progress and Fundamental for a New Paradigm. *SmartMat* **2025**, *6*, e1320.

(17) Gao, L.; Lin, J.; Wang, L.; Du, L. Machine Learning-Assisted Design of Advanced Polymeric Materials. *Accounts of Materials Research* **2024**, *5*, 571–584.

(18) Zhang, K.; Gong, X.; Jiang, Y. Machine Learning in Soft Matter: From Simulations to Experiments. *Advanced Functional Materials* **2024**, *34*, 2315177.

(19) Qiu, H.; Wang, J.; Qiu, X.; Dai, X.; Sun, Z.-Y. Heat-Resistant Polymer Discovery by Utilizing Interpretable Graph Neural Network with Small Data. *Macromolecules* **2024**, *57*, 3515–3528.

(20) Huang, H.; Barati Farimani, A. Multimodal learning of heat capacity based on transformers and crystallography pretraining. *Journal of Applied Physics* **2024**, *135*.

(21) Qiu, H.; Qiu, X.; Dai, X.; Sun, Z.-Y. Design of polyimides with targeted glass transition temperature using a graph neural network. *Journal of Materials Chemistry C* **2023**, *11*, 2930–2940.

(22) Thummalapalli, S. V.; Patil, D.; Ramanathan, A.; Ravichandran, D.; Zhu, Y.; Thippanna, V.; Sobczak, M. T.; Sajikumar, A.; Chambers, L. B.; Guo, S.; Kannan, A. M.; Song, K. Machine learning–enabled direct ink writing of conductive polymer composites for enhanced performance in thermal management and current protection. *Energy Storage Materials* **2024**, *71*, 103670.

(23) Cao, S.; Zhang, Z.; Song, S.; Lan, H.; Gao, L.; Lin, J.; Zhang, C.; Tang, W. On-demand design of materials with enhanced dielectric properties via a machine learning-assisted materials genome approach. *Journal of Materials Chemistry A* **2025**, *13*, 20531–20541.

(24) Liu, B.; Yan, Y.; Liu, M. Harnessing DFT and machine learning for accurate optical gap prediction in conjugated polymers. *Nanoscale* **2025**, *17*, 7865–7876.

(25) Yan, P.; Sun, J.; Zhao, Y.; Deng, W.; Zhang, M.; Li, Y.; Chen, X.; Hu, M.; Tang, J.; Wang, D. Machine Learning-Driven Optimization of Therapeutic Substance Composition for High-Hardness, Fast-Dissolving Microneedles for Androgenetic Alopecia Treatment. *ACS nano* **2025**, *19*, 29301–29315.

(26) Patel, R. A.; Webb, M. A. Data-driven design of polymer-based biomaterials: high-throughput simulation, experimentation, and machine learning. *ACS Applied Bio Materials* **2023**, *7*, 510–527.

(27) Qiu, H.; Zhao, W.; Pei, H.; Li, J.; Sun, Z.-Y. Highly accurate prediction of viscosity of epoxy resin and diluent at various temperatures utilizing machine learning. *Polymer* **2022**, *256*, 125216.

(28) Hu, W.; Jing, E.; Qiu, H.; Sun, Z.-Y. Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning. *Journal of Materials Informatics* **2025**, *5*, N–A.

(29) Zhang, K.; Gong, X.; Jiang, Y. Machine learning in soft matter: from simulations to experiments. *Advanced Functional Materials* **2024**, *34*, 2315177.

(30) Hocken, A.; Huang, S.; Ng, E.; Olsen, B. D. Improving the Optical Sorting of Polyester Bioplastics via Reflectance Spectroscopy and Machine Learning Classification Techniques. *ACS Applied Polymer Materials* **2024**, *6*, 14300–14308.

(31) Dong, Q.; Song, Q.; Tian, K.; Li, W. AI-driven automated construction of block copolymer phase diagrams. *Chinese J. Polym. Sci* **2025**,

(32) Batra, R.; Dai, H.; Huan, T. D.; Chen, L.; Kim, C.; Gutekunst, W. R.; Song, L.; Ramprasad, R. Polymers for Extreme Conditions Designed Using Syntax-Directed Variational Autoencoders. *Chemistry of Materials* **2020**, *32*, 10489–10500.

(33) Liu, Z.; Liu, Y.-X.; Yang, Y.; Li, J. Template Design for Complex Block Copolymer Patterns Using a Machine Learning Method. *ACS Applied Materials & Interfaces* **2023**, *15*, 31049–31056, PMID: 37335810.

(34) Qiu, H.; Sun, Z.-Y. On-demand reverse design of polymers with PolyTAO. *npj Computational Materials* **2024**, *10*, 273.

(35) Gurnani, R.; Kamal, D.; Tran, H.; Sahu, H.; Scharm, K.; Ashraf, U.; Ramprasad, R. polyG2G: A Novel Machine Learning Algorithm Applied to the Generative Design of Polymer Dielectrics. *Chemistry of Materials* **2021**, *33*, 7008–7016.

(36) Liu, D.-F.; Zhang, Y.-X.; Dong, W.-Z.; Feng, Q.-K.; Zhong, S.-L.; Dang, Z.-M. High-Temperature Polymer Dielectrics Designed Using an Invertible Molecular Graph Generative Model. *Journal of Chemical Information and Modeling* **2023**, *63*, 7669–7675, PMID: 38061777.

(37) Kim, S.; Schroeder, C. M.; Jackson, N. E. Open Macromolecular Genome: Generative Design of Synthetically Accessible Polymers. *ACS Polymers Au* **2023**, *3*, 318–330.

(38) Xu, C.; Wang, Y.; Barati Farimani, A. TransPolymer: a Transformer-based language model for polymer property predictions. *npj Computational Materials* **2023**, *9*, 64.

(39) Kuenneth, C.; Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nature communications* **2023**, *14*, 4099.

(40) Liu, Z.; Chai, Y.; Li, J. Toward Automated Simulation Research Workflow through LLM Prompt Engineering Design. *Journal of Chemical Information and Modeling* **2024**, *65*, 114–124.

(41) Ramos, M. C.; Collison, C. J.; White, A. D. A review of large language models and autonomous agents in chemistry. *Chemical science* **2025**, *16*, 2514–2572.

(42) Zhang, S.; Li, S.; Song, S.; Zhao, Y.; Gao, L.; Chen, H.; Li, H.; Lin, J. Deep Learning-Assisted Design of Novel Donor–Acceptor Combinations for Organic Photovoltaic Materials with Enhanced Efficiency. *Advanced Materials* **2025**, *37*, 2407613.

(43) Ma, R.; Luo, T. PI1M: a benchmark database for polymer informatics. *Journal of Chemical Information and Modeling* **2020**, *60*, 4684–4690.

(44) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100.

(45) Cheng, A. H.; Cai, A.; Miret, S.; Malkomes, G.; Phielipp, M.; Aspuru-Guzik, A. Group SELFIES: a robust fragment-based molecular string representation. *Digital Discovery* **2023**, *2*, 748–758.

(46) Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. SYBYL line notation (SLN): A versatile language for chemical structure representation. *Journal of chemical information and computer sciences* **1997**, *37*, 71–79.

(47) Gakh, A. A.; Burnett, M. N. Modular chemical descriptor language (MCDL): composition, connectivity, and supplementary modules. *Journal of chemical information and computer sciences* **2001**, *41*, 1494–1499.

(48) Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D. InChI, the IUPAC international chemical identifier. *Journal of cheminformatics* **2015**, *7*, 23.

(49) Schneider, L.; Walsh, D.; Olsen, B.; de Pablo, J. Generative BigSMILES: an extension for polymer informatics, computer simulations & ML/AI. *Digital Discovery* **2024**, *3*, 51–61.

(50) Cheng, A. H.; Cai, A.; Miret, S.; Malkomes, G.; Phielipp, M.; Aspuru-Guzik, A. Group SELFIES: a robust fragment-based molecular string representation. *Digital Discovery* **2023**, *2*, 748–758.

(51) Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Wang, Z.; Shysheya, A.; Crabbé, J.; Ueda, S.; others A generative model for inorganic materials design. *Nature* **2025**, *639*, 624–632.

(52) Vogel, G.; Weber, J. M. Inverse Design of Copolymers Including Stoichiometry and Chain Architecture. *Chem. Sci.* **2025**, *16*, 1161–1178.

(53) Zhang, S.; Li, S.; Song, S.; Zhao, Y.; Gao, L.; Chen, H.; Li, H.; Lin, J. Deep Learning-Assisted Design of Novel Donor–Acceptor Combinations for Organic Photovoltaic Materials with Enhanced Efficiency. *Advanced Materials* **2025**, *37*, 2407613.

(54) Ohno, M.; Hayashi, Y.; Zhang, Q.; Kaneko, Y.; Yoshida, R. SMiPoly: generation of a synthesizable polymer virtual library using rule-based polymerization reactions. *Journal of Chemical Information and Modeling* **2023**, *63*, 5539–5548.

(55) Yin, Q.; Qin, Y.; Lv, J.; Wang, X.; Luo, L.; Liu, X. Reducing Intermolecular Friction Work: Preparation of Polyimide Films with Ultralow Dielectric Loss from MHz to THz Frequency. *Industrial & Engineering Chemistry Research* **2022**, *61*, 17894–17903.

(56) Hasegawa, M.; Hishiki, T. Poly(ester imide)s Possessing Low Coefficients of Thermal

Expansion and Low Water Absorption (V). Effects of Ester-linked Diamines with Different Lengths and Substituents. *Polymers* **2020**, *12*.

(57) Zhang, M.; Miao, J.; Xu, Y.; Wang, Z.; Yan, J. Colorless Polyimides from Fluorinated Ladder Diamines Containing Norbornyl Benzocyclobutene Segments. *Macromolecules* **2022**, *55*, 7992–8001.

(58) Li, Y.; Sun, G.; Zhou, Y.; Liu, G.; Wang, J.; Han, S. Progress in low dielectric polyimide film–A review. *Progress in Organic Coatings* **2022**, *172*, 107103.

(59) Bei, R.; Chen, K.; Liu, Q.; He, Y.; Li, C.; Huang, H.; Guo, Q.; Chi, Z.; Xu, J.; Chen, Z.; Liu, S.; Zhang, Y. Relationship among the Water Adsorption, Polymer Structure, and High-Frequency Dissipation Factor: Precise Analysis of Water Adsorption of Low-Dielectric Constant Polyimide Films. *Macromolecules* **2024**, *57*, 2142–2153.

(60) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical review* **1964**, *136*, B864.

(61) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.

(62) Gajdoš, M.; Hummer, K.; Kresse, G.; Furthmüller, J.; Bechstedt, F. Linear optical properties in the projector-augmented wave methodology. *Physical Review B—Condensed Matter and Materials Physics* **2006**, *73*, 045112.

(63) Baroni, S.; Resta, R. Ab initio calculation of the macroscopic dielectric constant in silicon. *Physical Review B* **1986**, *33*, 7017.

(64) Blöchl, P. E. Projector augmented-wave method. *Physical review B* **1994**, *50*, 17953.

(65) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b* **1999**, *59*, 1758.

(66) Takahashi, A.; Kumagai, Y.; Miyamoto, J.; Mochizuki, Y.; Oba, F. Machine learning models for predicting the dielectric constants of oxides based on high-throughput first-principles calculations. *Physical Review Materials* **2020**, *4*, 103801.

(67) Wang, C.; Pilania, G.; Ramprasad, R. Dielectric properties of carbon-, silicon-, and germanium-based polymers: A first-principles study. *Physical Review B—Condensed Matter and Materials Physics* **2013**, *87*, 035103.

(68) Sihvola, A. H. *Electromagnetic mixing formulas and applications*; Iet, 1999.