JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

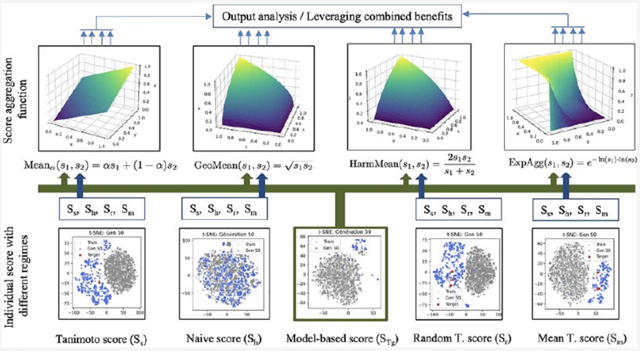
Multi-Score Reinforcement Learning for High-Tg Polyimide Design

Aymar Tchagoue,* Véronique Eglin, Jean-Marc Petit, Sébastien Pruvost, Jannick Duchet-Rumeau, and Jean-François Gérard

Cite This: https://doi.org/10.1021/acs.jcim.5c02807 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: This study explores strategies to guide the generation of polyimides with high glass transition temperatures (T_g > 750 K) through reinforcement learning. We present a systematic computational framework for analyzing and combining multiple scoring functions into a single score in reinforcement learning (RL) for molecular design. Rather than relying solely on a single scoring function based on a predictive model, we examine a range of complementary scores, including a novel naïve high-T_g score and various Tanimoto similarity-based scores. We analyze these scores both individually and in combination with the predictive model-based score in order to assess their influence on the structural diversity and quality of the generated polymers. In addition, we investigate several methods for combining scores, such as arithmetic, geometric, and harmonic means, as well as a novel exponential−logarithmic function, referred to as ExpAgg. We evaluate how these aggregation strategies affect the outcomes of molecular generation across different reinforcement learning configurations. Our findings show that the choice of score combination method significantly impacts both the quality and diversity of generated polymers. The proposed ExpAgg achieves superior performance in multiple settings, revealing nontrivial interactions between score compatibility and model convergence. While the predictive model exhibits underestimation in the out-of-distribution region (>800 K), our multiscore framework successfully generates chemically reasonable high-T_g candidates. Based on these insights, we provide practical guidelines for selecting aggregation functions when fusing two scores. This case study on high-T_g polyimide generation demonstrates how score aggregation strategies influence molecular RL outcomes; broader generalizability to other molecular design tasks remains to be investigated. This work emphasizes the importance of moving beyond simple weighted averages in order to enhance targeted molecular design.


Output analysis / Leveraging combined benefits
Score aggregation function
$\mathrm{Mean}_\alpha(s_1, s_2) = \alpha s_1 + (1-\alpha)s_2$
$\mathrm{GeoMean}(s_1, s_2) = \sqrt{s_1 s_2}$
$\mathrm{HarmMean}(s_1, s_2) = \frac{2s_1 s_2}{s_1 + s_2}$
$\mathrm{ExpAgg}(s_1, s_2) = e^{-\ln(s_1)\cdot\ln(s_2)}$
S_s, S_h, S_r, S_m
Individual score with different regimes
t-SNE: Gen 50
t-SNE: Génération 50
Train
Gen 50
Target
Tanimoto score (S_s)
Naive score (S_h)
Model-based score (S_{T_g})
Random T. score (S_r)
Mean T. score (S_m)

1. INTRODUCTION

The design of high-performance polymers with targeted properties, such as a high glass transition temperature (T_g), remains a long-standing challenge in materials science. Polyimides, in particular, are widely recognized for their thermal stability, mechanical strength, and chemical resistance, making them promising candidates for high-temperature applications in aerospace, electronics, and energy storage.[1,2]

Recent progress in deep learning and generative models has opened new possibilities for data-driven polymer design.[3,4] In this context, reinforcement learning (RL) has proven to be a powerful framework for optimizing candidate molecules toward specific properties.[5] Typically, this is achieved by defining a scoring function that guides molecular generation, often based on the predicted value of a key target property such as T_g. Designing an effective scoring function becomes particularly challenging when multiple molecular criteria must be considered simultaneously, such as synthetic accessibility, structural similarity, or novelty. Although multiobjective RL[6] has been proposed to address this challenge,[7,8] most existing approaches rely on scalarization with simple linear combinations of scores,[9,10] usually weighted averages, without thoroughly evaluating how the method of combining scores influences the quality and diversity of the generated molecules.

In this work, we systematically investigate how the scalarization (aggregation) function of multiple scores into a single score affects the generation of high-T_g polyimides (Figure 1). We consider several types of scores, including predictive model-based score, Tanimoto similarity scores, and Shapley-inspired importance score, each of which exhibits specific and complementary advantages. These scores are combined using various mean functions: arithmetic (*Mean*$_\alpha$), geometric (*GeoMean*), harmonic (*HarmMean*), and a novel exponential-

Received: November 16, 2025
Revised: January 20, 2026
Accepted: February 10, 2026

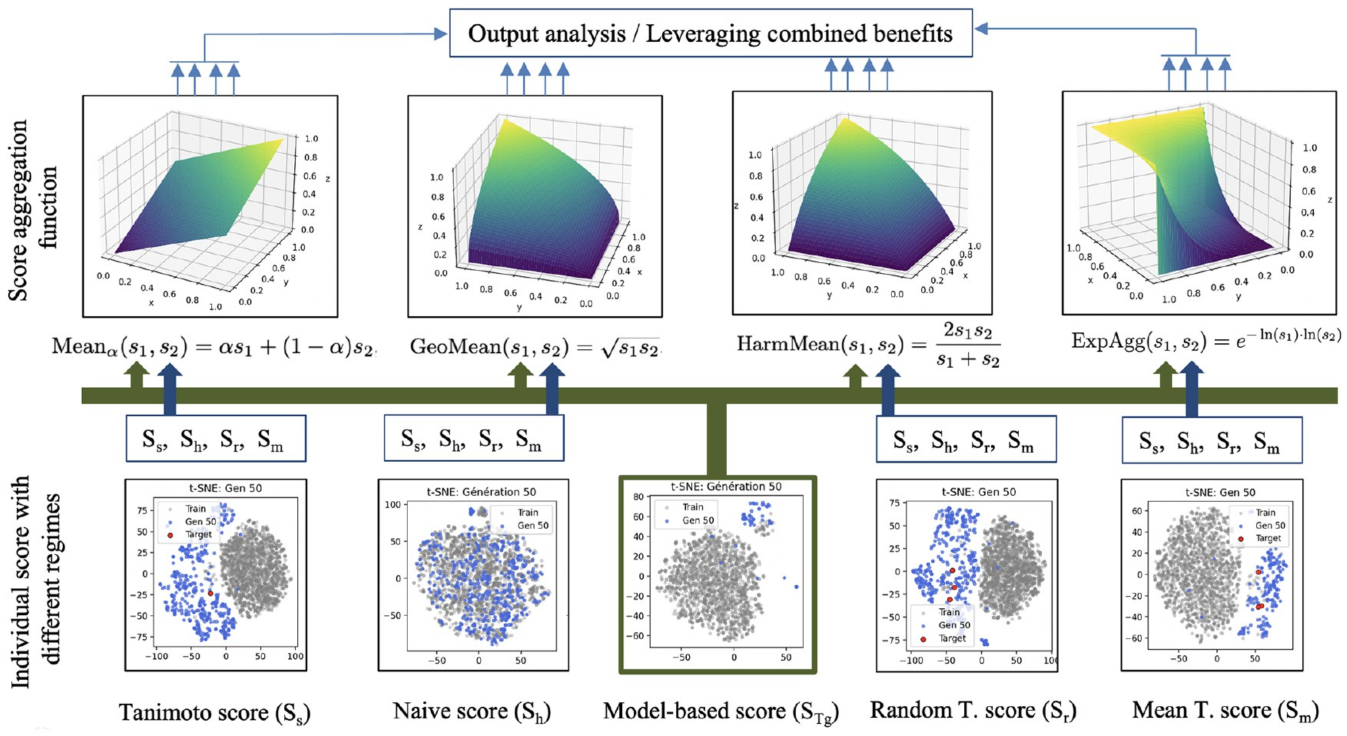

Output analysis / Leveraging combined benefits
Score aggregation function
$\mathrm{Mean}_{\alpha}(s_1, s_2) = \alpha s_1 + (1-\alpha) s_2$
$\mathrm{GeoMean}(s_1, s_2) = \sqrt{s_1 s_2}$
$\mathrm{HarmMean}(s_1, s_2) = \dfrac{2 s_1 s_2}{s_1 + s_2}$
$\mathrm{ExpAgg}(s_1, s_2) = e^{-\ln(s_1)\cdot\ln(s_2)}$
S_s, S_h, S_r, S_m
Individual score with different regimes
t-SNE: Gen 50
t-SNE: Génération 50
Train
Gen 50
Target
Tanimoto score (S_s)
Naive score (S_h)
Model-based score (S_{Tg})
Random T. score (S_r)
Mean T. score (S_m)

Figure 1. Graphical abstract: the model-based score (S_{Tg}) is combined with four alternative scoring functions, highlighting the benefits of combining complementary exploration and exploitation regimes, as well as the impact of the aggregation function choice.

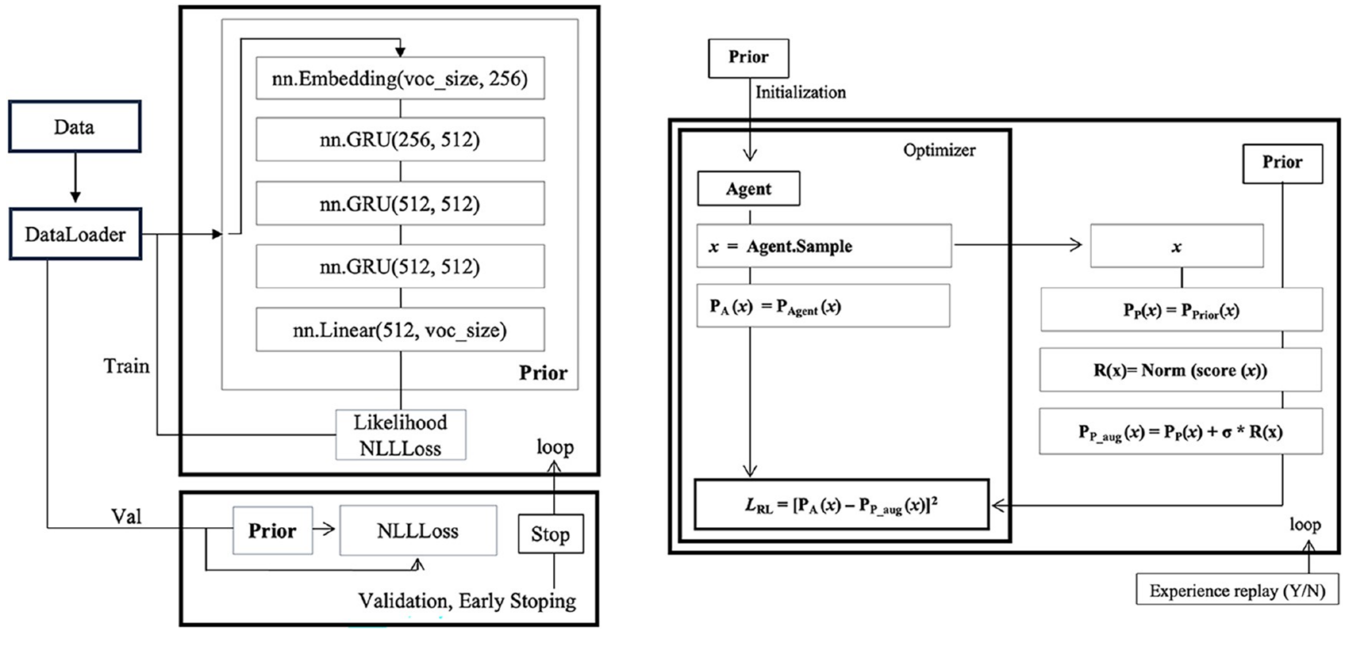

Data
DataLoader
nn.Embedding(voc_size, 256)
nn.GRU(256, 512)
nn.GRU(512, 512)
nn.GRU(512, 512)
nn.Linear(512, voc_size)
Prior
Likelihood
NLLLoss
Train
loop
Val
Prior
NLLLoss
Stop
Validation, Early Stoping
Prior
Initialization
Optimizer
Agent
x = Agent.Sample
$\mathrm{P_A}(x) = \mathrm{P_{Agent}}(x)$
$L_{\mathrm{RL}} = [\mathrm{P_A}(x) - \mathrm{P_{P\_aug}}(x)]^2$
Prior
x
$\mathrm{P_P}(x) = \mathrm{P_{Prior}}(x)$
$\mathrm{R}(x) = \mathrm{Norm\ (score\ }(x))$
$\mathrm{P_{P\_aug}}(x) = \mathrm{P_P}(x) + \sigma * \mathrm{R}(x)$
loop
Experience replay (Y/N)

Figure 2. Overview of the REINVENT architecture used in this study.

logarithmic aggregation *(ExpAgg)*, each of which plays a distinct role in balancing exploration and exploitation by modulating the influence of individual score components on the final aggregated value.

Our contribution is 2-fold: first, we demonstrate improved performance in high-T_g polyimide generation through multi-score optimization; second, we provide a systematic analysis of score aggregation strategies that may inform future molecular design studies. We acknowledge that the predictive model exhibits underestimation in the out-of-distribution region (>800 K), a limitation we discuss in Section 9.

The subsequent sections are organized as follows: we first review related literature in Section 2. The data set and T_g prediction model are presented in Section 3. In Section 4, we detail the various scoring strategies used to guide generation. The polyimides generation are analyzed in Section 5. Section 6 explores the impact of score combination methods. Section 7 provides a chemical characterization of the generated polyimides. Key findings are summarized in Section 8, followed by a discussion of implications in Section 9 and concluding remarks in Section 10. The data and software are accessible in the Declaration section 11.

2. RELATED WORK AND METHODOLOGICAL CONTEXT

2.1. Generative Models for Molecular Design

Most existing studies in molecular RL have focused on improving model architectures or on tuning reward functions for specific applications. By contrast, few have systematically investigated how the mathematical form of score aggregation itself influences learning dynamics, convergence, and molecular diversity.

To contextualize this problem, it is useful to review the main families of generative models applied to molecular design, including variational autoencoders (VAE), recurrent neural networks (RNNs, GRUs and LSTMs), chemical language models, and diffusion-based generators.[11−14] Sequence-based approaches (RNN, GRU, LSTM) have demonstrated strong performance in optimization tasks using RL for targeted property tuning.[15,16] Conversely, chemical language models and pretrained embeddings (such as PolyBERT) provide transferable molecular representations for property prediction but are not inherently generative for inverse-design applications.[12]

The comprehensive benchmarking study by Yue et al.[15] compared six generative models (VAE, AAE, ORGAN, CharRNN, REINVENT, and GraphINVENT) for inverse polyimide design targeting high-T_g. Their results highlight REINVENT as a particularly effective RL-based approach for generating high-T_g polyimides when sufficient data are available. Its architecture explicitly separates the Prior and Agent models, enabling stable policy optimization and efficient incorporation of task-specific reward functions. The next section provides an overview of the REINVENT framework and its adaptation to polymer design.

2.2. REINVENT Framework for Polyimide Generation

The REINVENT architecture was first introduced in[17] and later refined in REINVENT 4.0.[18] In this work, we adopt the REINVENT implementation from,[16] originally developed for drug-like molecules. It provides the foundation required to evaluate the impact of different scoring functions and aggregation strategies on molecular generation, which constitutes the main focus of this work.

2.2.1. Prior Model. Figure 2 illustrates the REINVENT pipeline. On the left, the *Prior* model consists of an embedding layer, three stacked GRU layers, and an output linear layer. The training objective is to embed SMILES sequences and reconstruct similar ones at the output, thereby enhancing the model's understanding of structural patterns in latent space. The prior model is trained to minimize the negative log-likelihood (NLL) (eq 1)

$$\mathcal{L}_{\mathrm{NLL}} = -\sum_{t=1}^{n} \log P(x_t|x_{<t}) \tag{1}$$

where $x = (x_1, x_2, ..., x_n)$ is the sequence of n tokens representing a polymer SMILES string, $x_{<t} = (x_1, x_2, ..., x_{t-1})$ are the preceding tokens, and $P(x_t \mid x_{<t})$ is the conditional probability of token x_t given its context. The training process is controlled with an early stopping implementation to prevent overfitting. Once trained, the *Prior* is capable of generating coherent and diverse polymers, reflecting the distribution of its training set. Its weights are then used to initialize the *Agent*, which is subsequently fine-tuned via RL to maximize a task-specific reward.

2.2.2. Score vs Reward. We distinguish between the **score** and the **reward** functions in this work. The score $S(x) \in [0, 1]$ of a polymer x measures how well it satisfies a specific property, ranging from 0 (not satisfied) to 1 (fully satisfied). The reward $R(x)$ is derived from the score using a normalization function (eq 2)

$$R(x) = \mathrm{Norm}(S(x)) \tag{2}$$

The normalization converts raw scores into meaningful reward values. Let S_g be the set of generated polymer scores at a given fine-tuning step, and $S_x \in S_g$ the score of a given polymer. With a slope parameter $\tau \in [0, 1]$, the normalized score is defined by the following sigmoid-based scaling (eq 3)

$$\mathrm{Norm}(S_x) = \frac{2}{1 + \exp(-\lambda(S_x - \overline{S_g}))} - 1,$$
$$\text{where}\lambda = -\frac{1}{\max(S_g) - \overline{S_g}} \cdot \ln\left(\frac{2}{\tau + 1} - 1\right) \tag{3}$$

2.2.3. Reinforcement Learning Loss. The negative log-likelihoods of a polymer x under the Prior and Agent models are denoted by $P_{\mathrm{Prior}}(x)$ and $P_{\mathrm{Agent}}(x)$, respectively. The augmented log-likelihood of the Prior model, computed using a weighting factor of σ, is denoted by $P_{\mathrm{Prior\_aug}}(x)$. The reinforcement learning loss ($\mathcal{L}_{\mathrm{RL}}$) is then defined as (eq 4)

$$P_{\mathrm{Prior}\_aug}(x) = P_{\mathrm{Prior}}(x) + \sigma \cdot R(x),$$
$$\mathcal{L}_{\mathrm{RL}} = [P_{\mathrm{Agent}}(x) - P_{\mathrm{Prior}\_aug}(x)]^2 \tag{4}$$

The optimization of the Agent's weights is performed after each generation step. The scores are therefore not interpreted in isolation, as normalization allows them to reflect the overall observations of the generation step by considering both the maximum ($max(S_g)$) and the mean value ($\overline{S_g}$).

The reinforcement learning loss defined above guides the Agent by comparing its predictions with the augmented Prior likelihood, which incorporates the normalized reward. This reward signal directly affects how the Agent balances exploring new regions of chemical space and exploiting known high-scoring molecules. Understanding this trade-off between exploration and exploitation is essential, as it determines both the diversity and quality of the generated polymers. We therefore next discuss these concepts in detail.

2.3. Exploration vs Exploitation in RL

In reinforcement learning, the process of discovering new information about the search space by deliberately selecting suboptimal actions is known as *exploration*, whereas relying on existing knowledge to obtain the best possible outcomes is referred to as *exploitation*.[19] For instance, suppose an agent is trained with a Tanimoto score to generate candidates that resemble an unknown target molecule within its chemical space. It adopts two complementary strategies. First, it generates structures that are already known within its learned space; the reward function subsequently guides it toward the formulations that best match the target, leading the agent to focus on regions of the chemical space that yield satisfactory results. Second, while concentrating on these relevant regions, the agent also probes beyond them, tentatively exploring more distant areas of the chemical landscape that might reveal new optimal solutions. These two behaviors, exploitation and exploration, operate jointly: the agent exploits promising regions to refine its generations while simultaneously exploring uncharted areas that may yield higher rewards. Achieving the right balance between these two learning modes is essential for effective optimization. If the agent focuses exclusively on exploiting known optima, it may fail to discover other potentially superior regions that remain unexplored; conversely, if it constantly explores without consolidation, it may forget what has already proven to be effective. Experience Replay (ER) stores past high-quality experiences for reuse during training, which can enhance learning by supporting a better balance between exploration and exploitation.[20] When multiple objectives are involved, such as optimizing both predicted T_g and structural diversity, the agent must balance exploration and exploitation across several scoring functions simultaneously. This motivates the study of multi-objective reinforcement learning and the strategies for combining scores, which are discussed in the next section.

2.4. Scoring and Multi-Objective Reward Design

In REINVENT, the Agent is fine-tuned via reinforcement learning using reward signals derived from scoring functions that measure how well generated molecules satisfy desired properties. The design and combination of these scores critically influence learning dynamics, molecular diversity, and the balance between exploration and exploitation. While previous work often focused on improving the Prior model or single-objective rewards, this study systematically investigates multiple scoring functions and their aggregation. A wide range of scoring functions has been proposed in the literature, depending on the target application. In drug discovery, popular metrics include Tanimoto similarity (eq 5), Jaccard distance, quantitative drug-likeness (QED) score, number of rings, and others. For instance, the Tanimoto similarity[21] between a generated molecule A and a target B is defined as

$$T(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{5}$$

where $|A \cap B|$ and $|A \cup B|$ represent, respectively, the number of shared active bits (intersection) and the total number of active bits (union) in their molecular fingerprints. These bits typically correspond to chemical substructures. In polymer design, predictive models are often directly integrated into the scoring function. In ref 15, the score is derived from a machine learning predictor of the T_g; it guides the generative model toward producing polymers with higher predicted T_g. Research in this area generally focuses on improving either the Prior model or the reward formulation. For example,[22] introduced a memory-assisted RL scheme, in which the score is weighted by historical performance to encourage molecular diversity. Similarly,[23] proposed a multiproperty predictor to build a composite reward function targeting multiple polymer attributes. Along the same lines, Xu et al.[24] developed multiobjective property predictors built from four machine-learning models, combined with a fragment-based generative architecture using chemically meaningful polyimide fragments as building blocks. Their results demonstrate that RL can efficiently generate polyimides with well-balanced and experimentally validated properties.

When multiple reward functions are required, a common strategy in reinforcement learning is to combine them into a single scalar reward, transforming a multiobjective problem into a single-objective one. The most widely used scalarization method is a simple weighted average, but this approach can be limited, as the model receives only one aggregated signal to manage exploration and exploitation across multiple objectives.

Depending on the nature of the scores and the model architecture, the scalarization process can affect optimization stability, the balance between exploration and exploitation, and the diversity of generated structures.[7,25,26] To overcome the limitations of linear scalarization, alternative aggregation schemes have been proposed. Pareto-based optimization, for instance, avoids the need for predefined weights and reveals trade-offs between competing objectives,[25,27] while nonlinear scalarization methods such as the Chebyshev approach have shown competitive or superior performance.[9]

In this work, we analyze both linear and nonlinear aggregation methods ($Mean_\alpha$, HarmMean, GeoMean, ExpAgg) of score functions into a single score which is normalized and used in the loss function as the reward (eq 4). Unlike traditional multiobjective formulations where distinct properties computed as separate rewards are optimized, all individual scores here target the same design objective. Our goal is to understand how different score aggregation formulas influence the balance between exploration and exploitation and, consequently, the overall generative performance.

To evaluate the impact of score aggregation strategies in polyimide design, we propose here a dedicated predictive model for the glass transition temperature (T_g). This model serves as the foundation for the scoring function that guides the REINVENT agent in the subsequent generative tasks.

3. POLYIMIDE T_G PREDICTION MODEL

Since the scoring function plays a central role in RL-based molecular generation, we first develop and validate a dedicated T_g predictor for polyimides. This model serves as the foundation of the scoring function used to guide the generation process.

3.1. Polyimide Data set Description

This study uses the data set introduced in,[28] which contains synthetic polyimide repeating units represented in SMILES format, along with their associated T_g values. We refer to this as "synthetic polymer data" because the structures, while chemically plausible, have not necessarily been synthesized experimentally. The data set was constructed in two main stages. First, large libraries of diamines and dianhydrides were generated using predefined combinatorial rules. These building blocks were then systematically combined to produce millions of distinct polyimide repeating units. Each unit consists of one diamine and one dianhydride, typically organized into alternating sequences of flexible linkers and bulky moieties. To ensure chemical relevance, the most frequently occurring linkers and moieties were identified and selected for the generation process. This strategy enabled the creation of a structurally diverse and chemically coherent library of candidate polyimides for both generation and property prediction tasks.

Furthermore, the T_g values of polyimides were calculated by Askadskii's computational scheme, implemented in the "Cascade" program.[29,30] Figure 3 shows that the T_g values span a wide range, reflecting high thermal variability. However, only a small fraction of the data set contains extreme values ($T_g >$ 800 K), with just 15 out of 500,000 samples. This highlights the rarity of the high-T_g molecules targeted in this work.

The average length of the SMILES strings in the data set is 140 characters, with a minimum of 81 and a maximum of 220. The polymer SMILES representation using the standard convention

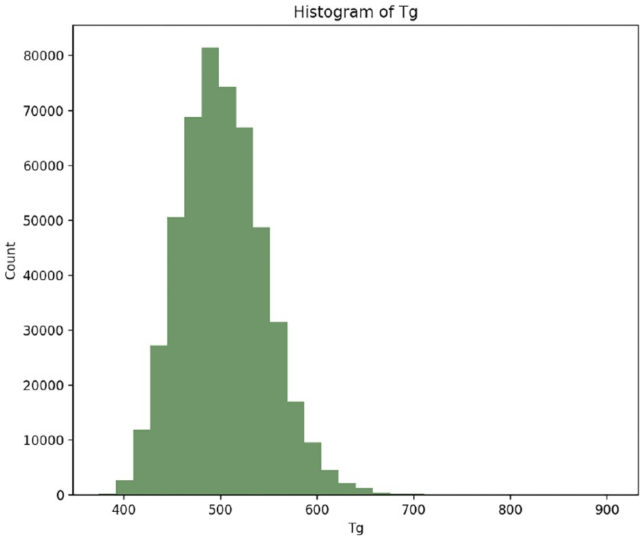


Figure 3. Histogram of T_g values in the global polyimide data set containing more than 6 million entries.

of ‘*’ is replaced by the symbol ‘I’ to indicate repeating unit boundaries, as used in the database of Volgin et al.[28] Since no iodine molecules are present, this does not pose a problem. The SMILES vocabulary used is given below (eq 6)

$$\text{Vocabulary} = [\%, (,), -, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \\ =, \text{C}, \text{F}, \text{I}, \text{L}, \text{N}, \text{O}, \text{S}, \text{c}, \text{n}, [\text{nH}]] \quad (6)$$

3.2. Predictive Model Description

Various strategies exist for predicting polymer properties directly from SMILES representations. Among them, embedding-based approaches using language models such as PolyBERT[12,31] have gained popularity in recent years. However, these models typically lack interoperability and interpretability. In this work, we deliberately opted for a more transparent and interpretable approach, employing a Random Forest regressor as the predictive model and Morgan fingerprints[32,33] as molecular descriptors, with parameters: radius $r = 2$ and 2048 bits.

The predictive model was trained on 80% of a subset of the data set, hereafter referred to as the “prediction set”, consisting of 50,000 representative polyimides and evaluated on the remaining 20%. The Random Forest uses 500 estimators and a random seed of 42 to ensure reproducibility. Figure 4 presents the model’s performance, the coefficient of determination (R^2 = 97%), the root mean squared error (RMSE = 7.08 K), and the mean absolute error (MAE = 5 K) on the test set.

We observe that the model provides accurate predictions across most of the T_g range. However, for very high values (above 800 K), it tends to systematically underestimate T_g, likely due to the limited number of such extreme cases in the data set. As a result, polyimides with high T_g values may have underestimated predictions, even within their uncertainty interval (±MAE = ±5 K).

To further assess the model’s ability to generalize to out-of-distribution, we conducted experiments using different training sets distributed across five T_g intervals: $A = [400, 500)$, $B = [500, 600)$, $C = [600, 700)$, and $D = [700, 800)$, and $E = [800, 900]$. Four scenarios were tested: the model was trained on A and tested on B; trained on $A + B$ and tested on C; then trained on $A + B + C$ and tested on D, and finally trained on $A + B + C + D$ and tested on E. Figure 5 illustrates the performance of the Random Forest model in this realistic setup, where the training data cover a limited thermal range and the model is challenged to extrapolate beyond its initial training distribution.

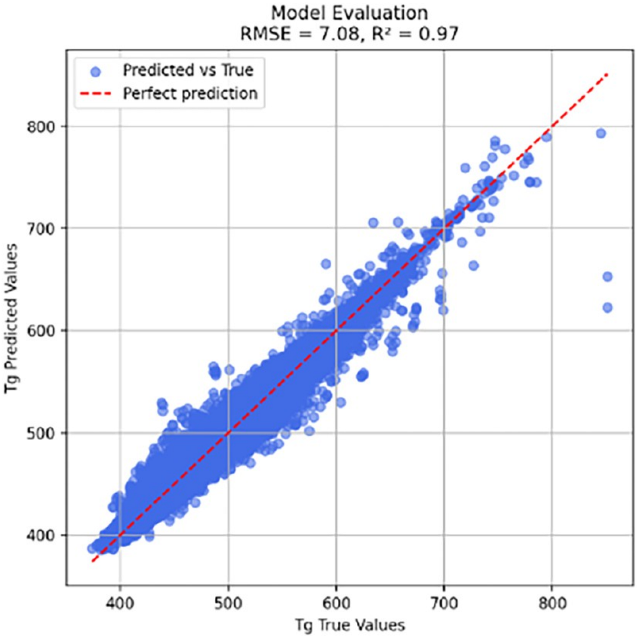


Figure 4. Performance of the Random Forest model in predicting T_g values of polyimides test set.

Overall, the extrapolation results (Figure 5) show that the predicted T_g values never exceed the highest T_g values observed during training. A similar trend was observed for models trained on PolyBERT embeddings. Therefore, for our baseline Random Forest model (Figure 4), the predictions remain accurate within the [400, 800] K range, but tend to lose reliability beyond this interval due to the limited amount of available data in the [800, 900] K range.

3.2.1. Evaluation of the Predictive Model Using External Experimental Polyimide Data.

Yu et al.[34] recently reported a Random Forest model (RMSE_{Yu} = 27.45 K) for predicting the T_g of polyimides using an experimental data set they provided. We evaluate here the performance of our predictive model, trained exclusively on synthetic data, on this independent experimental data. Unlike our data set vocabulary (eq 6), their experimental set contains additional elements (Br, F, P, and Si). Additionally, as their data set does not include the element ‘I’, consistency was ensured by mapping ‘*’ to ‘I’.

This evaluation must be interpreted with caution due to uncertainties in experimental T_g measurements (which were not provided) and the structural differences between the prediction set and the experimental data, as illustrated by the t-SNE projection in Figure 6. Since we used a global Morgan fingerprint for embedding, substructures not present in the training set were encoded as 0 in our Random Forest model during training.

To account for this domain shift, we evaluated model performance on subsets of the experimental data defined by increasing minimum Tanimoto similarity thresholds to the prediction set (Table 1), such that each experimental polyimide has at least one structural analogue in the prediction set above the given threshold.

As structural similarity increases, predictive performance systematically improves, with the RMSE decreasing from 82.37 to 30.74 K, corresponding to a 62.7% reduction. These results,

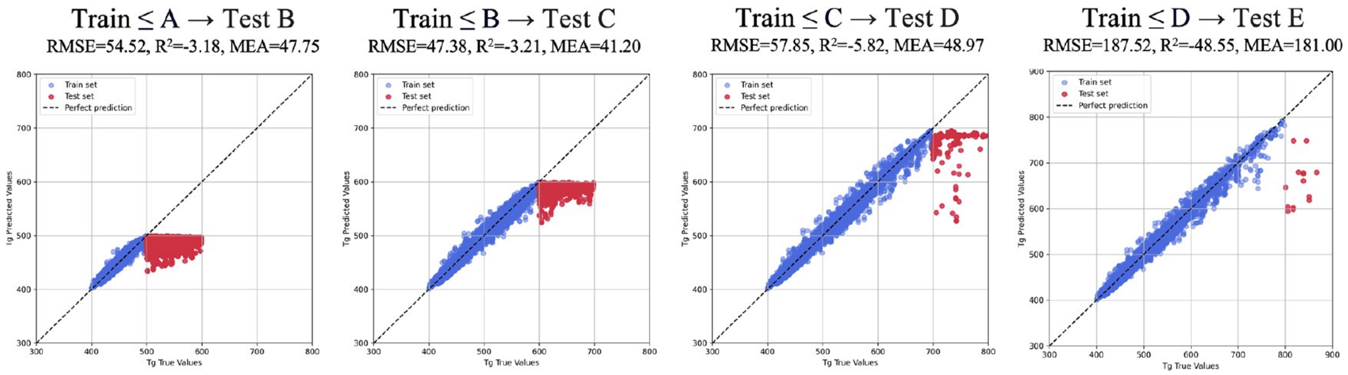


Figure 5. Out-of-distribution performance analysis of the Random Forest model.

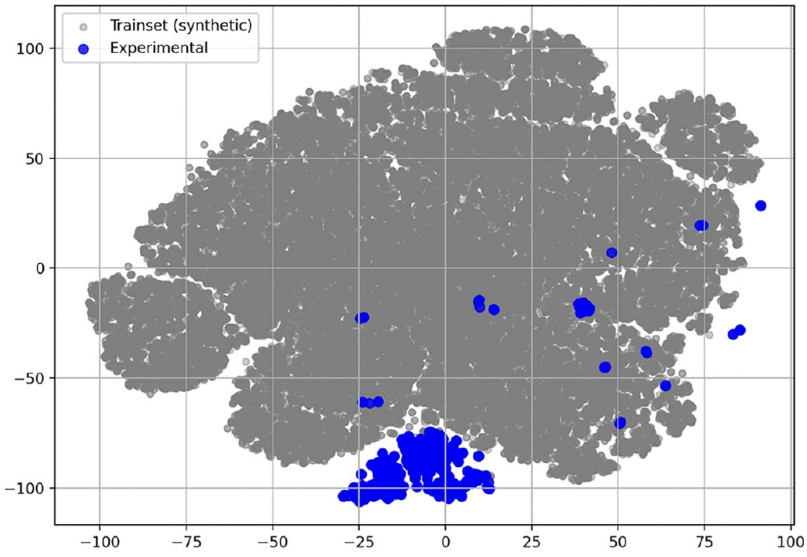


Figure 6. t-SNE projection of polyimide structures from the synthetic prediction set and the external experimental data set.

Table 1. Performance of Our Model on Subsets of External Experimental Polyimides at Varying Minimum Structural Similarity Thresholds

threshold	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Samples	1307	1307	1299	1274	1078	714	438	243	112	29	6
RMSE	**82.37**	82.37	81.60	80.54	66.01	53.75	48.02	46.71	42.50	39.50	**30.74**

obtained on previously unseen structures, indicate that, on one hand, the model is more reliable within its effective domain of applicability (i.e., the same types of substructures), and, on the other hand, the predictions on the experimental data show that the model is reasonably consistent with reality, as compared to a model trained on an actual experimental data set ($\mathrm{RMSE}_{\mathrm{Yu}}$ = 27.45 K). Even if a direct comparison is not possible, this provides an overall sense of accuracy.

3.2.2. Concrete Example of Predictions on External Experimental Polyimide Data. Table 2 provides representative examples of T_g predictions alongside their experimentally reported values from the literature.

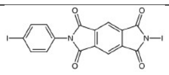
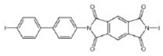
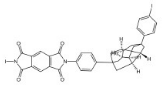
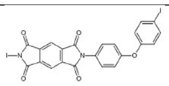
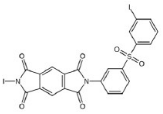
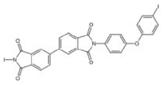
3.2.3. High-T_g Structures (T_g^{Exp} > 750 K). The first section of the table includes structures with high experimental T_g values. Specifically, structures (a) and (c) are reported only as lower bounds (>773.15 K),[35,36] while structure (b) is reported at 850.15 K.[37] As a result, these structures may fall outside the applicability domain of our predictive model, with true T_g potentially exceeding 800 K. Nevertheless, the predicted values remain good and consistent with the earlier discussion of model performance, and they lie within the targeted range of interest (>750 K).

3.2.4. Moderate-T_g Structures (T_g^{Exp} < 750 K). The second section of the table presents structures with lower experimental T_g values. The model predictions are in good agreement with the experimental data, demonstrating reliable performance.

3.3. Model Interpretation

To interpret the predictive model's decision-making process, we applied the SHAP (SHapley Additive exPlanations) framework,[38] which provides both local and global explanations by assigning importance values to each input feature, in this case, each bit of the Morgan fingerprint. The SHAP analysis yielded, for each of the 2048 fingerprint bits, both the *mean signed* and *mean absolute* contributions to the prediction. Figure 7 and 8 present the ten bits with the highest mean signed SHAP values, highlighting the most influential substructures driving the model's predictions. Although these substructures are difficult to analyze in isolation, they play a major role in the rigidity of the

Table 2. Concrete Example of Prediction on External Experimental Polyimide Data

	Structure	Dianhydride	T_g (Exp.) [K]	T_g (Pred.) [K]	$\|\Delta\|$
(a)		PMDA	> 773.15[35]	791.03	-
(b)		PMDA	850.15[37]	767.30	82.85
(c)		PMDA	> 773.15[36]	777.59	-
			T_g (Exp.) > 750 K		
(d)		PMDA	625.15[35]	603.01	22.14
(e)		PMDA	631.15[37]	646.56	15.41
(f)		BPDA	543.15[35]	548.16	5.01
			T_g (Exp.) < 750 K		

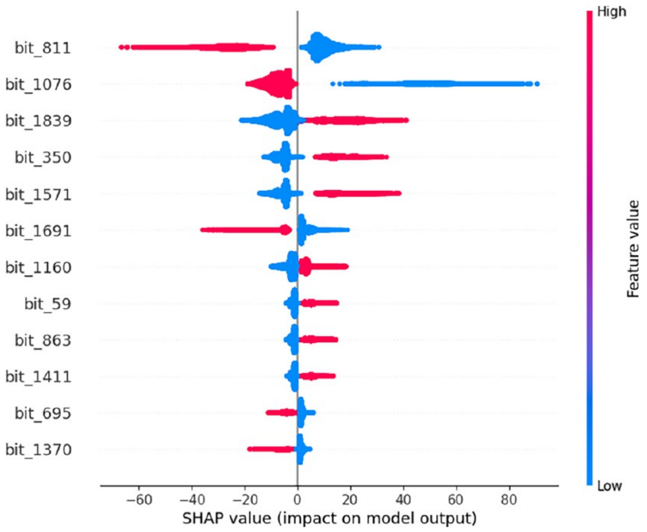


Figure 7. SHAP values indicating the contribution of fingerprint bits to T_g prediction.

molecular chain and serve as a foundation for the design of our new scoring function aimed at generating polymers with high T_g values.

This predictive model serves as foundation for the scoring function used to guide the REINVENT agent in generating high-T_g polyimides (Section 4).

4. SCORING FUNCTIONS FOR HIGH-T_G POLYIMIDES

Small structural variations can lead to large differences in the glass transition temperature. Even subtle modifications, such as the addition of a double bond or a specific functional group, can significantly increase chain rigidity, furthermore, structurally distinct polymers can also exhibit identical T_g values.[31] This nonlinear structure−property relationship implies that a score purely based on the predictive model (S_{T_g}) may be insufficient. Moreover, given the scarcity of high-T_g polyimides in the training set of the *Prior* model, there is a risk that the agent may overexplore in order to generate structurally diverse candidates.

To address this limitation, we introduce a composite scoring strategy combining complementary objectives. The model-based score (S_{T_g}) serves as the main optimization signal, quantifying the expected thermal performance of each generated polymer. In addition, a naïve score (S_h) is defined to promote substructures identified through SHAP analysis as contributors to high rigidity, while Tanimoto-based similarity scores (S_s, S_r, S_m) encourage the agent to explore new and chemically distinct regions of the design space. Although each score alone has limitations, their combination leverages their individual strengths, balancing exploitation of known high-performing motifs with exploration of novel structural possibilities. A summary of the score names, symbols, definitions, and formulas is provided in Table 3.

4.1. Design of the T_g-based Scoring Function: Tg-Score

Our objective is to generate novel polyimides with exceptionally high $T_g > 750$ K. To achieve this, we define a scoring function $S_{T_g}(x) \in [0, 1]$, derived from the T_g prediction model. Let x denote a generated polyimide and $T_g(x)$ its predicted T_g. The score $S_{T_g}(x)$ is defined such that it penalizes candidates whose predicted T_g falls below the target threshold $T_g^{\max} = 900$ K. Given the ambitious nature of this target (beyond the model's 800 K ceiling), a tolerance parameter $d_{\max} = 500$ K is introduced to define the range over which the score decreases linearly. The resulting scoring function is expressed as (eq 7)

$$S_{T_g}(x) = \begin{cases} 1 & \text{if } T_g(x) \geq T_g^{\max}, \\ 1 - \dfrac{T_g^{\max} - T_g(x)}{d_{\max}} & \text{if } T_g^{\max} - d_{\max} < T_g(x) < T_g^{\max}, \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

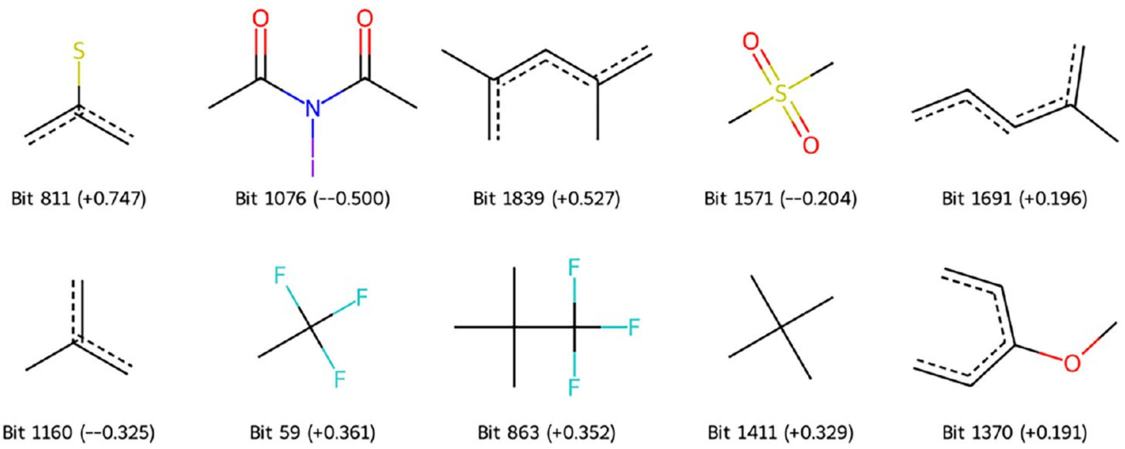


Figure 8. Top 10 Morgan fingerprint bits with the largest mean signed SHAP values for T_g prediction.

Table 3. Summary of Scoring Functions Used in the Reinforcement Learning Framework

score name	symbol	definition	formulas
T_g-Score	S_{T_g}	Penalizes candidates with predicted T_g < 900 K using linear decay over d_{max} = 500 K	eq 7
Naive High-T_g Score	S_h	SHAP-weighted fingerprint susceptibility	eq 10
Single Tanimoto Score	S_s	Similarity to a reference polymer P_1. $S_s = \min\left(\frac{T(x, P_1)}{0.7}, 1\right)$	eq 11
Random Tanimoto Score	S_r	Similarity to a randomly selected reference polymer from $P = \{P_1, P_2, P_3\}$	eq 12
Mean Tanimoto Score	S_m	Average similarity to all three references polymers: $\frac{1}{3}\sum_{i=1}^{3} S_s(x, P_i)$	eq 13

This function encourages the generation of polymers with predicted T_g values close to T_g^{max}, while still assigning partial credit to candidates within a reasonable proximity. Consequently, structures generated with predicted T_g in the [750, 800] K range are expected to have true T_g values either within this interval or potentially above it, due to the model's underestimation beyond 800 K.

4.2. Naive High-T_g Score Based on SHAP Values

While $S_{T_g}(x)$ quantifies the predicted thermal performance, it offers limited interpretability. To incorporate structural insights extracted from model explanations, we introduce a susceptibility function $h(x)$ built upon SHAP-derived fingerprint features associated with high-T_g values. This function quantifies the extent to which the molecular fingerprint of x is associated with high T_g. Specifically, $h(x)$ is computed as a weighted sum over the frequencies f_i of fingerprint bits present in x, modulated by their mean signed (s_i) and mean absolute (a_i) SHAP values. To reflect both the prevalence and the relevance of each fingerprint bit, we apply a TF-IDF weighting scheme, which prevents common substructures from dominating the score when they are

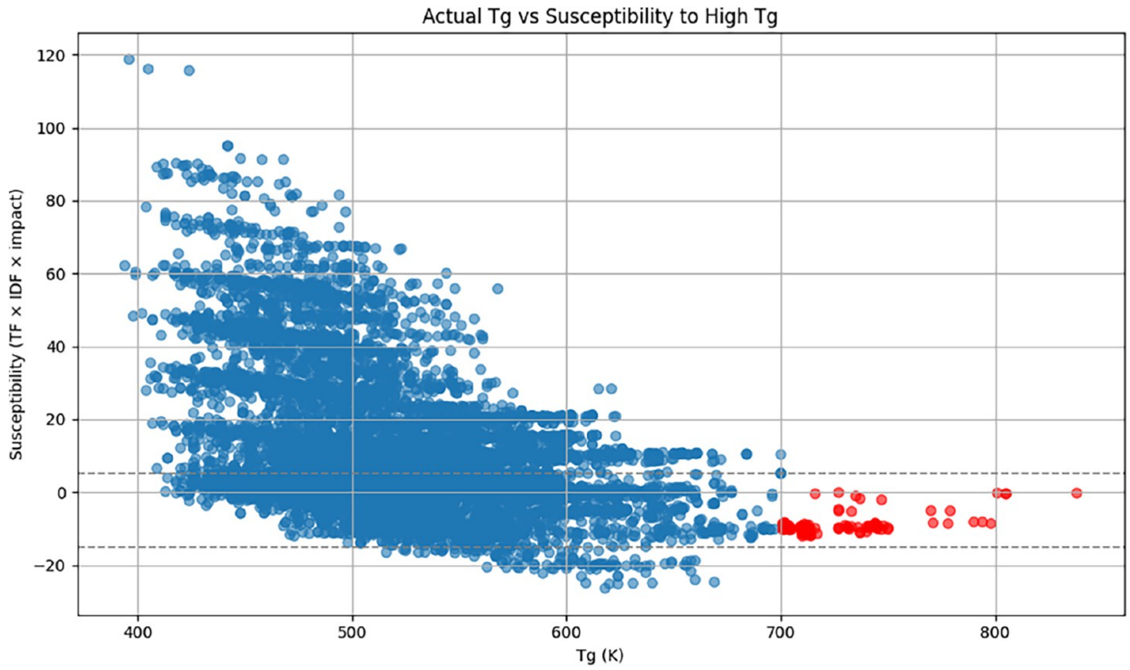


Figure 9. Polyimide T_g values plotted against their high-T_g susceptibility $h(x)$.

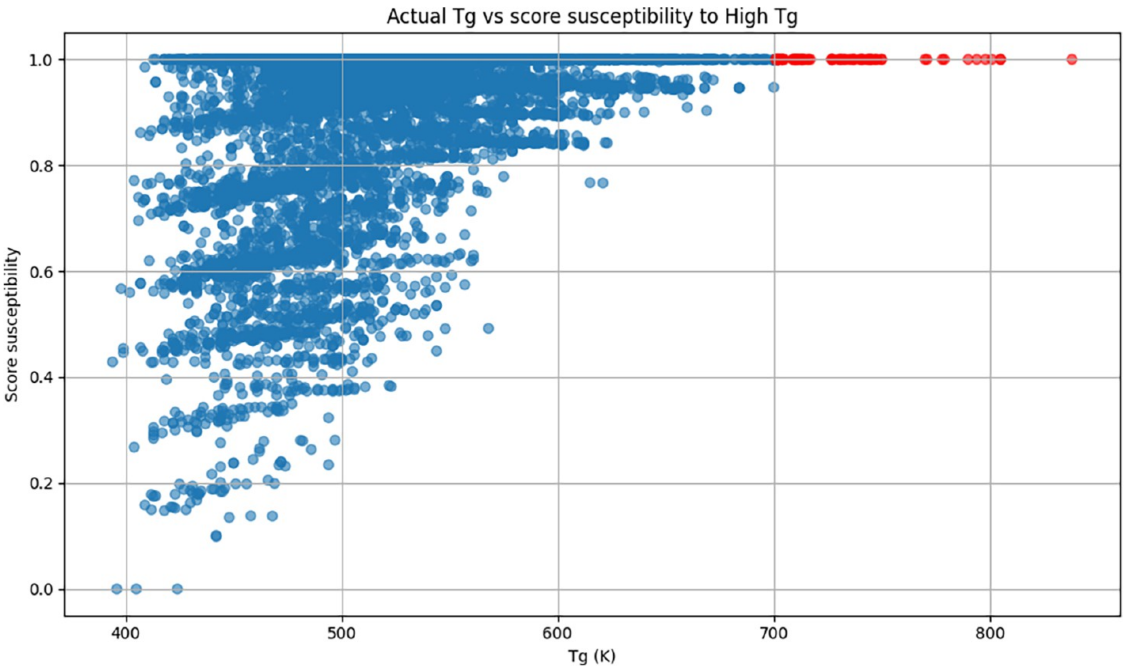


Figure 10. Polyimide T_g values versus susceptibility-based scores $S_h(x)$, illustrating filtering effect of the naive high-T_g score.

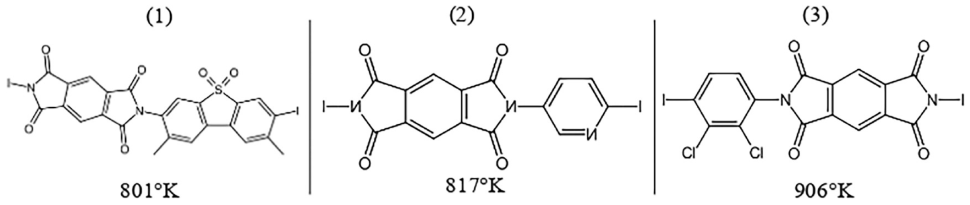


Figure 11. Reference polyimides used for Tanimoto similarity scoring. These molecules were selected based on their high glass transition temperatures and structural diversity.

not specifically indicative of high-Tg. The inverse document frequency (IDF) of bit i is defined as (eq 8)

$$\mathrm{IDF}_i = \ln\left(\frac{N}{d_i} + 1\right) \tag{8}$$

where N is the total number of molecules and d_i is the number of molecules containing bit i. The susceptibility function is thus given by (eq 9)

$$h(x) = \sum_{i \in x} f_i \cdot \mathrm{IDF}_i \cdot s_i \cdot a_i \tag{9}$$

Intuitively, $h(x)$ increases when a molecule contains fingerprint bits that the SHAP analysis identified as positively contributing to T_g. Figure 9 shows the polyimides T_g plotted against their susceptibility scores $h(x)$. This susceptibility score is termed *naive*, as it solely relies on Morgan fingerprint bits (radius 2) and does not account for higher-order or long-range molecular interactions. As such, a SHAP-weighted fingerprint approach may not fully capture the complexity underlying the T_g property. Nevertheless, Figure 9 shows that $h(x)$ provides a necessary, but not sufficient condition for high T_g. In particular, for $T_g > 700$ K, all observed susceptibility values lie within the range $[-15, 5]$. Based on this observation, we define a bounded score function $S_h(x)$ that quantifies how well a polymer aligns with the susceptibility range associated with high T_g (eq 10)

$$S_h(x) = \begin{cases} 1 & \text{if } \inf \le h(x) \le \sup\,, \\ \max\left(0,\ 1 - \dfrac{|h(x) - \inf|}{\theta}\right) & \text{if } h(x) < \inf\,, \\ \max\left(0,\ 1 - \dfrac{|h(x) - \sup|}{\theta}\right) & \text{if } h(x) > \sup \end{cases} \tag{10}$$

where inf = −15, sup = 5, and $\theta = 100$ is a sensitivity parameter controlling the decay rate of the score outside the target interval. This scoring function can be interpreted as a filtering mechanism: it penalizes polymers whose susceptibility values deviate significantly from the plausible high-T_g range. It is intended to be used in combination with more refined scoring functions. Figure 10 shows the resulting S_h values across the data set, illustrating that a large portion of candidates are excluded by this simple criterion.

4.3. Tanimoto Similarity-Based Scoring

As introduced in eq 5, the Tanimoto similarity score[21] quantifies the structural resemblance between molecules based on their substructures. Although the relationship between polymer structure and T_g is nonlinear, since similar structures can exhibit different T_g values,[31] it remains reasonable to assume that targeting structures resembling those of very high T_g polyimides may lead to the discovery of new polymers with similarly elevated T_g values. To this end, we define three Tanimoto-based scoring functions: *Single Tanimoto*, *Random Tanimoto*, and *Mean Tanimoto*. Each of these evaluates the structural similarity between a candidate polymer x and one or more reference polyimides with high observed T_g. We select three reference polyimides $P = (P_1, P_2, P_3)$ from our data set, with respective T_g values of 801, 817, and 906 K. Their molecular structures are shown in Figure 11.

4.3.1. Single Tanimoto Score. This score compares polyimide x to a single reference polyimide P_1. It is defined as (eq 11)

$$S_{\text{single}}(x) = S_s(x, P_1) = \min\left(\frac{T(x, P_1)}{k}, 1\right) \tag{11}$$

where $T(x, P_1)$ denotes the Tanimoto similarity and $k = 0.7$ is the satisfaction threshold. The similarity is capped at 0.7 to balance functional relevance and structural novelty. If k is too high, the agent will focus on the target and fail to explore other potentially valuable candidates.

4.3.2. Random Tanimoto Score. To promote structural diversity while targeting multiple high-T_g regions, this score compares x to a randomly selected reference polymer P_i from the list P. At each operation, the score is computed as (eq 12)

$$S_r(x) = \underset{i\in\{1,2,3\}}{\text{Random}} S_s(x, P_i) \tag{12}$$

This encourages exploration across several high-T_g structural basins during generation.

4.3.3. Mean Tanimoto Score. This score provides a smoother optimization signal by averaging the similarity between x and all three reference polymers. This formulation is particularly useful in scenarios where robustness across multiple high-T_g patterns is desired (eq 13)

$$S_{\text{mean}}(x) = S_m(x) = \frac{1}{3}\sum_{i=1}^{3} S_s(x, P_i) \tag{13}$$

Together, these three variants will enable different exploration modes: the Single Tanimoto focuses on targeted optimization, the Random variant promotes structural diversity, and the Mean variant provides stable convergence by averaging over multiple references.

5. HIGH-T_G POLYIMIDE GENERATION

In this section, we investigate the generation of high-T_g polyimides using the REINVENT framework, guided by the previously defined scoring functions, including the predictive T_g score, the SHAP-based susceptibility score, and Tanimoto similarity scores. We evaluate how these reward functions influence the balance between validity, novelty, and the discovery of high-T_g candidates. Our hypothesis is that Tg-driven scores maximize performance but reduce diversity, Tanimoto-based scores promote exploration within the space of high-T_g components, and the naive high-T_g score cannot perform effectively on its own, as it provides a necessary but insufficient condition.

5.1. Prior/Agent

As described in Section 2.2, the REINVENT architecture consists of two main components: a pretrained *Prior* model and a fine-tuned version referred to as the *Agent*. The Prior model is trained on a data set of 500,000 synthetic chemically valid polyimides, with 20% reserved for validation. Training is performed using a *learning rate* of 0.001, a *batch size* of 128, and early stopping with a patience of 5 epochs to prevent overfitting. Once pretrained, the Prior learns the characteristic connectivity patterns and token sequences associated with imide functional groups. As a consequence, it is capable of generating diverse and valid polyimides reflecting the distribution of the training set.

The Agent is initialized with the weights of the Prior and subsequently fine-tuned via reinforcement learning. Hyperparameters for fine-tuning include a learning rate of 0.0005, $\sigma = 60$, and a maximum of *100 generation steps*, because in most cases the Agent collapses around the 80th step. At each step, the sample function can generate up to 500 polymers, from which valid candidates are evaluated using the chosen *scoring function*.

Optionally, *experience replay* can be enabled, which reinjects a small number of top-scoring experiences from previous iterations into the current fine-tuning process. This approach helps mitigate temporal drift and promotes stable exploration around high-reward regions. This setting was compared with alternative configurations and retained because it consistently outperformed them.

5.2. Evaluation Metrics

To assess the quality and diversity of the generated polyimides, we define several key evaluation metrics. Some of these metrics rely on canonical SMILES to ensure that structurally identical molecules are uniquely identified, as different SMILES strings may represent the same compound. Canonicalization is performed using RDKit.[33]

5.2.1. Novelty. The proportion of generated molecules at step g that are not present in the training data set (eq 14)

$$\text{Novelty}(g) = \frac{|\mathcal{C}_g \setminus \mathcal{D}_{\text{train}}|}{|\mathcal{D}_g|} \tag{14}$$

where $\mathcal{C}_g$ is the set of unique canonical SMILES generated at generation g, $\mathcal{D}_g$ is the full set of generated SMILES at that step, and $\mathcal{D}_{\text{train}}$ is the set of canonical SMILES from the training data set. Thus, $|\mathcal{C}_g \setminus \mathcal{D}_{\text{train}}|$ counts the number of unique polyimides at step g that were not seen during training.

5.2.2. Uniqueness. The proportion of structurally unique SMILES within a single generation (eq 15)

$$\text{Uniqueness}(g) = \frac{|\mathcal{C}_g|}{|\mathcal{D}_g|} \tag{15}$$

5.2.3. Internal Diversity (IntDiv). Defined as (eq 16)

$$\text{IntDiv}(g) = 1 - \frac{1}{|\mathcal{D}_{g2}|}\sum_{(x,y)\in\mathcal{D}_{g2}} T(x, y) \tag{16}$$

where $\mathcal{D}_{g2}$ is the set of all unique unordered pairs of molecules in $\mathcal{D}_g$, and $T(x, y)$ is the Tanimoto similarity between polymers x and y. A value of IntDiv(g) close to 1 indicates high structural diversity among generated molecules.

5.2.4. Similarity to Nearest Neighbor (SNN). The average Tanimoto similarity between each generated molecule and its most similar counterpart in the training data set (eq 17):

$$\text{SNN}(g) = \frac{1}{|\mathcal{D}_g|}\sum_{x\in\mathcal{D}_g} \max_{y\in\mathcal{D}_{\text{train}}} T(x, y) \tag{17}$$

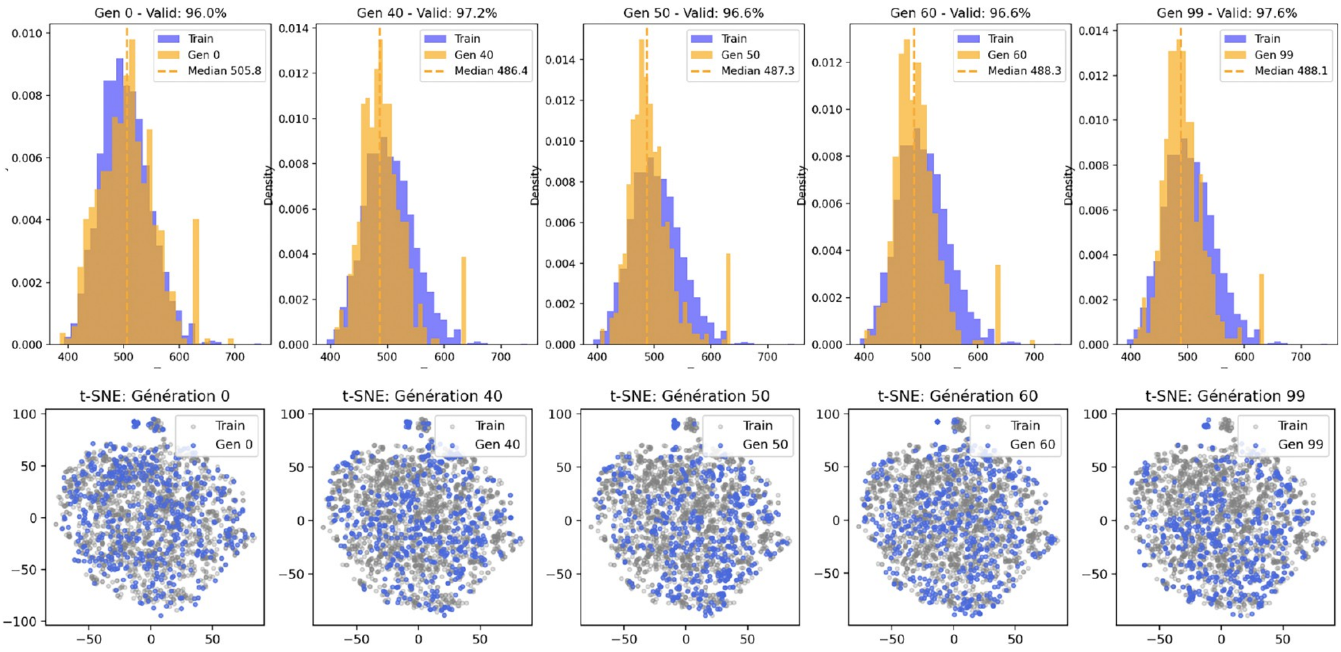


Figure 12. Naive High-T_g score: t-SNE and Tg density histogram of $\mathcal{D}_g$ vs $\mathcal{D}_{\text{train}}$. The generated polymers do not leave the training set region in the latent space.

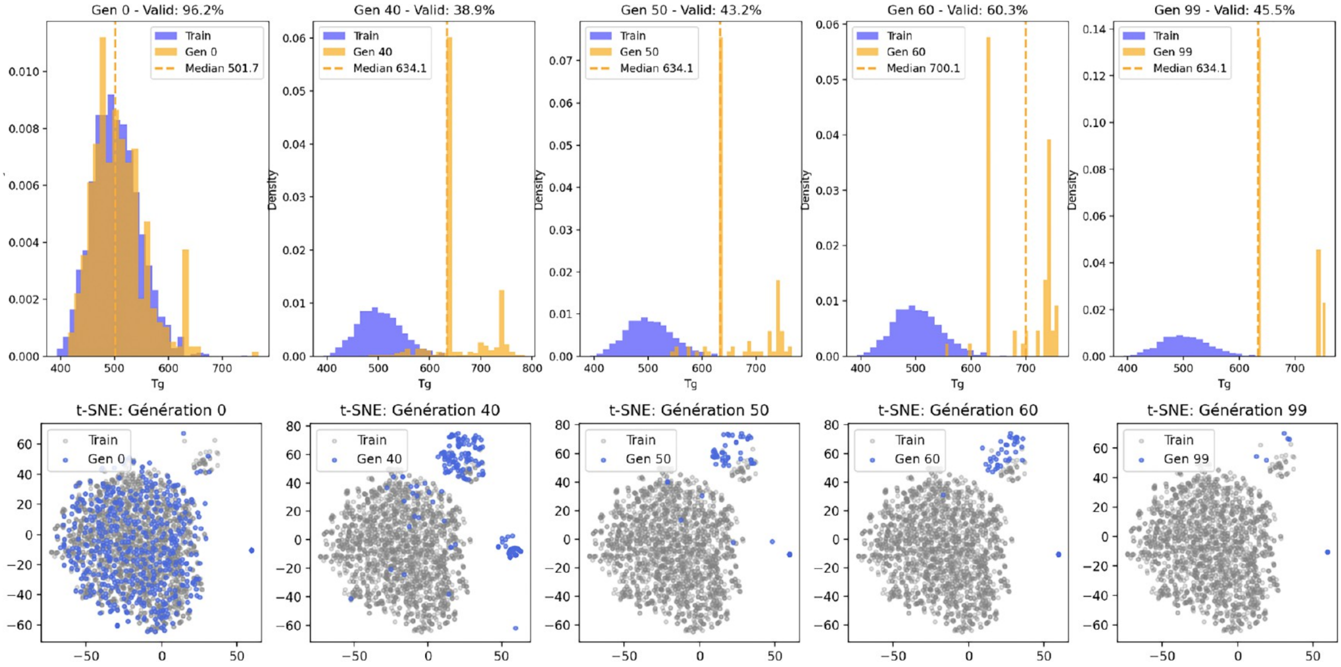


Figure 13. T_g score (S_{T_g}): t-SNE and Tg density histogram of $\mathcal{D}_g$ vs $\mathcal{D}_{\text{train}}$. From generation 40 onward, the generation focuses on a subset separate from the training data set.

A lower SNN value indicates greater structural novelty.

5.2.5. Validity and Valid SMILES Count.

Let n_g be the number of valid SMILES generated at generation g. The validity fraction is defined as (eq 18)

$$\text{Validity}(g) = \frac{n_g}{|\mathcal{D}_g|} \tag{18}$$

Validity is assessed using RDKit by attempting to parse each SMILES string into a chemically valid molecular object.

5.2.6. Global Generation Score.

To rank the 100 generation steps, we define an overall performance score Rate(g) (eq 19) as a weighted sum of the metrics, where the weights indicate the relative importance to each metric, and penalizing generations with very low molecular counts. Alternative configurations with $q_{n_g} \in (0, 0.1]$ were tested and produced qualitatively similar rankings.

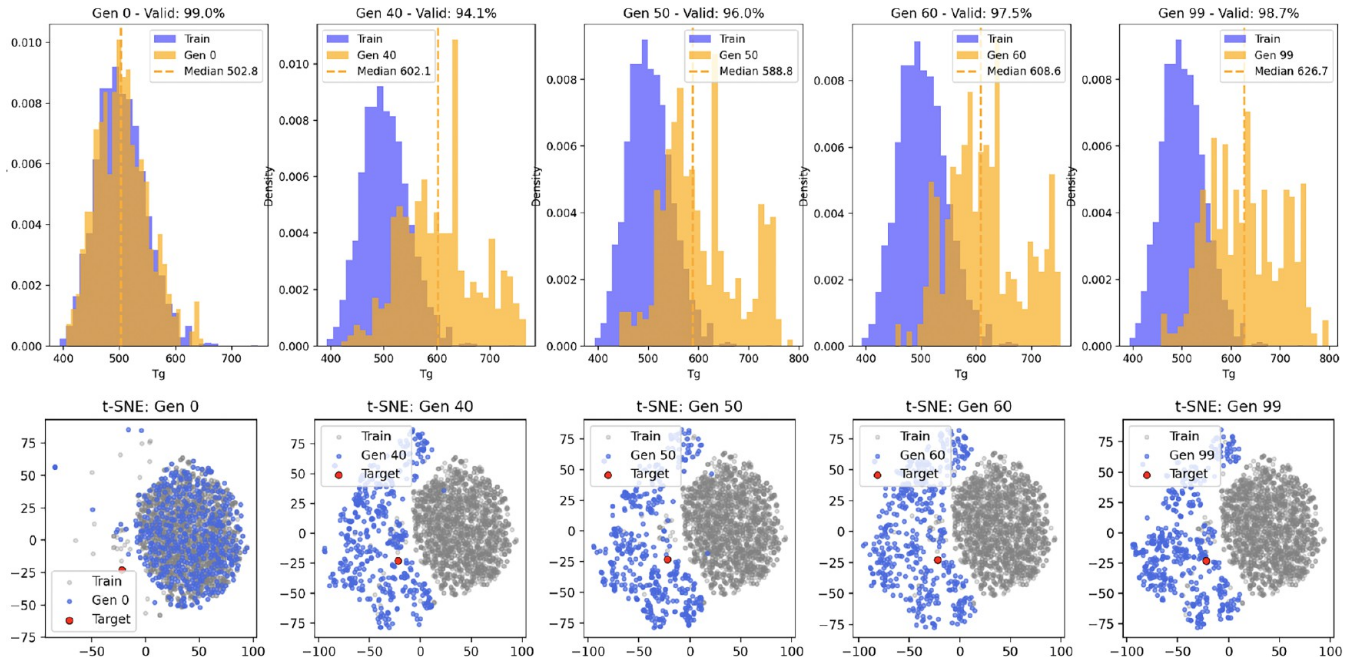


Figure 14. Single Tanimoto score (S_s): t-SNE and Tg density histogram of $\mathcal{D}_g$ vs $\mathcal{D}_{\text{train}}$. The generations are concentrated in a new, unexplored location, toward target polymer.

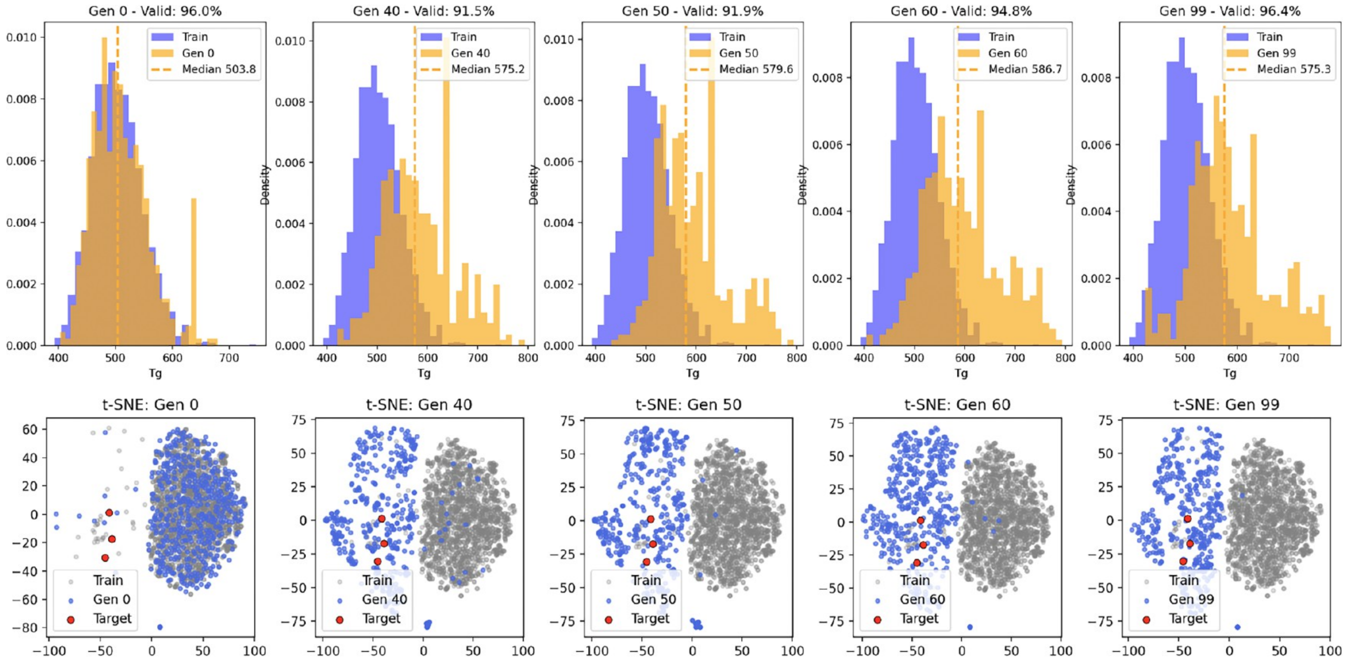


Figure 15. Random Tanimoto score (S_r): t-SNE and Tg density histogram of $\mathcal{D}_g$ vs $\mathcal{D}_{\text{train}}$. The generations are concentrating in a new and unexplored location, heading toward the 3 target polymers.

$$\begin{aligned}\text{Rate}(g) = {} & \frac{3}{10}\text{Uniqueness}(g) + \frac{2}{10}\text{IntDiv}(g) \\ & + \frac{3}{10}\text{Validity}(g) + \frac{1}{10}(1 - \text{SNN}(g)) \\ & + q_{n_g}(n_g{\cdot}\text{Novelty}(g))\end{aligned} \tag{19}$$

In this formulation, Uniqueness, IntDiv, Validity, and (1-SNN) are metrics ranging from 0 to 1, where (1-SNN) measures distance from the training set. The weights prioritize uniqueness (30%) and validity (30%) over diversity (20%), and diversity over proximity to the training set (10%). However, when considering only these four normalized ratio metrics, a generation producing a single novel polymer is favored over the one producing ten polymers.

To account for the absolute yield of novel structures generated at step g, we include the term (n_g · Novelty(g)). Experimentally, this term reaches a maximum value around 20. To maintain the significance of yield without overwhelming the other metrics, we assign it a weight of $q_{n_g} = \frac{1}{10}$. Thus, our

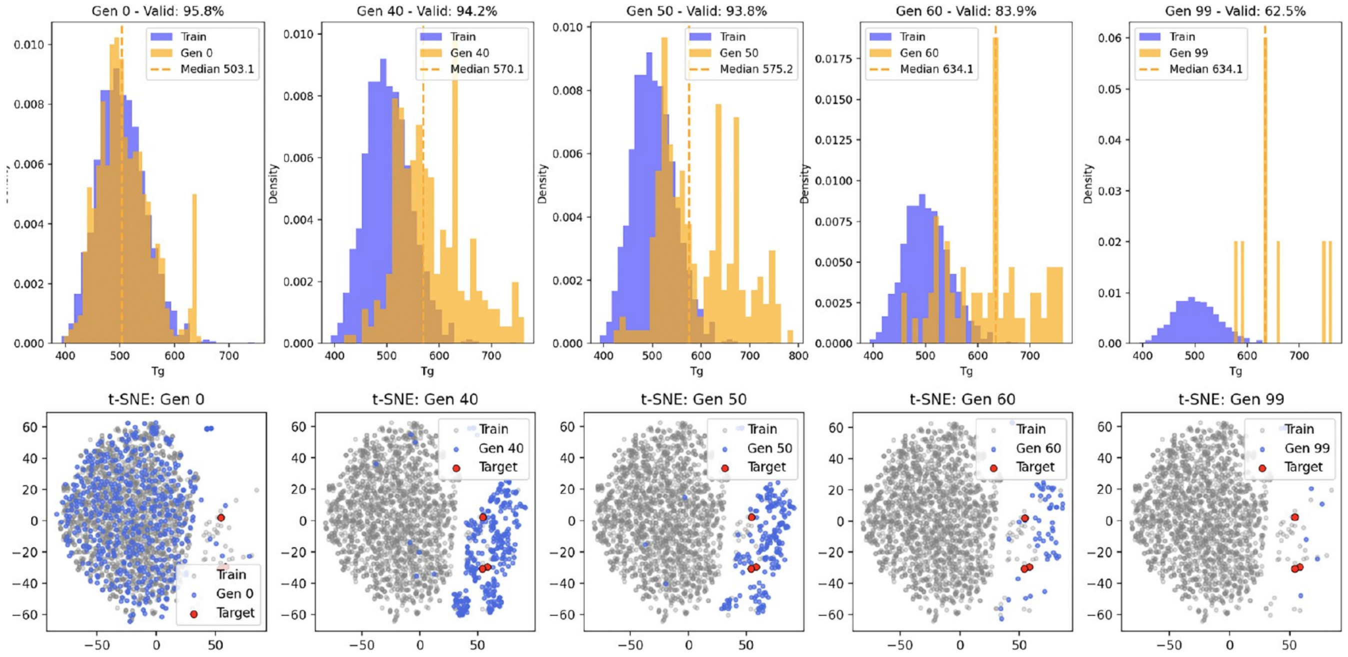


Figure 16. Mean Tanimoto score (S_m): t-SNE and Tg density histogram of $\mathcal{D}_g$ vs $\mathcal{D}_{\text{train}}$. The generations concentrate in a new and unexplored location, moving more selectively toward the 3 target polymers.

Rate(g) value is not normalized, but it allows generations to be ranked according to the criteria described above.

Example 1. For a near-optimal generation in our context, with Uniqueness(g) = 1, IntDiv(g) = 1, Validity(g) = 1, (1 − SNN) = 1, and (n_g · Novelty(g)) ≃ 20, we have (eq 20):

$$\begin{aligned}\text{Rate}(g) &\simeq (0.3 + 0.2 + 0.3 + 0.1) + 0.1 \times 20 \\ &= 0.9 + 2.0 = 2.9\end{aligned} \tag{20}$$

Other formulations with different weightings may yield different results; however, the outcomes obtained when considering the average over all generations (Figure 13) show very little impact on the overall conclusions derived from this rate.

The generation step with the highest Rate(g) is considered the best-balanced performer with respect to the priorities defined above in terms of chemical uniqueness, validity, novelty, diversity, and molecular yield. Among these metrics, the global score is the most important, as it summarizes the overall performance under the experimental conditions.

5.3. Polyimide Generation: Naive High-T_g Score, Tg Score, and Tanimoto-Based Scores

Figures 12, 13, 14, 15, and 16 illustrate the evolution of generated polymers at intermediate steps of Agent fine-tuning without experience replay (ER), under five different scoring functions: the naive high-T_g score S_h, the predictive T_g score S_{T_g}, and the Tanimoto-based scores, single S_s, random S_r, and mean S_m. Each figure presents both the t-distributed stochastic neighbor embedding (t-SNE) of Morgan fingerprints and the density histogram of the predicted glass transition temperatures for the generated set $\mathcal{D}_g$, compared against a representative subset of the training set $\mathcal{D}_{\text{train}}$. Only valid polymers are used in the t-SNE visualizations.

5.3.1. Global Analysis of Agent Generations. A preliminary analysis of the generations shown in Figures 12 to 16 confirms that, as expected, the earliest generations closely resemble the training distribution in both structure and predicted T_g.

5.3.1.1. Naive SHAP-Based Score (S_h). The naive high-T_g score (S_h) function (Figure 12) fails to generate polymers that differ significantly from the training set, as it performs **pure exploitation**, with no visible expansion into unexplored latent regions. This is consistent with the score's design: ($S_h(x) = 1$) is easily achieved for polymers x with T_g values in the range [400, 800], making it a useful complementary constraint but insufficient on its own to promote novelty or exploration.

5.3.1.2. Model-based Tg Score (S_{T_g}). By contrast, the model-based score (S_{T_g}), which is directly tied to a regression model, is difficult to satisfy and strongly biases the generation toward high-T_g candidates. As a side effect, this drives the Agent to **extreme exploration**, far from the training distribution, resulting in a high number of invalid molecules, at step 99 only about 45% of the polymers generated are valid (Figure 13).

5.3.1.3. Tanimoto Scores. The single Tanimoto score (S_s) avoids this issue with a good **balance exploration and exploitation**. At step 99, approximately 98% of the generated polymers are valid. This score effectively guides the Agent toward regions that are 70% structurally similar to the reference polymer P_1, leading to the synthesis of several high-T_g candidates, as illustrated by the histogram shift and t-SNE in Figure 14.

A similar trend is observed with the random Tanimoto score (S_r) in Figure 15, where the Agent alternates between the three reference polymers, encouraging **more exploration than exploitation** of chemically meaningful yet diverse regions in the latent space. Interestingly, the mean Tanimoto score (S_m) exhibits behavior that closely resembles that of the model-based score (S_{T_g}). By generation step 99, the model consistently generates polymers with high predicted T_g values (Figure 16) while producing a higher number of valid candidates. This outcome is surprising, given that individual Tanimoto scores are capped at 0.7 (see Section 4.3.1). However, averaging across the three references appears to enable the model to display **hybrid behavior**, thereby capturing structural patterns at the border of the embedding space that favor high T_g, even without directly using the predictive model.

While these global observations based on t-SNE plots and T_g histograms offer valuable insights, a complete evaluation requires more rigorous metrics such as uniqueness, diversity, novelty, and validity to fully assess generation quality.

5.3.2. Focused Analysis of Agent Generations: High-T_g Polyimides.

5.3.2.1. Performance by Scoring Function. To fairly evaluate the different Agent generations, we restrict our analysis to generated polymers with high T_g values. For each generation, the *Rate* score is computed on the subset of generated polymers with T_g > 750 K, allowing a focused assessment aligned with the design objective. We report evaluation metrics for the best-performing generations *(best g)* out of the 100 runs, both with and without experience replay (ER). The results are summarized in Table 4.

Table 4. Evaluation of the Best Polymer Generations (> 750, K) from Various Agent with their Associated Scoring Functions

ER	Metrics	Naive S_h	Tg S_{Tg}	Single S_s	Random S_r	Mean S_m
No	**Rate(g)**	0.600	1.108	1.625	**1.703**	1.010
	Novelty	0	0.364	0.692	**0.769**	0.750
	Uniqueness	1	0.917	1.000	1.000	1.000
	InDiv	0	0.444	**0.527**	0.443	0.486
	SNN ↓	1	0.921	**0.804**	0.858	0.867
	Validity	1	1.000	1.000	1.000	1.000
	$\lvert D_g \rvert$	1	12	**13**	**13**	4
	best g	2	29	78	83	26
	Highest $T_g(K)$		764.6	**774.3**	768.4	750
Yes	**Rate(g)**	0.714	**2.148**	1.599	1.116	1.221
	Novelty	1	0.560	0.600	**0.800**	0.714
	Uniqueness	1	0.962	1.000	1.000	1.000
	InDiv	0	0.458	0.432	0.501	**0.524**
	SNN ↓	0.823	0.880	0.872	0.846	**0.836**
	Validity	1	1.000	1.000	1.000	1.000
	$\lvert D_g \rvert$	1	**26**	15	5	7
	best g	73	52	37	9	66
	Highest $T_g(K)$	752	**797.7**	765.3	775.8	791.2

The analysis of Table 4 highlights several key observations. The generation step at which the best performance is achieved depends on the scoring function, and no consistent pattern emerges. In some cases, peak performance occurs early, between generations 9 and 29, while in others it is reached later, between generations 52 and 83.

As expected, the Agent trained using the naive high-T_g score S_h performs the weakest, generating no novel polymers (with T_g > 750 K) without ER, and only one when ER is applied.

Overall, the best-performing Agent is the one using the score S_{Tg} with ER. However, depending on the evaluation metric, other Agents also perform competitively. In particular, the Agent using the Random Tanimoto score S_r achieves the highest novelty, both with and without ER.

The highest T_g values reported in Table 4 correspond to novel polymers only. The best novel T_g value, 797.7 K, was achieved using the S_{Tg} score with ER.

5.3.2.2. Effect of Experience Replay. ER affects each scoring strategy differently. ER improves performance for Agents trained with the naive S_h (from rate 0.600 to 0.714), the predictive S_{Tg} (from rate 1.108 to 2.148), and the Mean Tanimoto S_m scores (from rate 1.010 to 1.221). In contrast, ER leads to decreased performance for the single Tanimoto S_s (from rate 1.625 to 1.599) and random Tanimoto S_r (from rate 1.703 to 1.116).

More specifically, on one hand, the novelty of S_{Tg} increased from 0.364 to 0.560 (+53.8%), and the internal diversity (InDiv) of S_m increased from 0.486 to 0.524 (+7.8%).

On the other hand, the novelty of S_s decreased from 0.692 to 0.600 (−13.3%), its InDiv dropped from 0.527 to 0.432 (−18.0%). and the number of generated structures decreased from 13 to 5 (−61.5%).

As a result, it is difficult to draw general conclusions about the influence of ER, other than that it can enhance performance in some cases depending on the scoring strategy, a summary is given in Table 5. It is important to note again that this analysis is limited to high-T_g polymers >750 K), which may explain discrepancies with findings in prior studies such as,[20] where ER consistently improved performance. A similar trend was observed with $q_{n_g} \in \{0.05, 0.08, 0.09\}$.

Table 5. Qualitative Summary of Exploration/Exploitation Regimes and the Effect of Experience Replay (ER) for Polymers with T_g > 750 K

Score	Regime	Without ER	With ER
S_{T_g}	Extreme Exploration	Many invalid candidates	More novelty and High-T_g
S_h	Collapsed Exploitation	No novelty	Slight gain in diversity
S_s	Balance Explor/Exploi	More diverse and novel	Less diverse and novel
S_r	More Explor/Less Exploi	Broader exploration	Reduced diversity
S_m	Hybrid	More novel but less High-T_g	More diverse and High-T_g

5.3.2.3. t-SNE Analysis and Structural Diversity. Figure 17 presents a t-SNE projection of the novel polymers (i.e., those not present in the training data set). The distribution of these molecules suggests that, for a common objective, both the scoring function and the use of ER influence the nature of the generated polymers. Notably, the structures of polymers with the highest T_g differ not only across scoring functions but also between ER settings for the same score. A chemical analysis of these polymers is provided in Section 7, along with the rest of the generated data set.

6. HIGH-T_G POLYIMIDE GENERATION: SCORE COMBINATION

6.1. Motivation and Aggregation Functions

So far, we have analyzed the impact of four individual scoring functions, each designed to guide the generation of high T_g polymers. The evaluation of their performance metrics and regime in Figure 5 suggests that these scoring functions exhibit distinct and complementary advantages. This observation motivates the exploration of score combinations, with the goal of leveraging their strengths to enhance the generation process.

A natural way to combine multiple scoring functions is to use a mean function. Beyond the standard weighted arithmetic mean, several alternative formulations can be employed to highlight different characteristics of the input scores. In this work, we examine three mean functions (eq 21, 22, 23) and a new

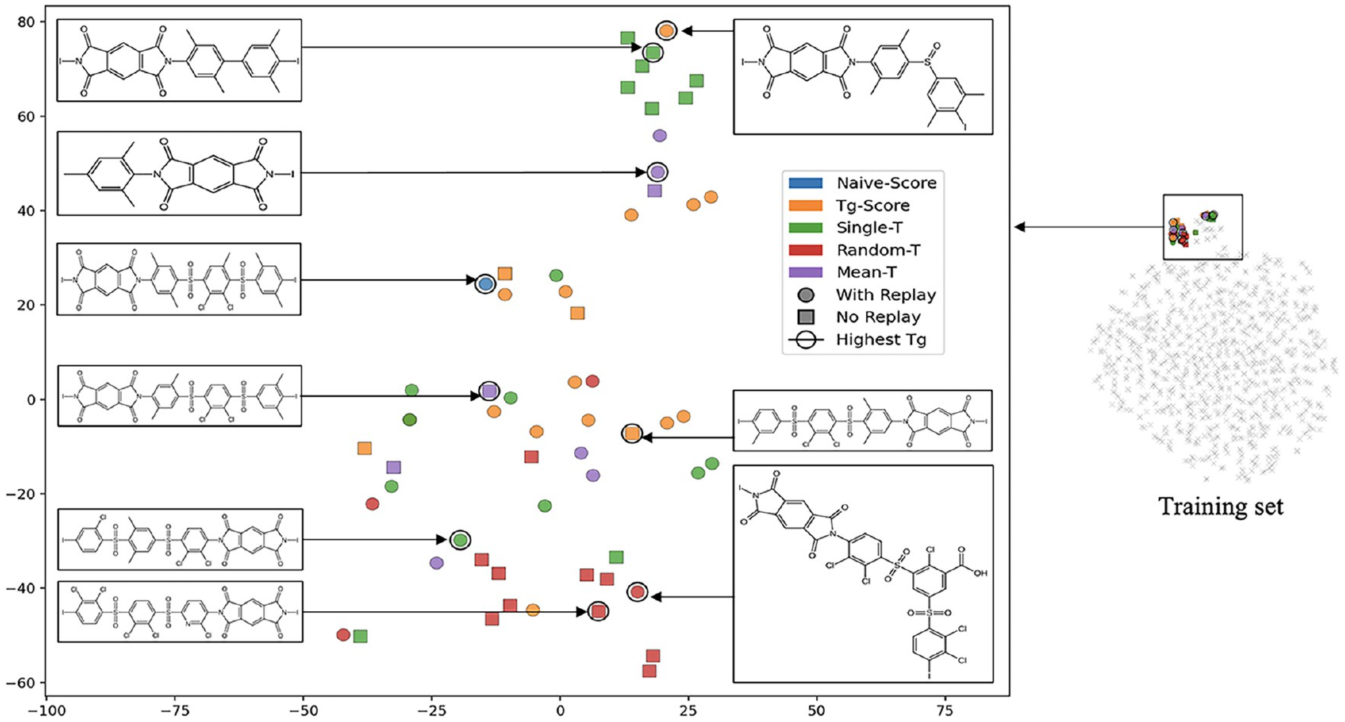


Figure 17. t-SNE projection of the best-generated polymers, with those obtained using ER shown as circles and those without ER shown as squares. The highest values are highlighted with a big circle.

aggregation function (eq 24). Given two score values $s_1, s_2 \in [0, 1]$, we define:

Weighted Arithmetic Mean.

$$\mathrm{Mean}_\alpha(s_1, s_2) = \alpha s_1 + (1 - \alpha)s_2 \tag{21}$$

where $\alpha \in]0,1[$ is the weighting parameter. For $\alpha = 0.5$, this reduces to the simple arithmetic mean, giving equal importance to both scores ($\mathrm{Mean}_{0.5} = \mathrm{Mean}$).

Geometric Mean.

$$\mathrm{GeoMean}(s_1, s_2) = \sqrt{s_1 s_2} \tag{22}$$

This mean gives more weight to the joint magnitude of the scores and slightly penalizes combinations where one score is low. Another study[39] on reinforcement learning, drawing insights from the theory of approachability in vector-valued stochastic games, also employed such mechanisms to combine multiobjective rewards.

Harmonic Mean. Defined as zero when $s_1 = s_2 = 0$, and otherwise as

$$\mathrm{HarmMean}(s_1, s_2) = \frac{2s_1s_2}{s_1 + s_2} \tag{23}$$

This formulation is commonly used when a reciprocal relationship exists between s_1 and s_2, emphasizing balance and penalizing disparity between the two values.

Proposed Exponential Aggregation. Defined as zero if either $s_1 = 0$ or $s_2 = 0$; otherwise:

$$\mathrm{ExpAgg}(s_1, s_2) = e^{-\ln(s_1)\cdot\ln(s_2)} \tag{24}$$

Among the functions studied, we introduce the **ExpAgg** as a novel candidate for score aggregation in reinforcement learning. Its formulation, which combines logarithmic and exponential components, strongly penalizes very low values, thereby emphasizing the importance of maintaining sufficiently high scores across both inputs simultaneously. We cannot refer to this function as an average, as it does not satisfy the properties of a standard mean. For example, it does not satisfy the identity property, since $\mathrm{ExpAgg}(s_1, s_1) \neq s_1$.

Let us consider the score function in the case where $s_1 = s_2$; the function $f(s_1) = \mathrm{ExpAgg}(s_1, s_1) - s_1$, plotted in Figure 18, illustrates the difference between the input and the output. As described in,[40] an aggregation function is said to exhibit *upward reinforcement* when its output exceeds the input scores, which occurs here on the interval [0.37, 1]; for example, ExpAgg(0.5, 0.5) = 0.61. Conversely, the function exhibits *downward reinforcement* when the output is lower than the input scores, as observed on the interval (0, 0.37); for instance, ExpAgg(0.2, 0.2) = 0.075. These observations highlight the reinforcement behavior of our aggregation function.

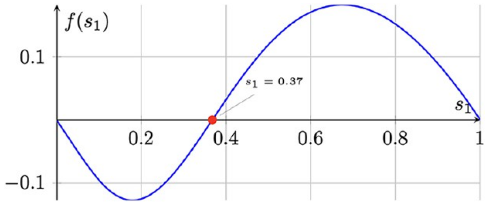
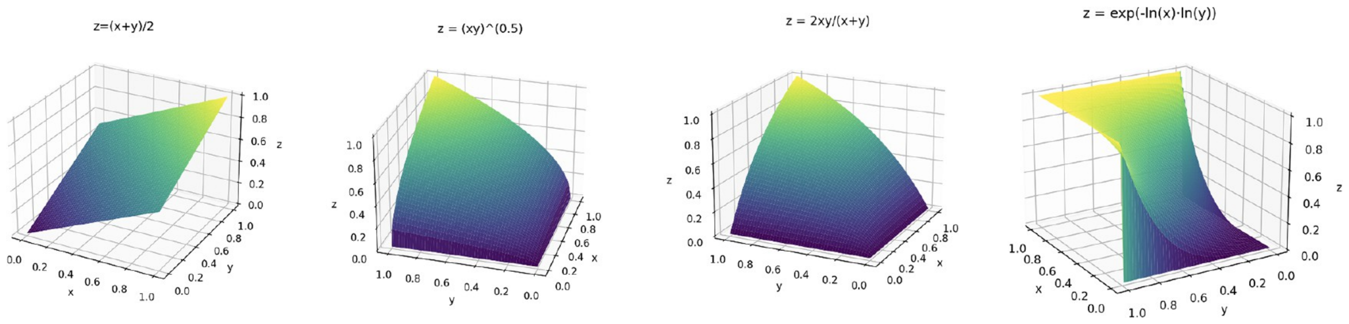


Figure 18. Plot of $f(s_1) = e^{-(\ln(s_1))^2} - s_1$ with an approximate root at $s_1 = 0.37$.

Figure 19 illustrates the range [0,1] of values produced by each function over the domain $[0, 1]^2$. The arithmetic mean generates a flat surface, while the geometric and harmonic means produce convex shapes. In contrast, the exponential aggregation displays a distinct concave profile, which is novel. We investigate how these curvature differences influence score aggregation within the reinforcement learning context.


z=(x+y)/2

z = (xy)^(0.5)

z = 2xy/(x+y)

z = exp(-ln(x)·ln(y))

Figure 19. Value surfaces of different mean functions applied to two input scores in the interval (0, 1].

Table 6. Performance Metrics for the Agent Combining S_{T_g} with Various Scoring Functions Using Different Mean Functions[a]

Scores (ER=No)	Metric	Mean	GeoMean	HarmMean	ExpAgg	$\sum$
Single Tanimoto and Tg score $S_s \& S_{Tg}$	Rate(g)	1.805	1.900	1.842	$\boxed{\mathbf{2.910}}$	$\boxed{\mathbf{8.457}}$
	Novelty	0.524	0.500	0.478	**0.786**	
	Uniqueness	**1.000**	**1.000**	0.958	**1.000**	
	$InDiv$	0.482	0.450	**0.486**	0.458	
	$SNN \downarrow$	0.909	0.903	0.907	**0.818**	
	Validity	1.000	1.000	1.000	1.000	
	$\|D_g\|$	21	24	24	**28**	
	best g	42	34	31	62	
	Highest T_g	769	769	769	**788**	
Random Tanimoto and Tg score $S_r \& S_{Tg}$	Rate(g)	2.203	$\boxed{\mathbf{2.504}}$	2.101	1.318	8.126
	Novelty	0.577	0.514	**0.609**	0.429	
	Uniqueness	**1.000**	**1.000**	**1.000**	0.933	
	$InDiv$	**0.464**	**0.464**	0.450	0.431	
	$SNN \downarrow$	0.893	0.892	**0.889**	0.905	
	Validity	1.000	1.000	1.000	1.000	
	$\|D_g\|$	26	**35**	23	15	
	best g	50	65	98	61	
	Highest T_g	769	**793**	768	775	
Mean Tanimoto and Tg score $S_m \& S_{Tg}$	Rate(g)	1.298	1.860	$\boxed{\mathbf{2.007}}$	1.551	6.716
	Novelty	0.400	0.407	0.565	**0.667**	
	Uniqueness	**1.000**	0.931	**1.000**	0.923	
	$InDiv$	0.451	0.460	**0.482**	0.468	
	$SNN \downarrow$	0.919	0.924	0.898	**0.866**	
	Validity	1.000	1.000	1.000	1.000	
	$\|D_g\|$	15	**29**	23	13	
	best g	42	37	40	38	
	Highest T_g	**793**	768	788	769	
Naive high-T_g and Tg score $S_h \& S_{Tg}$	Rate(g)	$\boxed{\mathbf{2.455}}$	1.989	1.335	0.600	6.379
	Novelty	**0.640**	0.565	0.545	0.000	
	Uniqueness	0.893	**1.000**	0.917	**1.000**	
	$InDiv$	0.412	0.393	**0.479**	0.000	
	$SNN \downarrow$	**0.875**	0.891	0.908	1	
	Validity	1.000	1.000	1.000	1.000	
	$\|D_g\|$	**28**	23	12	1	
	best g	80	69	45	35	
	Highest T_g	**783**	781	769	-	
	$\sum$ Best Rate	8.869	$\boxed{\mathbf{9.361}}$	8.393	7.487	

[a] Metrics are computed on generated polymers (T_g > 750 K). Experience replay = No.

6.2. Combining Two Scores: the T_g Score with Other Scores

It is intuitively reasonable to consider the T_g score, S_{Tg}, as a baseline for constructing two-score combinations, given its direct alignment with the generation objective. Therefore, we analyze the effect of combining S_{Tg} with other scoring functions, specifically the naive high-T_g score and the Tanimoto-based scores. The performance of these combinations is summarized in Table 6 for agents without experience replay (ER), and in Table 7 for those with ER.

Table 7. Performance Metrics for the Agent Combining S_{Tg} and Other Scores via Different Mean Functions[a]

Scores (ER=Yes)	Metric	Mean	GeoMean	HarmMean	ExpAgg	$\sum$
Single Tanimoto and Tg score $S_s\&S_{Tg}$	Rate(g)	1.598	1.586	2.450	**$\boxed{2.675}$**	**$\boxed{8.309}$**
	Novelty	0.500	0.667	0.630	**0.760**	
	Uniqueness	**1.000**	0.857	0.964	0.962	
	InDiv	0.440	0.414	0.421	**0.473**	
	SNN ↓	0.903	0.875	0.867	**0.839**	
	Validity	1.000	1.000	1.000	1.000	
	$\lvert D_g \rvert$	18	**14**	**28**	26	
	best g	44	74	41	85	
	Highest T_g	773	771	773	**775**	
Random Tanimoto and Tg score $S_r\&S_{Tg}$	Rate(g)	1.429	**$\boxed{2.155}$**	1.615	2.136	7.335
	Novelty	0.467	0.636	**0.692**	0.467	
	Uniqueness	0.938	0.957	**1.000**	0.968	
	InDiv	0.460	0.461	**0.511**	0.451	
	SNN ↓	0.911	0.874	**0.867**	0.911	
	Validity	1.000	1.000	1.000	1.000	
	$\lvert D_g \rvert$	16	23	13	**31**	
	best g	43	41	67	44	
	Highest T_g	**793**	771	772	775	
Mean Tanimoto and Tg score $S_m\&S_{Tg}$	Rate(g)	2.000	2.141	1.876	**$\boxed{2.391}$**	6.532
	Novelty	0.591	0.519	0.524	**0.708**	
	Uniqueness	**1.000**	0.964	0.913	**1.000**	
	InDiv	0.439	**0.451**	0.430	0.388	
	SNN ↓	0.882	0.901	0.0.891	**0.865**	
	Validity	1.000	1.000	1.000	1.000	
	$\lvert D_g \rvert$	22	**28**	23	24	
	best g	25	37	43	48	
	Highest T_g	**788**	769	779	781	
Naive high-Tg and Tg score $S_h\&S_{Tg}$	Rate(g)	2.301	**$\boxed{2.543}$**	1.693	0.717	7.254
	Novelty	0.615	0.514	0.435	**1.000**	
	Uniqueness	1.000	0.972	1.000	1.000	
	InDiv	0.432	**0.445**	0.425	0.000	
	SNN ↓	0.859	0.894	0.916	**0.833**	
	Validity	1.000	1.000	1.000	1.000	
	$\lvert D_g \rvert$	26	**36**	23	1	
	best g	74	64	33	3	
	Highest T_g	773	**775**	771	753	
	$\sum$Rate	9.476	$\boxed{10.573}$	9.782	10.067	

[a] Metrics computed on generated polymers (T_g > 750 K). Experience replay= Yes.

6.2.1. Without Experience Replay (Table 6).

The agent using only the T_g score S_{Tg} achieved a best rate of 1.108 without ER. All score combinations (regardless of the function used) outperformed the use of S_{Tg} alone, with the sole exception of *ExpAgg*(S_{Tg}, S_h). This indicates that, overall, score combination leads to a substantial gain in performance.

The single Tanimoto score S_s, which individually yields a rate of 1.625, produced the highest rate among all combinations (2.910) when paired with S_{Tg} via the *ExpAgg*, yielding the highest novelty (0.786) and SNN (0.818) across all experiments. This outstanding performance arises from two main factors: (1) while S_s provides a good balance between exploration and exploitation, S_{T_g}, in contrast, exhibits a highly explorative regime, (2) the exponential aggregation *ExpAgg* amplifies the influence of the two scores when they improve in tandem and exhibit relatively high individual performance.

Combining the naive high-T_g score S_h with S_{Tg} using *ExpAgg* leads to poor convergence. This can be attributed to the weak standalone performance of S_h, which drags down the overall score, as the ExpAgg heavily penalizes low-performing inputs.

The random Tanimoto score S_r, which individually achieves the highest rate among all solo scores (1.703), generally exerts a positive and stable influence on S_{Tg} when combined using the arithmetic *(Mean$_{0.5}$)*, geometric *(GeoMean)*, or harmonic *(HarmMean)* means. However, its effectiveness drops when using *ExpAgg* (1.318), likely due to the agent's early stage exploration of diverse target polymers. Because *ExpAgg* is sensitive to low values, it reduces the benefit of compensatory score effects. Additionally, *GeoMean*(S_{Tg}, S_r) produced the largest number of generated structures (35) across all experiments.

The mean Tanimoto score S_m, despite its modest individual performance (1.010), appears to also complement S_{Tg} well, yielding an average rate above 1.6. Interestingly, even the worst-performing individual score, the naive S_h, enhances S_{Tg}'s performance when combined via the arithmetic mean *(Mean$_{0.5}$)*, increasing the rate from 1.108 to 2.455.

Overall, the geometric mean *(GeoMean)* offers the best average rate performance across combinations, yielding the best combination in terms of priority across: validity, uniqueness, InDiv, SNN, and the number of novel structures. Moreover, the polymer with the highest T_g obtained without ER using a score combination is 793 K, in contrast to the best T_g achieved with a single scoring function without ER, which is 777.4 K.

6.2.2. With Experience Replay (Table 7). The best-performing agent using only the score S_{Tg} achieves a rate of 2.148 with ER. Among the 16 agents combining scores with ER, only six configurations outperformed this baseline. The remaining combinations failed to bring notable improvements, demonstrating that score aggregation does not universally lead to better results. Nonetheless, for each of the individual scores (S_s, S_r, S_m, and S_h), at least one aggregation function enabled a performance exceeding that of S_{Tg} alone, once again underscoring the importance of the aggregation function in effectively exploiting score complementarities.

As in the analysis without ER, the best overall performance with ER is achieved by combining S_{Tg} with the single Tanimoto score S_s, reaching a rate of 2.675, and the best novelty (0.760).

From the perspective of the *rate* metric, the arithmetic *(Mean$_{0.2}$)* and harmonic *(HarmMean)* means never yielded the top performance in any configuration. However, they consistently outperformed the baseline agent using only S_{Tg}, confirming their robustness and reliability across diverse settings. In addition, the *HarmMean* achieves the highest InDiv (0.511).

The exponential aggregation *(ExpAgg)* delivered the best overall performance with (S_{Tg}, S_s) and (S_{Tg}, S_m), maximizing their novelty (0.760, and 0.708) and SNN (0.839, and 0.865). Its poor result arose when combined with the naive high-T_g score S_h, as expected due to the low individual performance of S_h and ExpAgg's sensitivity to low values. Conversely, the geometric mean *(GeoMean)* exhibited near-optimal behavior across all score pairings. It consistently balanced the contributions of both inputs, maintaining performance without amplifying weaknesses. Its effectiveness in combination with the naive high-T_g score S_h further highlights this robustness, as it achieved the best performance in that case with the highest number of generated structures (36) and the best InDiv (0.445).

However, the polymer with the highest T_g was obtained using the single S_{Tg} score with ER (797.7 K), exceeding the best result from any score combination (793 K) with ER.

6.2.3. Key Insights in Score-combination Effects on Exploration and Exploitation (Table 8). Table 8 provides a summarized overview of the regimes observed under different score combinations, highlighting the best rate and the corresponding aggregation function. It can be seen that the *ExpAgg* outperforms the other aggregation functions, both with and without ER. The most robust pair of scores, which maintains good overall performance regardless of the aggregation function, consists of (S_{Tg}, S_s). This pair provides a well-balanced regime, where the balance behavior of S_s compensates for the highly exploratory nature of S_{Tg}. Concerning the maximal T_g values obtained, they all lie just below 800 K due to the limitation of the predictive model, which is effective only within the [400, 800]K range. As a result, the exploration cannot exceed this upper bound: even if the model identifies a polymer that would have a higher true T_g, it will inevitably be underestimated.

Table 8. Summarized Impact of Score Combinations on Exploration Vs. Exploitation, with and without Experience Replay (ER)[a]

Scores	Global Regime	ER = No	ER = Yes
S_{Tg} only	Extreme Exploration	1.108	2.148
S_{Tg} & S_h	S_h compensates S_{Tg} via stronger Exploi.	Mean, 2.455	GeoMean, 2.543
S_{Tg} & S_s	Balanced high-T_g Explor./Exploi.	**ExpAgg, 2.910**	**ExpAgg, 2.675**
S_{Tg} & S_r	More Explor. than Exploi.	GeoMean, 2.504	GeoMean, 2.155
S_{Tg} & S_m	Hybrid Explor./Exploi.	HarmMean, 2.007	ExpAgg, 2.391

[a] The table reports the best rate obtained and the corresponding aggregation function for each case.

6.2.4. Analysis of Generated Polyimides (Figure 20). Figure 20 shows the t-SNE projection of the Morgan fingerprint embeddings for all novel polyimides (i.e., not present in the training set) generated using score combination methods, with or without Experience Replay (ER), and with a predicted T_g > 750 K. A clear clustering can be observed, particularly for the polyimides with the highest T_g values. This indicates that the different score combinations guide the agent toward both diverse and similar regions of the latent space. Some of the top-performing polyimides are shared across multiple scoring combinations, and they overlap in Figure 20.

Moreover, when comparing these results to those obtained from agents trained with a single scoring function (with or without ER), a notable increase in diversity is observed: agents using single scores generated 54 unique novel polyimides, while those using score combinations generated 166, from which, only 27 are common between the two sets. This means that 50% of the polyimides generated using single scores are not found when using score combinations, and 83% of the polyimides obtained through score combinations are entirely novel. This highlights the enhanced exploration/exploitation of the latent space enabled by combining scoring strategies.

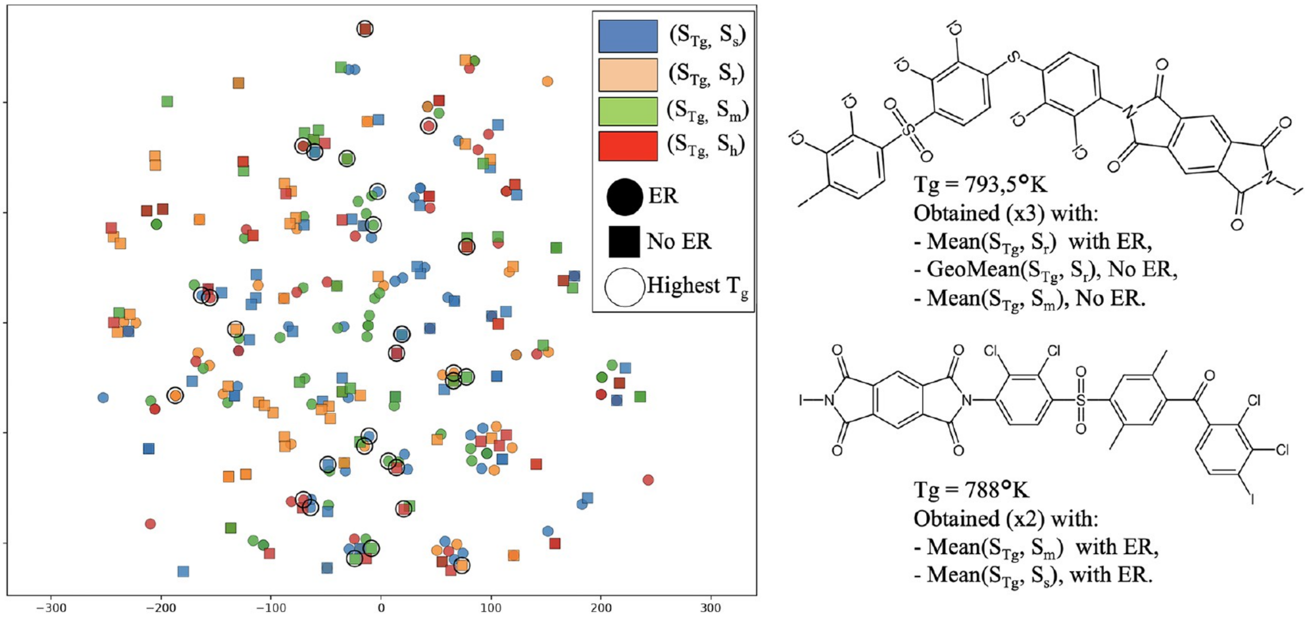
6.3. Further Analysis of the *Mean$_\alpha$* Function with $\alpha \neq 0.5$

So far, we have analyzed the weighted mean function *Mean$_\alpha$* using only $\alpha = 0.5$. However, it is important to study the impact of varying α on the results. From the experiments without ER (Table 6) and with ER (Table 7), we observe that the *Mean$_{0.5}$* function contributes most significantly when combining S_h with S_{Tg} in the **absence of ER** [*Mean$_{0.5}$*(S_h, S_{Tg}) = 2.455]. However, in this configuration, the aggregation gives equal weight to both scores, assuming that an equitable contribution provides a perfect balance between the high exploration of S_{Tg} and the high exploitation of S_h. We therefore focus on the score combination defined as (eq 25):

$$\text{Mean}_\alpha(S_h, S_{Tg}) = \alpha S_h + (1 - \alpha)S_{Tg} \tag{25}$$

and evaluate its performance for multiple values of $\alpha \in \{0.1, 0.2, ..., 0.9\}$. The results, presented in Table 9, reveal that

- The *Best Rate* does not follow a strictly monotonic trend for $\alpha \leq 0.6$; the highest performance is achieved at $\alpha = 0.2$, followed by a drop at $\alpha = 0.4$, before increasing again around $\alpha = 0.5$ and $\alpha = 0.6$.
- However, considering that S_h is meant to be complementary to S_{Tg}, it is reasonable to restrict α to values ≤ 0.5.



Tg = 793,5°K
Obtained (x3) with:
- Mean(S_{Tg}, S_r) with ER,
- GeoMean(S_{Tg}, S_r), No ER,
- Mean(S_{Tg}, S_m), No ER.

Tg = 788°K
Obtained (x2) with:
- Mean(S_{Tg}, S_m) with ER,
- Mean(S_{Tg}, S_s), with ER.

Figure 20. t-SNE projection of top-T_g (>750 K) polyimides generated with experience replay (ER, round markers) and without ER (squares), using score combinations. Colors indicate the score pairs used, independently of the mean function. The highest T_g in each colors is highlighted with a circle (the same optimum is sometimes found by two different mean functions).

Table 9. Weighted Mean Arithmetique Function for *α in*]0, 1[[a]

Metric	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Rate(g)	1.893	**2.753**	2.555	1.591	2.455	2.692	1.306	1.092	0.721
Novelty	0.545	0.625	0.472	0.473	0.640	**0.885**	0.666	1.000	1.000
Uniqueness	1.000	0.969	0.900	1.000	0.893	0.885	1.000	1.000	1.000
InDiv	0.420	0.436	0.440	0.411	0.420	0.414	0.461	0.359	0.000
SNN	0.902	0.868	0.918	0.908	0.875	0.881	0.853	0.795	0.788
Validity	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lvert D_g \rvert$	22	33	**40**	19	28	35	9	4	1
best g	39	50	44	52	80	92	92	63	59
Highest $T_g(K)$	775.2	781.7	781.7	776.7	**783**	778.2	771.1	764.6	766.56

[a] Metrics computed on generated polymers (T_g > 750 K). Experience replay = No.

Within this range, the achieved rates are consistently higher than those obtained from agents trained with S_h or S_{Tg} individually (which yield rates of 0.600 and 1.108, respectively).

- Surprisingly, for $\alpha = 0.6$, the rate is higher than for $\alpha = 0.5$, which is slightly counterintuitive, since S_h is expected to be only exploitative. Nevertheless, for $\alpha \geq 0.7$, the behavior becomes predictable, with a gradual decline in the rate.

As a result, using Mean$_\alpha$ requires varying the parameter α across all ER configurations to identify the best combination, which is tedious compared with the other methods. Additionally, it is interesting to investigate whether a triple combination among different scores is likely to improve overall performance, which we analyze in the following subsection.

6.4. Combining Three Scores

In the context of combining three scoring functions, it becomes essential to identify an appropriate strategy to effectively integrate multiple signals, given the heterogeneous nature of our aggregation functions. Previous results have highlighted that the best performances were achieved using the following two-score combinations (eq 26) **without ER**:

$$\mathrm{ExpAgg}(S_s, S_{Tg}) = 2.910, \ \mathrm{Mean}_{0.2}(S_h, S_{Tg}) = 2.753 \quad (26)$$

Given these promising outcomes, it is worth investigating whether a **triple score combination** could further enhance performance. In particular, we focus on the fusion strategies based on functions such as *ExpAgg*, *GeoMean*, and *Mean$_\alpha$* (with $\alpha = 0.2$, as it provides the best weighting).

We therefore define a set of triple-score combinations designed to reflect various fusion strategies. These include both direct extensions (S_I, S_{II}, S_{III}) of aggregation functions to three variables and nested compositions (S_I°, S_{II}°, S_{III}°) of pairwise functions. This allows us to explore not only additive blending but also how the structure of the combination influences the learning process (eq 27)

$$\begin{aligned}
S_I &= \mathrm{Mean}_{ex}(S_s, S_{Tg}, S_h) = \frac{S_s + S_{Tg} + S_h}{3}\\
S_I{}^\circ &= \mathrm{Mean}(S_s, \mathrm{Mean}_{0.2}(S_h, S_{Tg}))\\
&= \frac{S_s + (0.2S_h + 0.8S_{Tg})}{2}\\
S_{II} &= \mathrm{GeoMean}_{ex}(S_s, S_{Tg}, S_h) = \sqrt{S_s \cdot S_{Tg} \cdot S_h}\\
S_{II}{}^\circ &= \mathrm{GeoMean}(S_s, \mathrm{Mean}_{0.2}(S_h, S_{Tg}))\\
&= \sqrt{S_s \cdot (0.2S_h + 0.8S_{Tg})}\\
S_{III} &= \mathrm{ExpAgg}_{ex}(S_s, S_{Tg}, S_h) = \exp(\ln S_s \cdot \ln S_h \cdot \ln S_{Tg})\\
S_{III}{}^\circ &= \mathrm{ExpAgg}(S_s, \mathrm{Mean}_{0.2}(S_h, S_{Tg}))\\
&= \exp(-\ln S_s \cdot \ln(0.2S_h + 0.8S_{Tg}))
\end{aligned} \tag{27}$$

The results presented in Table 10 indicate that a simple arithmetic mean (S_I) of the three scoring functions does not yield the best performance, although it still provides reasonably good results. In contrast, the weighted mean derived from the composite form (S_I°) leads to less favorable outcomes.

Table 10. Combination of Three Scoring Fonction with Different Strategies[a]

Metric	S_I	S_I°	S_{II}	S_{II}°	S_{III}	S_{III}°
Rate(g)	2.101	1.764	**2.654**	2.502	0.717	2.005
Novelty	0.666	0.913	0.633	0.562	1.000	0.764
Uniqueness	1.000	0.913	0.967	1.000	1.000	1.000
InDiv	0.436	0.426	0.437	**0.455**	0.000	0.448
SNN ↓	0.863	0.898	0.866	0.891	0.821	0.837
Validity	1.000	1.000	1.000	1.000	1.000	1.000
$\lvert D_g \rvert$	21	23	31	**32**	1	17
best g	54	50	50	58	21	91
Highest $T_g(K)$	784.4	769.1	**793.4**	783.7	769.6	782.6

[a] Metrics computed only on generated polymers (T_g > 750K). Experience replay= No.

As expected, the geometric mean (S_{II}) achieves the most significant performance improvements. However, when composed with another mean function as in (S_{II}°), its effectiveness slightly declines.

On the other hand, extending the *ExpAgg* function to three variables (S_{III}) results in poor performance. Nevertheless, its composite form (S_{III}°), where *ExpAgg* is composed with $Mean_{0.2}$, achieves better results.

In conclusion, combining **two scores** proves to be more effective overall than merging three scores. The model tends to benefit more from targeted pairwise combinations that preserve the complementarity between scoring objectives.

6.5. Comparison Between the Aggregation Functions and the Chebyshev Multi-Objective Optimization

In order to evaluate our result with classical approch inspire from multiobjectif optimization, to ensure comparability, we apply the Chebyshev function at the score aggregation level rather than at the reward level.

The Chebyshev scalarization[9,41] is a nonlinear aggregation method originally introduced in multiobjective optimization and later adopted in multiobjective RL. Unlike linear scalarization, which can only recover Pareto-optimal solutions in convex regions of the Pareto front, the Chebyshev method can identify optimal solutions regardless of the front's shape and generally provides a better spread of Pareto-optimal solutions while being less sensitive to weight selection.

The Chebyshev scalarization is based on an L_∞ metric that measures the distance between a point in the multiobjective space and a reference point z^*. For a set of n scores $S_1, S_2, ..., S_n$ with corresponding weights $w_1, w_2, ..., w_n$ (with $\sum_{i=1}^{n} w_i = 1$) and reference point $z^* = (z_1^*, ..., z_n^*)$ (typically chosen as the ideal point), the Chebyshev scalarized score is defined as (eq 28):

$$\mathrm{ChebAgg}(S_1(x), ..., S_n(x)) = \max_{i=1,...,n} \{w_i \cdot \lvert z_i^* - S_i(x) \rvert\} \tag{28}$$

Since all scores are normalized in [0, 1] and the objective is maximization, we set $z^* = (1, ..., 1)$ and compute $1 - \mathrm{ChebAgg}(x)$ to transform the distance minimization into a maximization problem. In this formulation, the aggregated score is driven by the lowest score at each generation, thereby constraining the model to balance both objectives simultaneously and enforcing an exploratory regime on the fused scores. We evaluated this approach by comparing it to our best-performing aggregation strategy, $\mathrm{ExpAgg}(S_s, S_{T_g})$, by computing $\mathrm{ChebAgg}(S_s, S_{T_g})$.

The results (Table 11) show that Chebyshev scalarization does not surpass our best aggregation method, which outperforms it in terms of Rate (2.910 vs 2.200), novelty (0.786 vs 0.424), and IntDiv (0.473 vs 0.429), with or without ER.

Table 11. Score Combination (S_s, S_{Tg}) Comparison with Chebyshev Multi-Objectif Formulation

Metric	ER = No		ER = Yes	
	S_{Cheby}	*ExpAgg*	S_{Cheby}	*ExpAgg*
Rate(g)	2.200	**2.910**	1.493	**2.675**
Novelty	0.424	**0.786**	0.444	**0.760**
Uniqueness	0.916	**1.000**	**1.000**	0.962
InDiv	0.446	**0.458**	0.429	**0.473**
SNN ↓	0.913	**0.866**	**0.818**	0.839
Validity	1.000	1.000	1.000	1.000
$\lvert D_g \rvert$	**36**	28	18	**26**
best g	44	62	41	85
Highest $T_g(K)$	789.5	788	**796.2**	775

However, Chebyshev generates more polymers without ER (36 vs 28), but with a high level of redundancy due to its lower novelty. If we consider only novel polymers, this corresponds to approximately 15 candidates (36 × 0.424) for ChebAgg versus 22 candidates (28 × 0.786) for ExpAgg, which changes the balance of the observation. Furthermore, with ER, Chebyshev achieves better results in terms of SNN and Uniqueness.

That being said, this comparison provides a useful reference with respect to aggregation strategies used in multiobjective RL and opens the way for future work.

7. HIGH-T_G POLYIMIDE GENERATION: CHEMICAL ANALYSIS

Figure 21 presents the most promising and novel generated polyimide, denoted as P^*, which exhibits the highest predicted T_g. It is absent from major polymer databases such as PolyInfo[42] and PubChem,[43] highlighting its novelty.

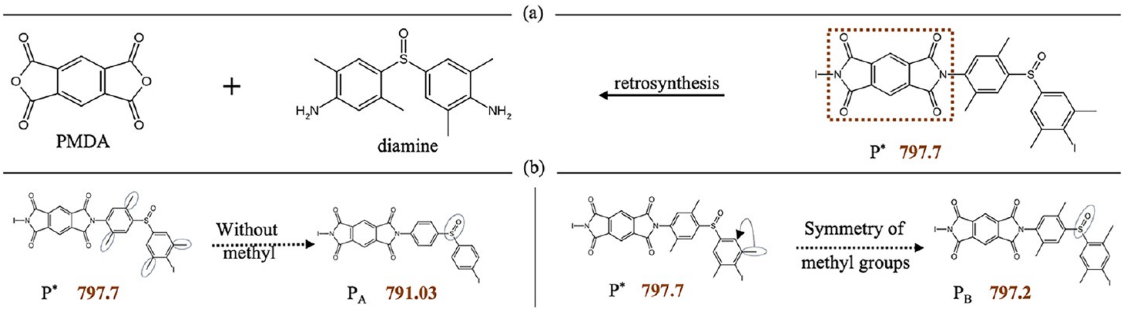


Figure 21. Retrosynthesis of the best generated polyimide (P*).

Table 12. Strategic Guide for Selecting Mean Functions Based on Generation Objectives and Performance Profiles

Mean Function	Robustness	Maximize peak performance	Score balancing	Chemical diversity	Scalability (3+ scores)	Requires tuning
$Mean_\alpha$	**	**	***	**	**	Yes
GeoMean	***	***	***	***	***	No
HarmMean	**	**	**	**	-	No
ExpAgg	**	***	*	**	*	No

The substructure highlighted in the square originates from the dianhydride PMDA (pyromellitic dianhydride), which consistently appears in all generated high T_g polyimides despite representing only 1.3% of the training data set (Section 3.1). Its presence was expected, as PMDA is well-known for its strong ability to induce elevated T_g compared to other dianhydrides.[44,45] This suggests that the generative model has automatically identified PMDA as a critical structural factor.

Regarding the associated diamine, it represents the limiting factor for synthesizability, as it is not symmetric. We can therefore search for diamine variants that are more synthesizable while maintaining the same Tg level. We propose either removing the methyl groups or making them symmetric. Figure 21 b shows the impact of these transformations on Tg, and we retain the two approximate variants (P_A, and P_B).

Moreover, the 193 generated structures are highly diverse, as the experimental parameters, such as experience replay (ER), scoring function and aggregation method, allow the model to exploit key regions while still exploring other areas of the latent space (Figure 20). Future work will analyze the synthesizability of the generated polyimides through retrosynthetic analysis, particularly focusing on the symmetry of methyl groups and the functional bonds (SO and SO_2) of the associated diamines, and will introduce a dedicated synthesizability score.

8. SUMMARY OF RESULTS

The results in Tables 7 and 6 show the best-performing generation (highest Rate(g)) among 100 for each experiment. Since each aggregation function was tested with 4 score pairs under two conditions (with and without ER), this yields 8 experiments per function, each selecting the top generation from 100. Table 13 provides a complementary statistical evaluation by averaging all metrics across all 800 generations (8 experiments × 100 generations).

Table 12 and Figure 22 summarize the key findings from our study of score combinations and mean functions for the generation of high T_g polyimides (T_g > 750 K). We use a star-based scale to indicate practical suitability, where *** denotes highly effective or recommended, ** indicates moderate or context-dependent suitability, and * refers to limited applicability or low performance under most conditions. Several patterns emerge:

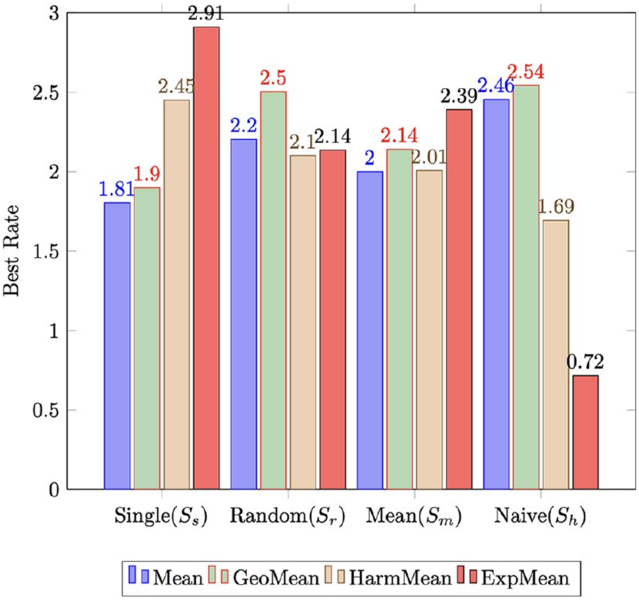


Figure 22. Comparison of “Best Rate” Based on S_{Tg} Score Combinations with Other Scores Using Different Aggregation Functions Across All best Setups.

- The geometric mean, *GeoMean* emerges as the most suitable method for combining scores, consistently demonstrating the highest effectiveness across all setups with the best average Rate. Even in three-score combinations, it successfully compensates for the mutual weaknesses of individual scores (Tables 12 and 13).
- The exponential aggregation, *ExpAgg*, provides the best performance **only** when both input scores are already strong (Figure 22). It heavily penalizes poorly scoring components, as evidenced by its weak performance when combined with the naive high-T_g score (S_h). This explains why it exhibits the lowest average values across most metrics over the 800 generations (see Table 13), with the exception of SNN, for which it achieves the best performance, indicating that it generates structures that are the most distant from the training set. Additionally, it can be observed that *ExpAgg* exhibits the highest standard

Table 13. Statistical Evaluation of All 800 Generation (with T_g > 750*K*), with and without ER (See Tables 6 and 7)

Mean ± std	*$Mean_{0.5}$*	**GeoMean**	HarmMean	ExpAgg
$\overline{\text{Rate}}$	0.894 ± 0.404	**0.928** ± 0.456	0.837 ± 0.360	0.855 ± **0.611**
$\overline{\text{Novelty}}$	0.237 ± 0.230	**0.277** ± 0.240	0.245 ± 0.238	0.266 ± **0.261**
$\overline{\text{Uniqueness}}$	**0.914** ± 0.258	0.888 ± 0.287	0.894 ± 0.286	0.702 ± **0.436**
$\overline{\text{InDiv}}$	**0.367** ± 0.183	0.359 ± 0.182	0.355 ± 0.184	0.308 ± **0.221**
$\overline{\text{SNN}}$↓	0.885 ± 0.247	0.854 ± 0.275	0.863 ± 0.273	**0.670** ± **0.415**
$\overline{\text{Validity}}$	**0.930** ± 0.255	0.909 ± 0.288	0.911 ± 0.285	0.725 ± **0.447**
$|\overline{D_g}|$	7.701 ± **7.899**	**8.223** ± 7.882	6.315 ± 5.561	7.414 ± 7.614

deviation (std) across all metrics, except for the number of generated polymers. This suggests that *ExpAgg* delivers very strong performance only under certain configurations.

- The weighted average, *$Mean_\alpha$*, with manually tuned score (α), lead to improvement in performance. It slightly outperforms the simple arithmetic mean ($\text{Mean}_{0.2}$ = 2.753 vs $\text{Mean}_{0.5}$ = 2.455), highlighting the potential benefit of adjusting score weights. In addition, it exhibits the highest average values observed over the 800 generations in terms of uniqueness, internal diversity (InDiv), and validity (Table 13).
- When aiming to maximize chemical diversity or novelty, strategies based on random Tanimoto, S_r combined with either the *HarmMean* or *GeoMean* are especially effective, as they lead to the highest *InDiv* values for most configurations.
- When combined with S_{Tg}, the Single Tanimoto score S_s exhibits the highest overall performance across all configurations, both with and without experience replay (Figure 22).
- Experience replay globally improves the metric parameters, but in an uneven manner across different configurations and metrics.

9. DISCUSSION

This study focused on the generation of polymers, specifically polyimides with high glass transition temperatures (T_g > 750*K*). Beyond the standard approach of defining a single score to guide the fine-tuning of an agent via reinforcement learning, we explored a range of alternative strategies aimed at constructing more diverse and informative scoring functions. Among these strategies, we introduced the *naïve* high-T_g score, based solely on the weighted sum of Shapley contributions from the T_g prediction model. Other approaches relied on Tanimoto similarity in three distinct modalities: a *simple* version (direct comparison to a target polymer), a *randomized* version (alternating between several targets), and a *mean* version (average similarity with a list of reference polymers).

These strategies enabled the generation of polymers with notable structural diversity while maintaining high T_g values. We further analyzed the impact of score combination methods, on both the quantity and quality of generated structures. Our findings show that score combinations tend to enhance structural diversity by balancing exploration and exploitation, albeit sometimes at the cost of maximizing T_g. Additionally, we introduced a novel score aggregation function based on exponential and logarithmic transformations. This *ExpAgg* function proved effective when combining scores that evolve at the same pace but remained fragile when applied to unbalanced pairs, where one score is significantly lower than the other at certain stages of training, hence the high standard deviations observed in the statistical study.

Limitation

However, a limitation of this study lies in the predictive range of the T_g model, which remains reliable only within the [400, 800] K interval. This restriction arises from the intrinsic behavior of tree-based ensemble methods, such as Random Forests, which are highly performant but have difficulties extrapolating beyond the distribution of their training data. Because these models partition the feature space into discrete regions, they tend to predict conservative values near the boundaries of the observed data, resulting in a "ceiling effect."

As a consequence, if the generative model proposes a polymer with a true T_g around 900 K, it is likely to be underestimated by the scoring function S_{Tg} and thus perceived as less favorable. This constraint limits the determination of which polymers have exact predicted T_g values in the range [750,800] or exceed it, explaining why the generated high-T_g candidates in this work have predicted T_g values mostly confined to the upper bound of the model's valid range.

However, the multiscore approach mitigates this bias: for structures with identical S_{Tg} values, Tanimoto-based scores provide clear structural differentiation, allowing the selection of the desired outcome based on a set of target polyimides.

Nevertheless, by targeting polyimides within the [400,800] K interval, the developed approach does not suffer from this specific limitation.

Perspectives

Future work will explore alternative predictive models with enhanced extrapolation capabilities, including piecewise approaches or mechanisms to safely handle predictions beyond the training range. In order to evaluate the influence of the Askadskii and Random Forest approximations, we performed the evaluation of the model on external experimental data; the results show promising outcomes on structures similar to the training set.

Future work will focus on performing MD simulations and laboratory synthesis to continue the assessment of the models. Although these results were demonstrated on high-T_g polyimide generation, the proposed framework may also be applicable to other molecular design tasks requiring multiobjective optimization through scalarization, pending further investigation.

We also investigated the use of the Chebyshev multiobjective optimization function, at the score level in our framework. The results show interesting outcomes but do not outperform our best results. This experimental analysis has been provided to pave the way for future work that will compare the proposed score combination strategies with multiobjective RL approaches,[6] where each score is implemented as an individual reward.

10. CONCLUSION

The design of high-performance polymers, particularly high-T_g polyimides, remains a significant challenge in materials science due to the vast chemical space and complex structure−property relationships. In this work, we systematically investigated how different score aggregation strategies influence molecular generation outcomes in reinforcement learning, using high-T_g polyimide design as a case study. A key strength of our approach lies in its modularity: the scoring function variants and the aggregation functions are independent, making the framework potentially transferable to other applications.

Our investigation of various combination strategies reveals that the choice of aggregation method significantly impacts both the novelty and diversity of generated high-T_g polyimides. Among the approaches explored, the *GeoMean* is the most robust, but our proposed exponential aggregation function (*ExpAgg*) demonstrated superior performance across several experimental settings. However, its effectiveness proved sensitive to the compatibility of the combined scores, making it less suitable in configurations where the scores are not well aligned. Given that it has been evaluated on our application, its performance on other structures and different RL architectures remains an open question for the community.

Our predictive model exhibits underestimation in the out-of-distribution high-T_g region (>800 K), a limitation inherent to tree-based ensemble methods. Despite this constraint, our multiscore framework successfully generated chemically reasonable high-T_g candidates, as validated through structural analysis and comparison with experimental literature on PMDA-based polyimides.

Our findings underscore the importance of carefully selecting aggregation methods when combining multiple scores in molecular RL frameworks. Future research will focus on evaluating score fusion strategies within a multiobjective reinforcement learning setting, with the aim of optimizing polymer generation across multiple simultaneous properties. Additionally, experimental validation through molecular dynamics simulations and synthesis represents an essential next step to verify the true T_g values of our top-performing candidates.

ASSOCIATED CONTENT

Data Availability Statement

The code and data used in this article are available on https://github.com/AMETHYST-2025/Multi-Score-RL-for-High-Tg-Polyimide-Design.

AUTHOR INFORMATION

Corresponding Author

Aymar Tchagoue − *INSA Lyon, UCBL, CNRS, LIRIS UMR 5205, 69100 Villeurbanne, France; INSA Lyon, UCBL, Université Jean Monnet, CNRS, IMP UMR 5223, 69100 Villeurbanne, France;* orcid.org/0009-0002-0250-2330; Email: aymar-lebeau.tchagoue-tchagoue@insa-lyon.fr

Authors

Véronique Eglin − *INSA Lyon, UCBL, CNRS, LIRIS UMR 5205, 69100 Villeurbanne, France*

Jean-Marc Petit − *INSA Lyon, UCBL, CNRS, LIRIS UMR 5205, 69100 Villeurbanne, France*

Sébastien Pruvost − *INSA Lyon, UCBL, Université Jean Monnet, CNRS, IMP UMR 5223, 69100 Villeurbanne, France;* orcid.org/0000-0002-3270-5668

Jannick Duchet-Rumeau − *INSA Lyon, UCBL, Université Jean Monnet, CNRS, IMP UMR 5223, 69100 Villeurbanne, France*

Jean-François Gérard − *INSA Lyon, UCBL, Université Jean Monnet, CNRS, IMP UMR 5223, 69100 Villeurbanne, France;* orcid.org/0000-0002-3096-2767

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.5c02807

Author Contributions

This work emerged from a collaboration between two research teams. The LIRIS team (A.T., V.E., and J.-M.P.) contributed to the design and development of the machine learning framework, including reinforcement learning strategies, scoring function combination methods, and computational experiments. The IMP team (A.T., S.P., J.D.-R., and J.-F.G.) contributed essential polymer domain knowledge, validation of chemical structures and property predictions, and interdisciplinary insights on polyimide design principles. Aymar Tchagoue bridged both teams, integrating machine learning methodologies with polymer science expertise. All authors contributed to the interpretation of results, manuscript preparation, and provided valuable feedback throughout the revision process.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the French government's France 2030 initiative through the DIADEM program (AMETHYST Project), which is managed by the Agence Nationale de la Recherche (AMETHYST, ANR-22-PEXD-0004).

REFERENCES

(1) Sroog, C. E. Polyimides. *Prog. Polym. Sci.* **1991**, *16*, 561−694.

(2) Matsuura, T.; Ando, S.; Sasaki, S. *Fluoropolymers 2*; Hougham, G.; Cassidy, P. E.; Johns, K.; Davidson, T., Eds.; Topics in Applied Chemistry; Springer: Boston, MA, 2002; pp 373−393.

(3) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360−365.

(4) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828−849.

(5) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de novo design through deep reinforcement learning. *J. Cheminf.* **2017**, *9*, 48.

(6) Zhang, L.; Qi, Z.; Shi, Y. Multi-objective Reinforcement Learning Concept, Approaches and Applications. *Procedia Computer Science* **2023**, *221*, 526−532.

(7) Al-Jumaily, A.; Mukaidaisi, M.; Vu, A.; Tchagang, A.; Li, Y. Examining multi-objective deep reinforcement learning frameworks for molecular design. *Biosystems* **2023**, *232*, No. 104989.

(8) Winter, R.; Montanari, F.; Noé, F.; Clevert, D.-A.et al. Efficient multi-objective molecular optimization in a continuous latent space. *Chem. Sci.* **2019**10,. DOI: 10.1039/C9SC01928F.

(9) Moffaert, K. V.; Drugan, M. M.; Nowé, A. Scalarized Multi-objective Reinforcement Learning: Novel Design Techniques. In *2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*; Singapore, 2013; pp 191−199.

(10) Kong, L.; Yang, C.; Neufang, S.; Beyan, O. D.; Boukhers, Z. EMORL: Ensemble Multi- Objective Reinforcement Learning for

Efficient and Flexible LLM Fine-Tuning. 2025, arXiv:2505.02579. arXiv.org e-Print archive https://arxiv.org/abs/2505.02579.

(11) Merz, K. M. J.; De Fabritiis, G.; Wei, G.-W. Generative Models for Molecular Design. *J. Chem. Inf. Model.* **2020**, *60*, 5635−5636.

(12) Kuenneth, C.; Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat. Commun.* **2023**, *14*, No. 4099.

(13) Morehead, A.; Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **2024**, *7*, 150 DOI: 10.1038/s42004-024-01233-z.

(14) Cornet, F.et al. Equivariant Neural Diffusion for Molecule Generation. In *NeurIPS 2024 Proceedings*, 2024.

(15) Yue, T.; Tao, L.; Varshney, V.; Li, Y. Benchmarking study of deep generative models for inverse polymer design. *Digital Discovery* **2025**, *4*, 910−926.

(16) Nigam, A.; Pollice, R.; Tom, G.; Jorner, K.; Willes, J.; Thiede, L. A.; Kundaje, A.; Aspuru-Guzik, A. TARTARUS: A Benchmarking Platform for Realistic And Practical Inverse Molecular Design. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks*, 2023; Open-Review/arXiv preprint arXiv:2209.12487.

(17) Blaschke, T.; Ars-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for De Novo Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 5918−5922.

(18) Loeffler, H. H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L. H.; Engkvist, O. Reinvent 4: Modern AI−driven generative molecule design. *J. Cheminf.* **2024**, *16*, 20.

(19) Sutton, R. S.; Barto, A. G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, 2015.

(20) Guo, J.; Schwaller, P. Augmented Memory: Sample-Efficient Generative Molecular Design with Reinforcement LearningClick to copy article link. *J. Am. Chem. Soc.* **2024**, *4*, 2160.

(21) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminf.* **2015**, *7*, 20.

(22) Blaschke, T.; Engkvist, O.; Bajorath, J.; Chen, H. Memory-assisted reinforcement learning for diverse molecular de novo design. *J. Cheminf.* **2020**, *12*, 68.

(23) Li, W.; Li, Y.; Lei, Q.; Wang, Z.; Wang, X. PolyRL: Reinforcement Learning-Guided Polymer Generation for Multi-Objective Polymer Discovery. ChemRxiv, Cambridge Open Engage, Version 1, 20252025; Working paper, posted 05 June 2025.

(24) Xu, Y.; Feng, W.; Gao, L.; Wang, L.; Lin, J.; Ye, X.; Liang, L.; Du, L. De Novo Design of Polyimides Leveraging Deep Reinforcement Learning Agent. *Adv. Mater.* **2025**, *38*, e11099 DOI: 10.1002/adma.202511099.

(25) Fromer, J. C.; Coley, C. W. Computer-aided multi-objective optimization in small molecule discovery. *Patterns* **2023**, *4*, 100678.

(26) Jumaily, A. A.; Mukaidaisi, M.; Vu, A.et al. Exploring Multi-Objective Deep Reinforcement Learning Methods for Drug Design. In *Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2022.

(27) Wang, J.; Zhu, F. Multi-objective molecular generation via clustered Pareto-based reinforcement learning. *Neural Networks* **2024**, *179*, No. 106596.

(28) Volgin, I. V.; Batyr, P. A.; Matseevich, A. V.; Dobrovskiy, A. Y.; Andreeva, M. V.; Nazarychev, V. M.; Larin, S. V.; Goikhman, M. Y.; Vizilter, Y. V.; Askadskii, A. A.; Lyulin, S. V. Machine Learning with Enormous "Synthetic" Data Sets: Predicting Glass Transition Temperature of Polyimides Using Graph Convolutional Neural Networks. *ACS Omega* **2022**, *7*, 43678−43691.

(29) Askadskii, A. A. Methods for Calculating the Physical Properties of Polymers. *Rev. J. Phys. Chem. B* **2015**, *5*, 83−142.

(30) Askadskii, A. A. *Computational Materials Science of Polymers*, 1st ed.; Cambridge International Science Publishing: Cambridge, 2003.

(31) Tchagoue, A.; Eglin, V.; Petit, J.-M.; Pruvost, S.; Rumeau, J.; Gerard, J.-F. Dual Embedding: AFine-Tuned Language Model Approach for Accurate Polymer Glass Transition Temperature Prediction. *J. Chem. Inf. Model.* **2025**, *65*, 12342 DOI: 10.1021/acs.jcim.5c02469.

(32) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.

(33) Landrum, G. et al. RDKit: Open-source cheminformatics. https://github.com/rdkit/rdkit, 2020; https://www.rdkit.org.

(34) Yu, X.; Fang, Z.; Wu, F. Random Forest Based Approach for Predictions of Glass Transition Temperatures in Polymers. *Polym. Eng. Sci.* **2025**, *65*, 6238 DOI: 10.1002/pen.70128.

(35) Hsiao, S.-H.; Chen, Y.-J. Structure-Property Study of Polyimides Derived from PMDA and BPDA Dianhydrides with Structurally Different Diamines. *Eur. Polym. J.* **2002**, *38*, 815−828.

(36) Chern, Y.-T.; Shiue, H.-C. High Subglass Transition Temperatures and Low Dielectric Constants of Polyimides Derived from 4,9-Bis(4-aminophenyl)diamantane. *Chem. Mater.* **1998**, *10*, 210−216.

(37) Sroog, C. E. Polyimides. *Prog. Polym. Sci.* **1991**, *16*, 561−694.

(38) Lundberg, S. M.; Lee, S.-I.A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)* 2017; pp 4765−4774.

(39) Mannor, S.; Shimkin, N. A Geometric Approach to Multi-Criterion Reinforcement Learning. *J. Machine Learning Res.* **2004**, *5*, 325−360.

(40) Cheniti, M.; Akhtar, Z.; Adak, C.; Siddique, K. An Approach for Full Reinforcement-Based Biometric Score Fusion. *IEEE Access* **2024**, *12*, 49779 DOI: 10.1109/ACCESS.2024.3384544.

(41) Chugh, T.Scalarizing Functions in Bayesian Multiobjective Optimization. In *Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC)*, 2020; pp 1−8.

(42) Ishii, M.; Takemura, T.; Tanifuji, M.PoLyInfo RDF: A semantically reinforced polymer database for materials informatics. In *CEUR Workshop Proceedings* 2019; Vol. 2456, p 18.

(43) Kim, S.; Thiessen, P. A.; Cheng, T.; Zhang, J.; Gindulyte, A. et al. PubChem in 2021: New data content and improved web interfaces, 2021.

(44) Zhang, P.; Zhang, K.; Dou, S.; Zhao, J.; Yan, X.; Li, Y. Mechanical, Dielectric, and Thermal Attributes of Polyimides Stemmed Out of 4,4'-Diaminodiphenyl Ether. *Crystals* **2020**, *10*, 173.

(45) Kim, S. I.; Pyo, S. M.; Kim, K.; Ree, M. Investigation of glass transition behaviours in aromatic poly(amic acid) precursors with various chain rigidities by oscillating differential scanning calorimetry. *Polymer* **1998**, *39*, 3063−3072.