

<https://doi.org/10.1038/s41524-025-01683-6>

An informatics framework for the design of sustainable, chemically recyclable, synthetically accessible, and durable polymers

Joseph Kern¹, Yong-Liang Su², Will Gutekunst² & Rampi Ramprasad¹ ✉

We present a novel approach to designing durable and chemically recyclable ring-opening polymerization (ROP) class polymers. This approach employs digital reactions using virtual forward synthesis (VFS) to generate over 7 million ROP polymers and machine learning techniques to rapidly predict thermal, thermodynamic, and mechanical properties crucial for performance and recyclability. This methodology enables the generation and evaluation of millions of hypothetical ROP polymers from known and commercially available molecules, guiding the selection of approximately 35,000 candidates with optimal features for sustainability and utility. Three of these recommended candidates have passed validation tests in the physical lab — two of the three by others, as published previously elsewhere, and one of them is a new thiocane polymer synthesized, tested, and reported here. This paper highlights the potential of VFS and machine learning to enable a large-scale search of the polymer universe and advance the development of recyclable and environmentally benign polymers.

Plastics, central to everyday life and pivotal in diverse applications ranging from food packaging to electronic components, cause a concerning amount of environmental pollution. Studies reveal pervasive microplastic contamination globally that negatively impacts humans, plants, and animals^{1–5}. The quest for sustainable alternatives that balance the beneficial attributes of plastics (such as cost-effectiveness, durability, and performance) with environmental considerations (such as recyclability and reduced ecological footprint) is a significant focus in contemporary materials development.

Creation of new plastics amenable to chemical recycling, i.e., transforming them back to monomers at the end of their life, will be enormously beneficial. Traditional mechanical recycling methods suffer from degradation limits⁶, whereas chemical recycling promises near-infinite recyclability. However, polymers that can undergo chemical recycling must still meet the demands of their application needs. A new take-out container that is recyclable sounds attractive, but consumers will not use it if it breaks after holding just a single item or if it is too expensive.

As depicted in Fig. 1a, polymer design can be complex. Various classes of properties must be considered as they profoundly influence the polymer's performance within consumer and industrial environments. These encompass thermal properties, such as the glass transition temperature (T_g) and melting temperature (T_m), which not only dictate stability at operational

temperatures but also affect processing conditions. Additionally, mechanical properties like Young's modulus (E), tensile strength at break (σ_b), and elongation at break (ϵ_b) play crucial roles in determining the polymer's stiffness, strength, and stretchability. Certain applications, such as take-out containers for food, demand thermal insulation properties to prevent heat transfer or minimize the risk of burn injuries upon contact, while others, such as plastic food wrap, necessitate specific gas permeabilities. Adding thermodynamic properties such as the enthalpy of polymerization (ΔH) and ceiling temperature (T_c), essential to assess chemical recyclability, complicates the design process further.

In this work, our focus is on the design of suitable chemically recyclable polymers for a specific application, namely, a replacement material for polystyrene (PS) used in containers. This focus is significant considering PS's notable presence in plastic production, its recyclability challenges, and its associated environmental and health concerns. PS constituted 6.8% of plastic production in Europe in 2019⁷, while the U.S. alone saw the creation of 220 thousand tons of PS containers, bags, sacks, and wraps in 2018⁸. Such substantial volumes are required for recyclable alternatives, as economies of scale often dictate affordability⁹. While PS is technically recyclable, it is not commonly recycled due to prohibitive costs^{10,11}. Styrene, the monomer used in its production, is also classified as “reasonably anticipated to be a human

¹School of Materials Science and Engineering, College of Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ²School of Chemistry & Biochemistry, College of Sciences, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: rampi.ramprasad@mse.gatech.edu

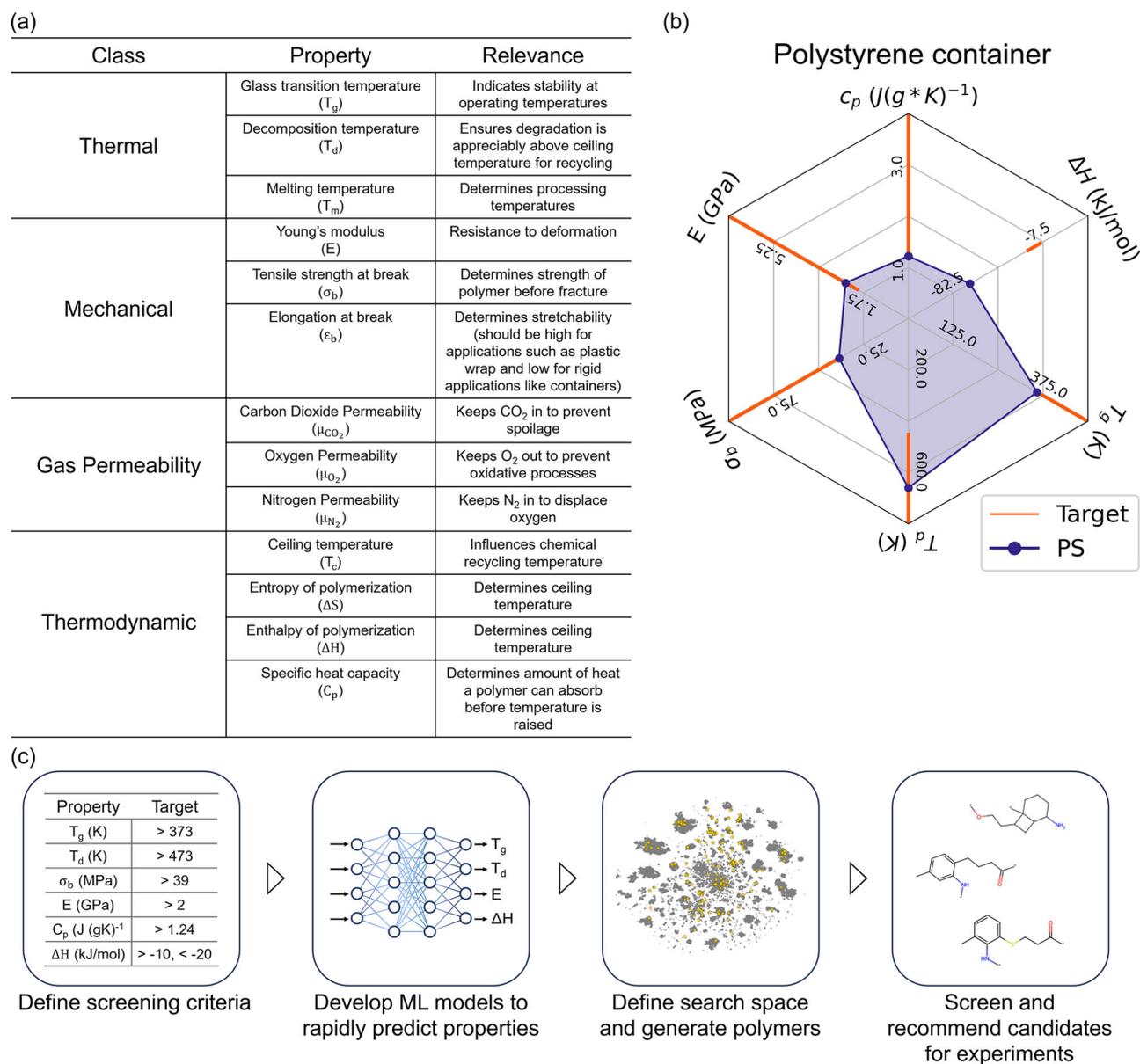


Fig. 1 | Informatics-driven polymer design approach. **a** Classification of key polymer properties, including specific properties within each class, and their relevance. **b** Radar chart illustrating the design targets for a polystyrene container, highlighting polystyrene properties in blue and target values in orange^{17,72,73}.

c Overview of the material informatics workflow, starting with the definition of screening criteria, followed by the development of ML models, defining of the polymer search space being explored, and ending with the screening and recommendation of suitable candidates for further experimental validation and testing.

carcinogen” by the Department of Health and Human Services (DHHS) and the National Toxicology Program (NTP)¹². PS microplastics have also been identified as potential immune system stimulants and toxic to freshwater organisms^{13,14}.

Although compostable or recyclable alternatives to PS, such as poly(lactic acid) (PLA) and poly(butyl succinate) (PBS), are available, they often exhibit limitations that render them unsuitable for specific applications. Notably, PLA’s glass T_g of 331 K and PBS’s even lower T_g range (230–263 K) make them ill-suited for moderate heat exposure, highlighting the need for more robust recyclable alternatives^{15,16}.

In Fig. 1b, a radar chart illustrating the properties of PS (depicted as blue dots) alongside the specific property targets set for our design (represented by orange lines) is presented. These targets, explicitly outlined in the first box of the informatics workflow of Fig. 1c, were carefully selected through a comprehensive analysis of PS properties, coupled with considerations of the typical operating conditions of a container.

Concerning thermal properties, our design prioritizes a T_g value surpassing the boiling point of water (373 K), ensuring the container’s integrity under operational conditions typical of PS containers. Additionally, maintaining the polymer in a glassy state below this temperature is essential for structural rigidity. A decomposition temperature (T_d) set 100 K above the boiling point of water is chosen to prevent decomposition during thermally induced chemical recycling.

Addressing mechanical considerations, we establish a σ_b exceeding 39 MPa to ensure the container’s resistance to breakage when subjected to typical loads. Similarly, a minimum E exceeding 2 GPa is stipulated to mitigate excessive bending when loaded with contents, aligning closely with the properties exhibited by PS¹⁷.

In assessing thermodynamic attributes, a heat capacity (C_p) akin to that of PS is desired, emphasizing the necessity for thermal insulation to prevent burns from hot contents. While considerations such as thermal conductivity are desirable, rapid and accurate models to predict this

property for polymers are not available. Unlike other properties where exceeding a threshold is sought, the ΔH is constrained within a narrow range of -10 to -20 kJ/mol. It's noteworthy that the true quantity of interest that determines the polymerization/depolymerization equilibrium is the T_c , which is defined as the ratio of ΔH to the entropy of polymerization (ΔS) when the monomer concentration is one. However, predicting ΔS or T_c remains challenging due to the limited availability of large datasets, their strong dependence on environmental factors (such as solvents, initial monomer concentration and other experimental variables) and the lack of established computational methods, such as density functional theory (DFT) or molecular dynamics (MD), to accurately predict these properties over a range of realistic conditions adopted in the physical lab. Consequently, we focus on ΔH , leveraging its proportional relationship to T_c to infer that excessively negative ΔH values may hinder depolymerization, while negative values close to zero may impede polymerization.

We limited our analysis to these six properties due to the complexity of designing a new material that meets multiple property requirements. Incorporating additional properties like transparency and solubility, which are crucial for polymer processing and certain applications, would further restrict the search space, potentially resulting in no suitable or synthesizable options. Ultimately, we believe that virtual or informatics-based approaches are primarily intended to alert us to hidden opportunities and point us in the right direction to accelerate design progress. This consideration guided our decision to focus on a manageable set of key properties.

For each of these properties, we developed rapid and accurate machine learning (ML) models to predict them from a polymer's molecular structure, as displayed in the informatics workflow's second box in Fig. 1c^{18,19}. These models offer a significant advantage over physics-based simulation techniques such as DFT and classical MD due to their remarkable speed, enabling them to evaluate the millions of polymers in our search space efficiently.

Notably, the modeling of polymer recyclability remains an emerging field. As such, our exploration, illustrated by the sample Uniform Manifold Approximation and Projection in the third box of Fig. 1c, concentrated on the search space of ROP polymers. We chose to narrow our focus to ROP polymers because of their demonstrated potential to meet the necessary thermodynamic criteria for facilitating chemical recycling to monomers^{20,21}. Moreover, ROP polymers have garnered significant attention in the pharmaceutical industry owing to their customizable properties, biocompatibility, and biodegradability²²⁻²⁴. Our ΔH model is specifically trained on a dataset tailored to ROP, thereby improving its accuracy in predicting outcomes related to ROP polymers^{19,25}.

Subsequently, after a thorough exploration of the molecular space to find candidates that could undergo ROP, we screened over seven million hypothetical polymer designs (encompassing 9 polymer classes) to identify promising candidates that met our stringent screening criteria. The screened candidates are then recommended for further in-depth lab-based studies, as depicted in the final box of Fig. 1d. Given the vastness of this search space, traditional Edisonian trial-and-error methods would prove impractical. While experienced chemists are capable of dramatically narrowing this search space through literature reviews and chemistry first principles, an informatics-based approach offers several additional benefits, including accelerated discovery, reduced reliance on expert knowledge, and the ability to systematically explore complex chemical spaces. Consequently, there is a pressing need for computational techniques capable of swiftly predicting polymer properties and identifying promising candidates, as has been done here. In the subsequent sections, we delineate our polymer design process and spotlight designs deemed most promising for environmentally friendly alternatives to PS. We also highlight three experimentally validated designs, two of which have been previously published by others, and a third, a novel thiocane polymer, which we have synthesized, tested, and report on here for the first time.

Results

Virtual forward synthesis (VFS) reactions

As stated previously, in the pursuit of our design goals, we narrowed our focus to ROP polymers, because of their potential to meet the requisite thermodynamic criteria for facilitating chemical recycling to monomers^{20,21}. The polymer should undergo depolymerization into its monomer ring form above its T_c , while remaining stable in polymer form below it.

To initiate the search for viable polymer candidates, we employ a VFS approach. This method involves the systematic generation of hypothetical polymers from a database of initial monomers, following established reaction pathways. Although the technique dates to the early 2000s²⁶, recent advancements in this area, exemplified by initiatives such as the Open Macromolecular Genome, SMiPoly, and polyVERSE²⁷⁻²⁹, have underscored the potential of this methodology when integrated with ML. Unlike these predecessors, our approach places a significant emphasis on database design, as detailed in the "Database and molecule data" section, and supports the integration of multi-step reaction pathways.

Figure 2 depicts a high-level flowchart of the VFS procedure. Our technique interfaces heavily with a database, as can be indicated by the database icons within the figure, indicating when data is stored. Our workflow begins with a computational chemist creating a virtual reaction using reaction smarts from known reaction pathways.

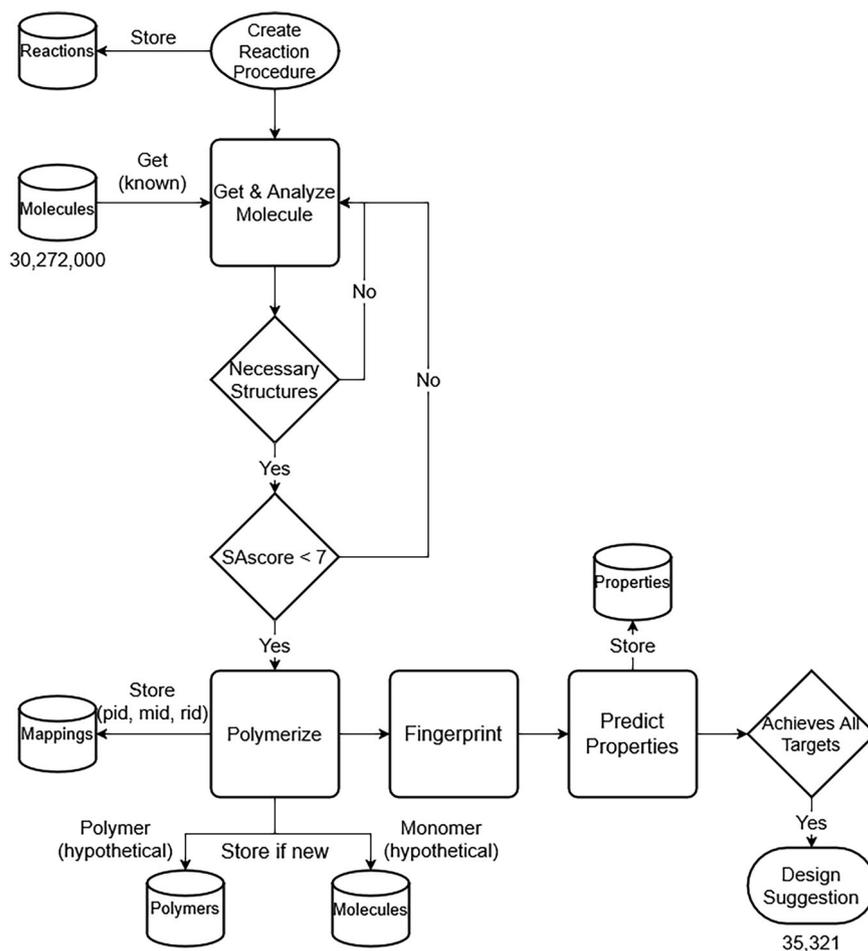
Within our database, a structured query language (SQL) table houses the simplified molecular-input line-entry system (SMILES) of both known and hypothetical molecules. Known molecules were gathered from multiple sources, including ZINC15, ChemBL, compounds sourced from literature, eMolecules, and VWR³⁰⁻³³. During each VFS reaction procedure, we search this repository for known molecules possessing the required substructures via SMILES arbitrary target specification (SMARTS) substructure queries³⁴. Subsequently, to ensure synthetically tractable molecules, we apply a stringent filter, excluding those with a synthetic accessibility score (SAscore) above 7, which is substantially higher than the average score of approximately 3 commonly observed for synthetic molecules.

We selected the SAscore as an initial screening metric due to its well-established validation on a large corpus of synthesized molecules, high computational efficiency, and seamless integration with RDKit. This score is calculated by incorporating penalties for atom count, chiral center count, and ring complexity (including spiro and bridge atoms). Notably, the RDKit implementation also accounts for molecular symmetry, rewarding highly symmetrical structures with lower scores, which are indicative of enhanced synthesizability. Although our focus on ringed monomers was expected to yield higher scores on average, due to the inherent penalties associated with ringed motifs, we deemed this metric suitable for preliminary filtration due to its efficiency advantages over more computationally expensive fingerprinting techniques, which can be prohibitively slow for complex molecules (<https://www.rdkit.org/>)^{35,36}.

Following this filtration process, we subject the retained molecules to the reaction procedure to generate virtual polymers (denoted by the polymerize process in the flowchart). Additionally, for multi-step reactions, we generate virtual monomers, essential for ΔH predictions. Subsequently, both monomers and polymers are converted to SMILES format and subjected to canonicalization, ensuring that molecules with the same structure output the same string³⁷. For example, polyethylene can be encoded as "[*]CC[*]" and "[*]C[*]" and canonicalization will ensure these are both recorded as "[*]C[*]". This canonicalization step is crucial for database querying efficiency, as it enables us to perform string comparisons on the canonical_smiles column in $O(\log(n))$ search complexity due to B-Tree indexing. This allows us to assess if the polymer is "known" already or "hypothetical". While RDKit's CanonSmiles function is used for monomer canonicalization (<https://www.rdkit.org/>), the canonicalize_psmiles package^{18,38} serves this purpose for polymers.

We query the molecules table for the monomer's molecule ID (mid in the flowchart) and the polymers table for the polymer ID (pid in the flowchart). If a monomer or polymer already exists within our database, we extract its ID. Otherwise, we will add it to the database with a unique ID and

Fig. 2 | VFS workflow for polymer design. Flowchart of the VFS design workflow. Arrows indicate the direction of data flow. Cylindrical elements represent database tables accessed during the process. A total of 30,272,000 molecules were analyzed, yielding 35,321 polymer designs.



flag it as a “hypothetical” structure. If the monomer or polymer already exists within the database, it will have been flagged as “known” if it originated from literature reviews, database dumps, or previous experimental characterization, or “hypothetical” if it was generated through our VFS technique and not corroborated by existing literature or experimental data. Upon identification or generation of IDs, a mapping table is used to store the pertinent associations among molecule ID, reaction procedure ID (rid in the flowchart), and polymer ID. This facilitates the swift retrieval of reactants necessary for promising polymer synthesis, alongside the proposed reaction pathway. The identification of promising polymers entails fingerprinting the polymers (and monomers for enthalpy predictions) and then leveraging ML techniques to predict their properties and subsequently evaluating them against predefined screening criteria, as identified in Fig. 1c. The efficiency of virtual polymerization, canonicalization-enhanced databasing, and ML has shifted the bottleneck to fingerprinting, which requires approximately 1000 compute seconds to process 10,000 polymers. We estimate this step is at least an order of magnitude slower than the combined time of the other processes¹⁸.

In Table 1, a simplified version of the ROP reaction is displayed, along with the class of monomer ring being opened, the number of polymers generated from each class, and how many successfully met all target criteria. Our investigation encompasses ring-opening reactions of ethers, thioethers, esters, thioesters, thionoesters, amides, cycloalkenes, carbonates, and thiocanes. For cycloalkenes, we assume the polymerization and depolymerization process would occur through a ring-opening metathesis reaction³⁹. For thioethers, thioesters, and thionoesters, in addition to exploring commercially available options, we also explored hypothetical monomer designs by swapping the appropriate oxygen in an ether or ester with a sulfur. This approach resulted in a

substantially larger explored chemical space than would be expected from a sole search of these structures in the database.

Similarly, for thiocanes, because so few commercially available molecules with the thiocane structure existed, we developed a two-step reaction procedure guided by the expertise of our polymer chemists. This two-step procedure had two variants, R₁ and R₂, as outlined in Table 1. The variants differ in their functionalization step: R₁ involves modifying the vinyl sulfide’s vinyl component with a terminal alkyne-containing molecule, whereas R₂ involves functionalizing the ring’s ketone component with a bromine-containing molecule. Notably, both the alkyne and bromine groups are eliminated during this initial reaction step, leaving behind the R-group attached to the thiocane. The resulting thiocane can be either the R₁ or R₂ variant, but not both. Both variants undergo ring-opening in the second step, yielding a novel polythiocane.

All 30,272,000 molecules in our database were screened using SMARTS to assess if they contained specific substructures relevant to our reaction procedures, as outlined in Table 1. For example, 2,321,545 molecules contained the ether ring structure (SMARTS of “[C; R; !\$(C=O)][O; R][C; R; !\$(C=O)]”). These molecules are eligible for use in both our ether-to-polyether ring-opening reaction procedure and the ether-to-thioether-to-polythioether ring-opening reaction procedure. From the 30,272,000 molecules, we generated a total of 7,301,681 polymers using nine distinct classes of reaction procedures. The full list of necessary substructures searched using SMARTS can be found in the “reaction_procedures.json” located on the GitHub page linked in the “Fitness assessment” section. This file contains detailed single and multi-step reaction procedures, which constitute the structures searched for within each molecule using SMARTS nomenclature, the step at which the monomer is generated (if distinct from the initial molecule), and the reaction SMARTS employed at each step. For

Table 1 | ROP VFS results

Reaction (X = S, O)	Monomer Class	Polymers Generated	Promising Candidates
	Thioether	2,797,024	12,100 (0.43%)
	Ether	2,321,545	8,850 (0.38%)
	Cycloalkene	1,164,917	2,876 (0.25%)
	Thiocane	306,177	126 (0.04%)
	Amide	256,223	6,827 (2.66%)
	Thioester	225,649	1,177 (0.52%)
	Thionoester	130,230	1,777 (1.36%)
	Ester	80,607	1,578 (1.96%)
	Carbonate	821	10 (1.22%)
	Total	7,301,681	35,321 (0.48%)

With the exception of thiocanes, the inclusion of 'R' in the reaction signifies that the ring can be of any size, and the 'R' ring atoms can be any element with any substituents attached. For thiocanes, 'R₁' and 'R₂' involve distinct modifications to the molecule, with 'R₁' adding a terminal alkyne-containing molecule to the vinyl component and 'R₂' functionalizing the ketone component with a bromine-containing molecule, resulting in either the 'R₁' or 'R₂' thiocane variant after the elimination of the alkyne and bromine groups.

additional details on the database schema and VFS, refer to the Supplementary Section "VFS and the Database Schema". For a detailed example of creating reaction procedures and multi-step reactions, see the example presented in Supplementary Fig. S14 and the corresponding lines 251–299 of "reaction_procedures.json", as well as the "Reaction procedures" section.

Predictive models

Polymers generated in this study had their properties predicted using two subsets of previously developed and published models^{18,19}, as described in the "Predictive models for polymer properties: Gaussian process regression and multitask neural network" section: a Gaussian process regression (GPR) model to predict ΔH based on the polymer and monomer¹⁹, and a multitask

neural network (MTNN) trained on homo and copolymer data to predict all other properties¹⁸. Parity plots illustrating the MTNN models' performance on test data of known polymers (red dots) and our dataset of known ROP polymers (black stars) are depicted in Fig. 3. The count and root mean squared error (RMSE) shown represents firstly the test dataset size and model performance on the test data, then on the known ROP polymers. For ΔH , the black stars and labeled RMSE and count indicate the performance on test ROP data and the size of the test dataset, while the colored circles represent the training data.

With the exception of ΔH , the models generally perform worse for known ROP polymers than their known test dataset chemistries. This outcome is unsurprising given the limited variety of ROP chemistries in the

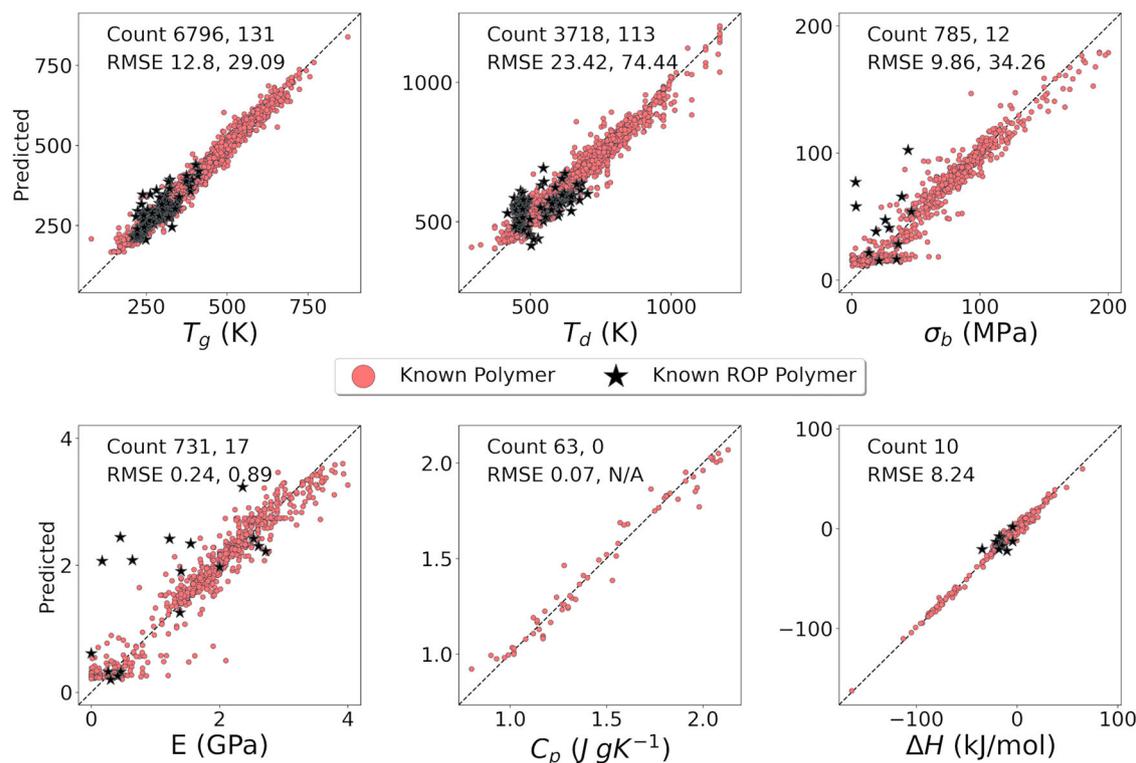


Fig. 3 | Parity plot for property predictions. Parity plot illustrating model performance across different properties. The first set of values in the top left showcase counts and RMSE on the model test data of known polymers, which are represented by red circles. The second set delineates the dataset size and model performance

specifically on known ROP polymers, denoted by black stars. For the enthalpy of polymerization, dots depict training data, while black stars indicate test data, as the model was exclusively trained on ROP polymers.

original training data. Of the 138 unique ROP polymers with thermal property data available, only eight were seen in the training data, and of the twenty with mechanical property data available, none were in the training data. However, predictions for thermal properties tend to align closely with the parity line, indicating acceptable performance. In contrast, for mechanical properties, the models tend to over-predict performance. In addition to the lack of ROP data, this discrepancy could be attributed to the training process, as the models only consider molecular structure, while factors such as molecular weight, which contribute significantly to mechanical properties, are not considered. It is possible that while the ROP molecular structure enhances strength, the molecular weight remains insufficient⁴⁰. Unfortunately, due to the scarcity of reliable datasets that include molecular structure, mechanical property values, and molecular weight, we were unable to train a machine learning model that effectively utilizes molecular weight information. The predicted mechanical properties should thus be viewed as a stochastic average over morphological and molecular weight variations, and should be used as a guide that points us in the right direction.

This observation is supported by two outlier polymers in the σ_b plot. These outliers exhibit predicted values of 58 and 77 MPa, whereas their actual values are 3.4 and 3.1 MPa, respectively. Notably, the molecular weights associated with these polymers are very low at 18 and 10.8 kDa, compared to other ROP polymers in the dataset, which range from 109–126 kDa (44 MPa), 198 kDa (39 MPa), to 266–438 kDa (46.4 MPa)^{21,41,42}.

Similarly, in the case of E , one outlier exhibits a predicted value of 2.07 GPa, while its true value is 0.17 GPa, with a low molecular weight of 18 kDa²¹. However, this theory fails to account for some other inaccuracies of the model, as evidenced by two other polymers with predicted values of 2.44 and 2.08 GPa. Their true values were considerably lower at 0.45 and 0.64 GPa, despite higher molecular weights of 85.6 kDa and 69.3 kDa.^{42,43} These instances suggest that the model has not encountered enough similar chemistry.

Figure 4 shows the histograms representing the distribution of the predicted polymer properties for the entire pool of over seven million hypothetical ROP polymers generated here (green), as well as for the measured properties of presently known ROP polymers (black). Note that the y-axis is on a log scale. Based on Fig. 4, it can be concluded that existing ROP polymers generally have low mechanical and thermal properties. In contrast, our models predict that many hypothetical ROP polymers could potentially surpass these limitations, achieving higher performance values.

Promising polymers

To determine which hypothetical polymers are most promising, we employed a three-step fitness function defined mathematically in the “Fitness assessment” section. This assessment multiplies scaled property predictions, each normalized between zero and one, to compute a polymer’s fitness score. The scaling ensures that properties with different units and scales, such as T_g and C_p , are treated equally, preventing any one property from dominating the fitness score due to its magnitude. We chose a multiplicative approach to emphasize the importance of meeting all target properties, as a single property falling short can significantly impact the overall performance of the polymer. Polymers achieving a perfect fitness score of one met all predefined property requirements. The fitness distribution among the polymers and their classes is illustrated in the stacked bar plot in Fig. 5a.

The multiplicative approach employed in our fitness assessment heavily penalizes polymers that significantly fall short of achieving a single property or moderately miss multiple targets. This enables us to prioritize polymers that most closely match all target properties, even when not all properties are perfectly met. The effect of this approach is reflected in the right-skewed distribution of polymer classes in the left plot of Fig. 5a, characterized by a sharp decline at 0, a gradual decrease between 0.2 and 0.4, and a more pronounced drop-off thereafter. Furthermore, due to the

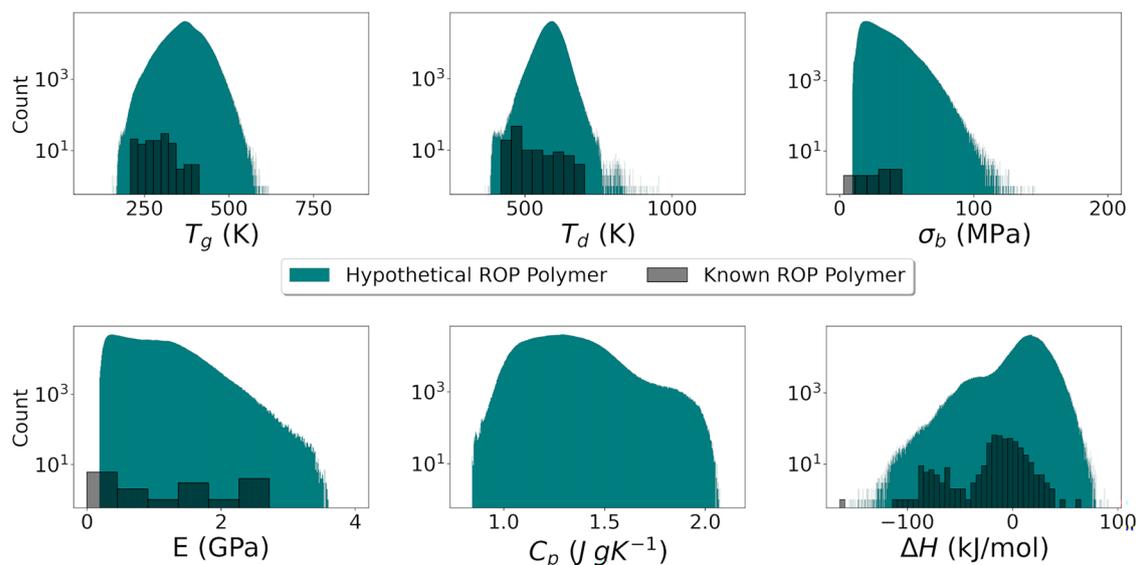
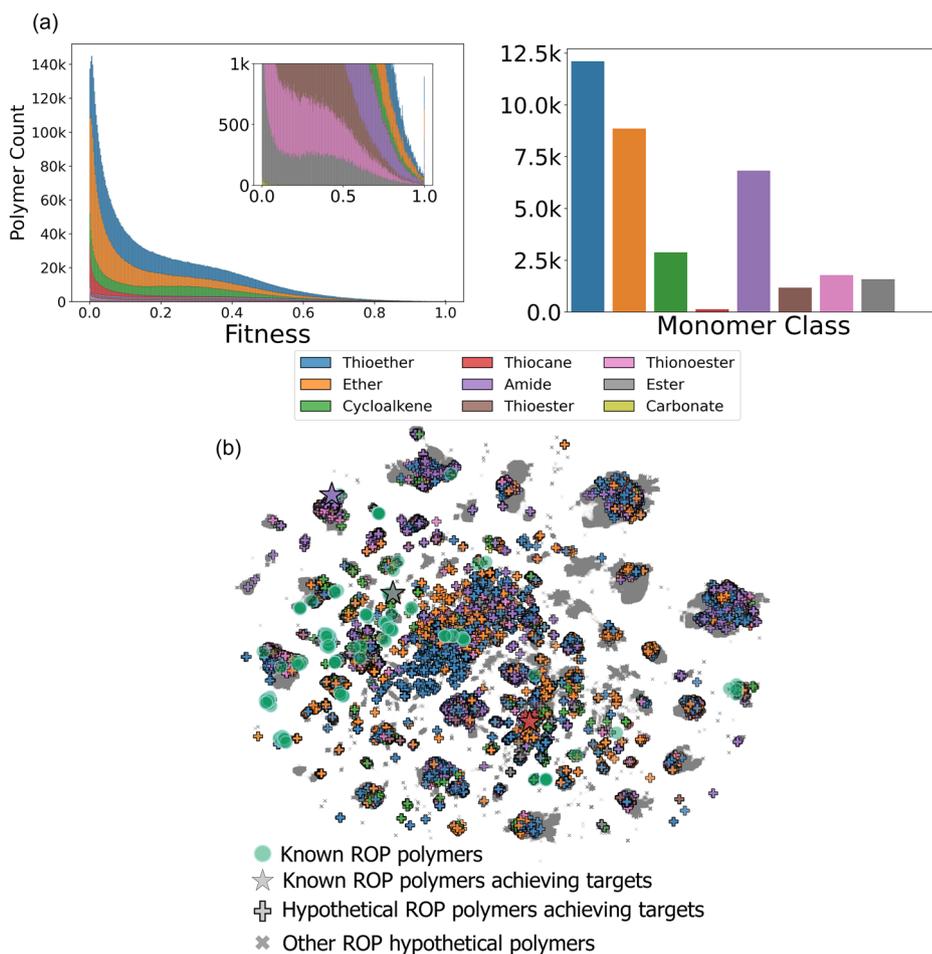


Fig. 4 | Histograms of predicted and measured ROP polymer properties. Histograms illustrating predicted property distributions for hypothetical ROP polymers (green) and measured property distributions for known ROP polymers (black).

Fig. 5 | Polymer fitness and fingerprint embedding. **a** Histogram illustrating the distribution of fitness values among hypothetical polymers (left) and a count plot for polymers achieving fitness score >0.8 for each monomer class (right). **b** Uniform manifold approximation and projection (UMAP) visualization of high-dimensional polymer fingerprints condensed into two dimensions, with class colors matching (a) and (b) to indicate target-achieving polymers. The stars correspond to the bottom three polymers shown in Table 2.



capping of values above the target at 1, a clustering of polymers at the maximum fitness value of 1 is observed.

To account for modeling errors, we established a fitness score threshold of 0.8 to identify promising polymer candidates. Despite generating millions of polymers, only 817 achieved all property targets, while 35,321 met the 0.8

cutoff, underscoring the intricate challenge of designing polymers with multiple desired properties. The successful polymers were predominantly thioethers, ethers, and amides, with smaller subsets of cycloalkenes, thionoesters, thioesters, esters, and a negligible number of thiocanes and carbonates, as can be seen in the right plot of Fig. 5a. However, amides, esters,

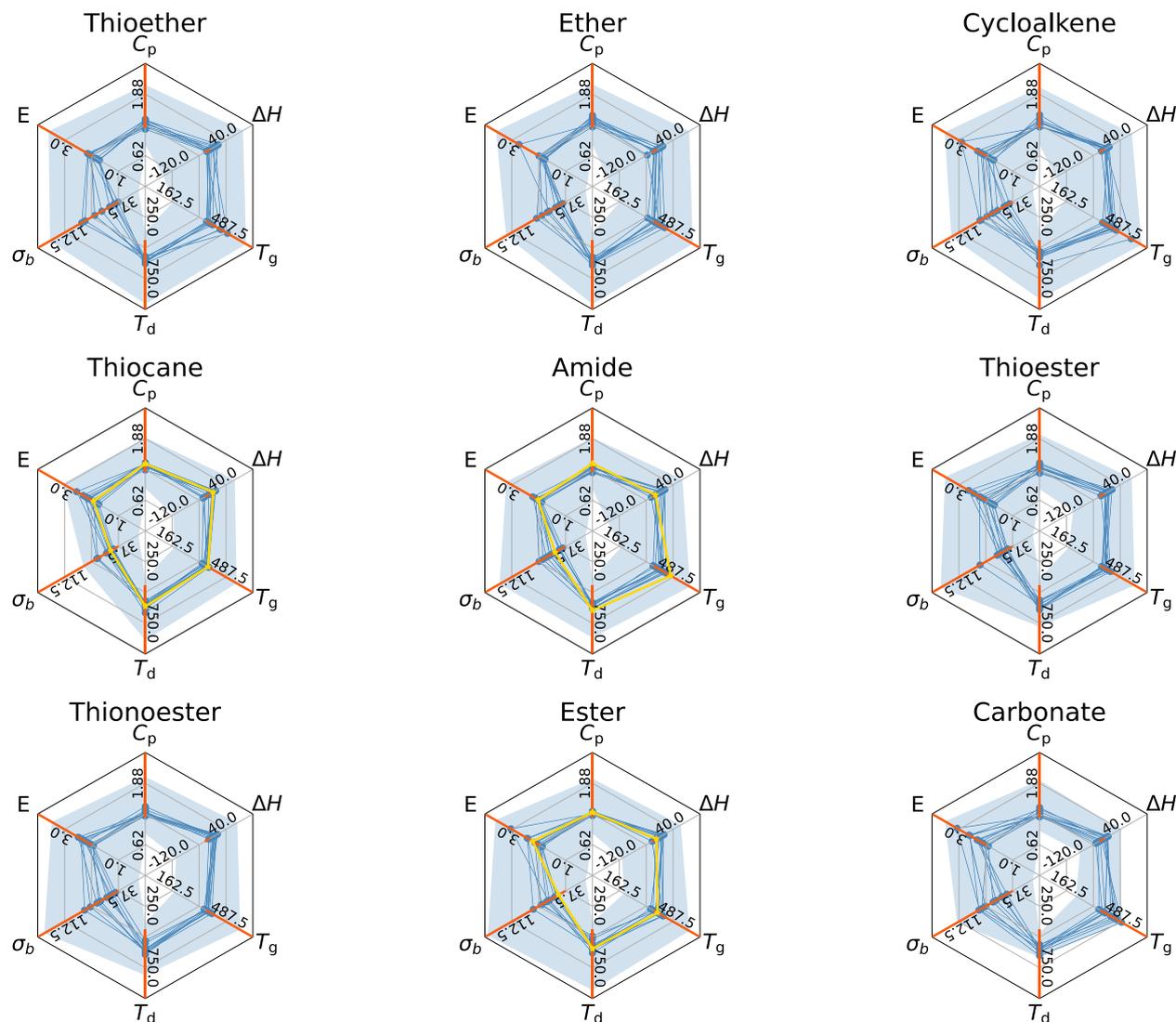


Fig. 6 | Property ranges for polymer classes and values for achieving candidates. Radar plots showcasing ranges for C_p ($\text{J (g}^\circ\text{K)}^{-1}$), ΔH (kJ/mol), T_g (K), T_d (K), σ_b (MPa), and E (GPa) of different polymer classes. Each shaded area delineates the minimum to maximum range of predicted properties within the respective polymer class, while the orange line denotes the targeted design region. Additionally,

individual lines trace ten randomly chosen polymers that successfully meet the fitness objective (>0.8). The gold lines in the 'Amide', 'Ester', and 'Thiocane' charts delineate the predicted properties for the known ROP polymers within these classes that achieved the fitness objective.

thionoesters and carbonates exhibited a higher frequency of meeting the target criteria, relative to their representation in the population, as shown in Table 1.

We visualized the high-dimensional polymer fingerprint, as described in the “Fingerprinting techniques for polymers and molecules” section, using a UMAP (see the “Uniform manifold approximation and projection (UMAP)” section for an overview of the method)⁴⁴. Employing a cosine similarity metric, we configured the UMAP with 200 nearest neighbors and a minimum distance setting of 0.25, optimizing for a balance between local and global manifold structures and ensuring adequate spacing of data points⁴⁵. The result is shown in Fig. 5b. Known ROP polymers are denoted by green circles, while known ROP polymers achieving the desired fitness score, discussed in the “Synthetic validation & design guidelines” section, are marked with colored stars. Hypothetical ROP polymers achieving the fitness score are represented by colored plus symbols, contrasting with the gray x's representing other hypothetical ROP polymers. The color scheme for stars and plus symbols in (b) corresponds to the class colors in (a), facilitating intuitive comparison.

The UMAP visualization unveils discernible clusters, particularly noticeable on the upper right side, where hypothetical polymers exhibit substantial deviations from current ROP chemistries, hinting at uncharted realms with untapped potential. However, given that these chemistries are novel to the model, the extrapolation may not be accurate. Therefore, we recommend validating these predictions using more rigorous computational models, such as DFT, before pursuing experimental synthesis. On the other hand, there are some promising polymers situated near established ROP domains, particularly in the left portion of the map. This proximity suggests the existence of novel candidate polymers that may be able to effectively leverage known ROP reaction catalysts, solvents, and temperatures.

To comprehensively examine the distinct properties of the polymer classes, we devised radar plots for each category, as depicted in Fig. 6. Within each plot, the shaded region delineates the minimum to maximum range of predicted properties for the polymer class and the solid lines represent the properties for ten randomly chosen polymers achieving the fitness threshold within the class. Additionally, the predicted properties for known ROP polymer within a class are colored gold.

Table 2 | Predicted and measured properties of promising ROP polymer candidates

ID	Class	T_g (K)	T_d (K)	E (GPa)	σ_b (MPa)	ΔH (kJ/mol)	C_p (J(gK) ⁻¹)
1	Amide	397	612	2.21	80.6	-18.45	1.42
2	Ether	381	580	2.71	54.72	-10.26	1.62
3	Ether	386	558	2.07	41.66	-15.31	1.66
4	Ether	394	511	2.68	57.84	-16.22	1.48
5	Ether	384	568	2.44	51.17	-16.45	1.6
6	Amide	479	725	2.24	80.06	-10.15	1.34
7	Ether	392	530	2.27	50.06	-16.14	1.59
8	Thioether	470	689	2.14	77.52	-13.55	1.26
9	Ether	385	525	2.73	57.99	-13.91	1.41
10	Ether	419	622	2.47	60.76	-11.78	1.51
11	Ether	434	623	2.42	63.63	-14.72	1.42
12	Ether	393	619	2.27	48.17	-16.45	1.43
13	Ester (Known)	393 (322)	595 (613)	2.22 (2.7)	47.24 (54.7)	-19.63 (-20)	1.28
14	Amide (Known)	473 (419–472)	646 (630)	2.02 (2.28)	52.15	-12.48 (-10)	1.37
15	Thiocane (Known)	379 (352)	613 (563)	1.93	48.1	1.92	1.37

Polymers 1–12 represent hypothetical polymers, while polymers 13–15 have been synthesized and tested^{55,56}. The selection process for the hypothetical polymers was based on the length of the SMILES, serving as a heuristic for polymer complexity, with the 12 shortest strings presented here. For synthesized polymers, the measured properties are shown in parentheses below the predicted values. The complete list of hypothetical polymers, including their corresponding fitness values, is available in our data repository, referenced in the “Fitness assessment” section. Images corresponding to these polymers are displayed in Fig. 7.

The plot reveals that the majority of polymers hover around or slightly exceed the target thresholds for each property. Notably, some polymers excel in specific properties like E or T_g , while barely meeting the other targets. Conversely, a subset of polymers falls short of the target region for certain properties, particularly ΔH . However, these deviations are marginal, and therefore, the fitness function imposes only minimal penalties.

Table 2 showcases a selected group of hypothetical polymers that achieved all targeted properties, along with their predicted properties. The corresponding reactant monomers and polymer repeat units are illustrated in Fig. 7. A common structural motif among all promising polymers is the presence of cyclic elements within their backbones or substituents, which is known to enhance polymer rigidity and consequently improve thermal stability, stiffness, and strength^{46,47}. In our dataset of known polymers, we see a statistically significant upward shift in the histograms of T_g , E , and σ_b for polymers with either aliphatic or aromatic rings. Moreover, there is a notable further shift toward higher values when these rings are incorporated within the backbone, as illustrated in Supplementary Fig. S1. For the hypothetical polymers, those with the ring in the backbone also tended to have higher T_g .

Among the 35,321 selected polymers, only 416 lacked any ring structures entirely. Notably, 323 of these polymers featured alcohol, (thio)imide, or amide moieties with no or short substituents attached to the polymer backbone. We hypothesize that the observed enhanced performance of these interesting cases may stem from the potential for hydrogen bonding within these structures. Indeed, hydrogen bonding has been recognized to augment mechanical properties by promoting inter-chain interactions that enhance polymer rigidity^{48–52}, so long as intrachain bonding that can increase chain flexibility is avoided⁵⁰. Moreover, substituents and side chains are known to influence polymer properties significantly. Increasing their size typically leads to greater free volume and reduced packing density, factors known to lower both the T_g and mechanical strength^{33,54}.

Mirroring the stiffening trend observed in ring-containing polymers, we once again noted a rightward shift in the histograms depicting the

property distributions of known polymers containing nitrogen, as shown in Supplementary Fig. S2. While the presence of amines and amides seemed to have little impact on mechanical properties, thermal properties saw a small effect. Furthermore, both categories of properties exhibited a statistically significant rightward shift when featuring an imide structure. In general, we observed that an increase in the number of hydrogen-bond acceptor atoms in the chain corresponded to an increase in both properties, as seen in Fig. S2c. Moreover, a distinct positive correlation between T_g and σ_b was discerned in our experimental dataset, as illustrated in Fig. S3. This is likely because both properties are influenced by chain stiffness.

The downside to functional groups that induce hydrogen bonding, however, is their tendency to be inherently reactive. Chemical intuition dictates caution against the inclusion of amines, hydroxyls, carboxylic acids, and acidic methylene groups in the monomer structures. While these groups have the potential to enhance thermal and mechanical properties through hydrogen bonding, their high reactivity could lead to undesirable side reactions during synthesis, ultimately diminishing the likelihood of successful ring-opening polymerization. Filtering out molecules containing these functional groups yields a list of 6477 promising polymer candidates.

Synthetic validation & design guidelines

Our design process identified two previously known polymers (IDs 13 & 14) as potential candidates to replace PS (these were not part of the training set used to create our non-enthalpy property prediction ML models). Significantly, both featured rings in their backbone structures, with one additionally incorporating an amide moiety. Polymer 13 was synthesized from a γ -butyrolactone derivative and successfully met all target properties except for the C_p , which was not determined, and the T_g , which we predicted to be 393 K but was measured at 322 K. The deviation in T_g could potentially be attributed to the polymer’s low crystallization rate or variations in stereochemistry that are not adequately represented by the models⁵⁵. The other values listed were a ΔH of -20 kJ/mol, a T_d of 613 K, a E of 2.7 GPa, and a σ_b of 54.7 MPa, all slightly higher than our model predictions, but close.

For polymer 14, experimental measurements revealed a T_d of 630 K and a T_g spanning from 419 K to 472 K. Impressively, the polymer demonstrated outstanding chemical recycling capabilities, boasting a remarkable mass recovery rate of 93–98%. Moreover, upon copolymerization with nylon 4, it achieved an impressive E value of 2.28 GPa. Although literature lacks information regarding its C_p or σ_b , the other properties were consistent with both our predictions and design targets⁵⁶.

The discovery of these polymers marks a compelling initial validation of our informatics-based approach. Moreover, they exhibit chemical motifs akin to other polymers flagged by our models. This promising discovery bodes well for unveiling more mechanically and thermally resilient, chemically recyclable polymers (poised to serve as ideal substitutes for PS) within our catalog.

In addition to the above “designed” candidates previously identified in the literature, we endeavored to synthesize the other promising, entirely new candidates. Focusing on the thiocane class, we selected one functionalized with an alkyne-containing, comparatively low-cost dimethyl fluorene, polymer 15 in Table 2. In selecting this candidate, we utilized polymerization efficacy heuristics; dimethyl fluorene lacked additional functional groups that could potentially disrupt the polymerization process. The polymer was synthesized according to the procedure outlined in the Supplementary Section “Thiocane Synthesis” and our recent publication⁵⁷. Excitingly, the T_d of the synthesized polymer met our target, and while the T_g did not fall precisely within our desired range, it reached 352 K—only 27 K lower than the predicted value of 379 K. This discrepancy between predicted and measured values may be attributed to the polymer’s relatively low molecular weight (6 kDa), suggesting potential improvement with higher molecular weights based on the Flory-Fox equation.

Unfortunately, we encountered difficulties during the polymer synthesis. Low yield in monomer preparation posed an initial challenge, limiting our ability to produce sufficient quantities of monomer for polymerization. Additionally, we were unable to exceed a molecular weight of

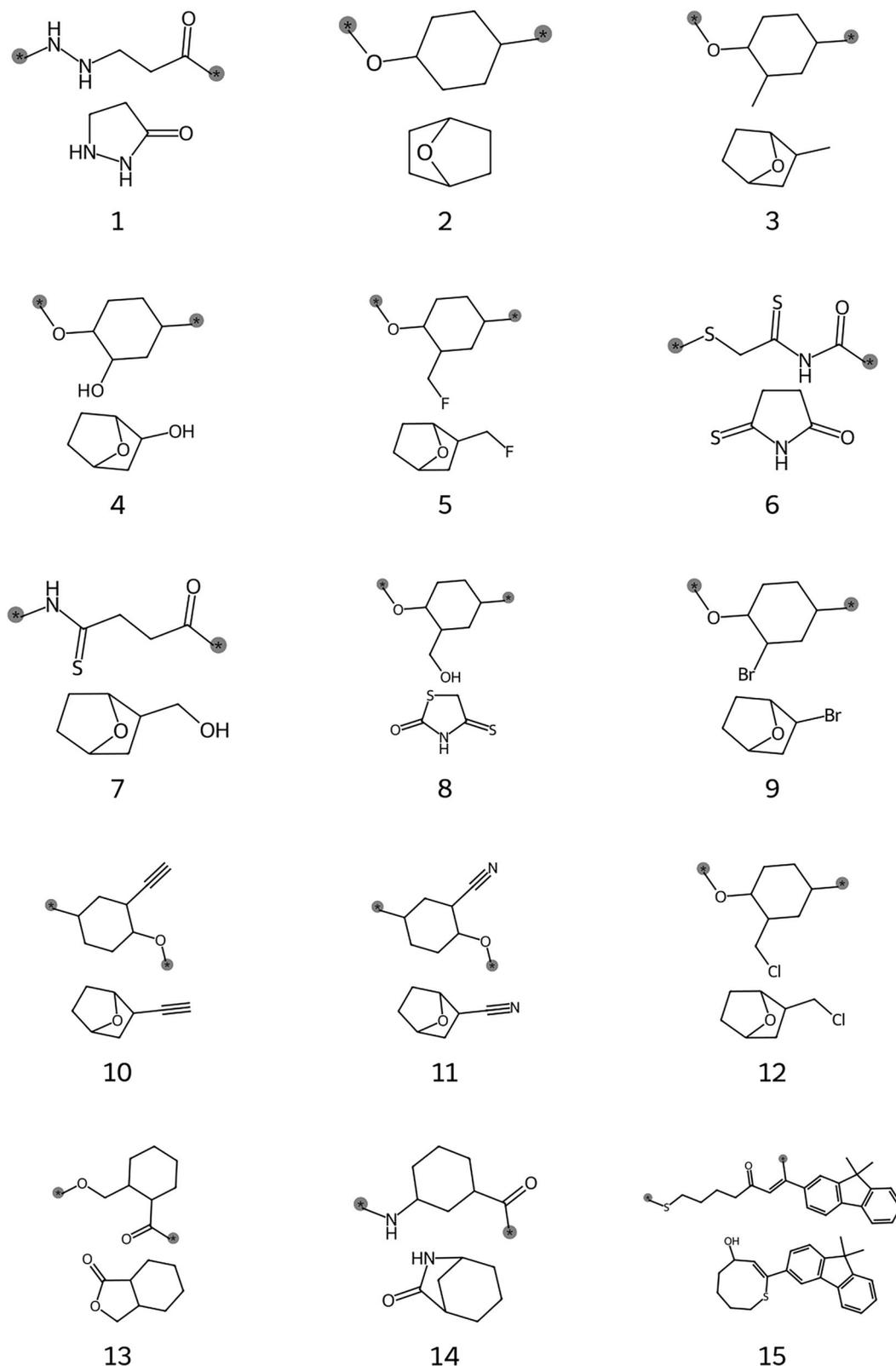


Fig. 7 | Promising polymer structures. Polymer (top) and monomer (bottom) structures for promising candidates, labeled with corresponding IDs from Table 2. Polymer bonding locations are indicated by gray-highlighted stars.

6 kDa as a result of solubility problems, which hindered measurements of mechanical properties. Despite numerous adjustments to the polymerization solvent, these challenges persisted. Nonetheless, this experience serves as robust validation of the VFS technique and underscores the inherent

complexity of translating polymer “design” to polymer synthesis⁵⁷. Among thousands of designs, only a select few are predicted to achieve the requisite property targets, and fewer still prove viable for synthesis due to potential interference from certain functional groups in the reaction pathway. Even

when a viable candidate shows initial promise, achieving high monomer yields and large polymer molecular weights remains a formidable task. Thus, although employing informatics approaches may not lead to an immediate synthesizable hit, it can be instrumental in prioritizing efforts.

A comprehensive catalog of the promising polymers, along with their predicted properties, precursor materials, and suggested synthesis routes, is provided in Supplementary Materials. Typically, the most promising candidates exhibit the following characteristics:

1. Preference for the heterocycle being opened to have a size of 4, 5, 6, or 7. These comprised 6%, 46%, 35%, and 11% of the down-selected polymers, respectively.
2. Presence of rings within the backbone structure, observed in 64% of down-selected polymers.
3. Inclusion of rings somewhere in the polymer structure (either the backbone or side groups), with 99.3% of down-selected polymers meeting this criterion.
4. Absence of side chains or presence of short substituents or those containing bulky rings.
5. Ability to hydrogen bond (however, care is needed to ensure the bonding group does not interfere with the polymerization chemistry).

Discussion

In this study, informatics advancements were employed to generate over seven million hypothetical ROP polymer candidates from a dataset comprising 30,272,000 known and commercially available molecules, with the aim of identifying recyclable alternatives to conventional plastics that are both thermally and mechanically durable. Our methodology yielded over 35,000 promising recyclable candidates demonstrating predicted mechanical and thermal durability close to PS. With input from polymer chemists, we further narrowed down this number to 6477. Furthermore, we identified two known chemically recyclable polymers, not initially included in our thermal and mechanical property training datasets, yet meeting the criteria for PS substitutes alongside our hypothetical polymers, serving as validation for our informatics approach. Additionally, we synthesized a novel thiocane polymer designed using our methodology with enhanced thermal properties. This achievement represents a significant milestone, as our methodology not only designs novel polymers but also provides clear reaction pathways, enabling rapid experimental synthesis - a rarity in the field of generative polymer design^{7,58,59}. Investigation of all promising candidates revealed that suitable substitutes for PS containers shared similar chemical characteristics, notably involving the opening of heterocycles ranging from 4 to 7 atoms, the presence of a ring within the polymer backbone, the absence of side chains, or the incorporation of bulky constituents within them. Additionally, numerous polymers contained atoms conducive to hydrogen bonding.

While VFS shows promise in generating synthetically feasible polymer candidates, only a minute fraction (<0.5%) of the polymers produced meet all desired properties. This low success rate highlights the need for accelerated discovery methods, particularly when expanding to copolymers for recyclable designs or complex reactions with multiple reactants^{21,60-64}. In such cases, the search space grows exponentially with the number of molecules involved. For instance, a reaction with two reactants (n and m options, respectively) would have $n \times m$ possible combinations. To efficiently navigate this vast space, we propose leveraging GAs in conjunction with VFS to identify promising reaction procedures and reactants and prioritize their exploration, thereby streamlining the discovery process^{58,59,65}.

The reliance on ML models warrants caution, however, especially during extrapolation to unseen chemical spaces. Enhancements in these models, particularly by using multi-fidelity methods that combine lower fidelity simulation data with experimental observations, will be crucial to extrapolating to unexplored chemical regions^{19,66-68}. Regardless, informatics approaches, when executed synergistically and iteratively with physical experiments, have the potential to rapidly accelerate the discovery of novel and beneficial materials.

Despite the promise of informatics-discovered, recyclable polymer alternatives, significant economic hurdles must be overcome to facilitate their widespread adoption, particularly in replacing commodity plastics. A notable example in this work is the substantial cost disparity between the dimethyl fluorene functionalizing component (\$39.6 per gram) used in our synthesized polythiocane and polystyrene (approximately \$1 per gram). Notably, this was one of the cheaper competitive options available to us, as can be seen in Supplementary Fig. S4. This cost disparity renders it non-competitive with polystyrene. Bridging this economic gap will require future research efforts to focus on identifying more economically viable components, developing process optimizations, scaling up manufacturing, and implementing other cost-reduction strategies to enhance the financial competitiveness of proposed recyclable polymer alternatives.

Methods

Database and molecule data

A PostgreSQL version 12.17 on an Ubuntu 20.04.6 LTS system was used to store all molecule, reaction, and polymer data. The full details of the database schema can be found in the Supplementary Section “Database Schema”.

The molecule section of the database features a “category” column that delineates between known and hypothetical molecules. Known molecules, crucial for generating synthetically accessible polymers, are drawn from five distinct subsets: ZINC15, ChemBL, compounds sourced from literature, an eMolecules database dump from December 19th, 2020, and data from a VWR database, harvested through a tailored webscraper (details available in Supplementary Section “Webscraper”)³⁰⁻³³. Molecules from the ZINC15 and ChemBL datasets are reportedly “commercially available,” although in reality, they can be challenging to procure. Those sourced from literature theoretically can be synthesized based on the referenced procedures. Molecules in the eMolecules database are likely available commercially, though without assurance. However, each eMolecules ID is cataloged, facilitating easy cross-referencing from the eMolecules site. Entries from the VWR database, covering the years 2023–2024, are also presumed to be commercially accessible.

Reaction procedures

The Supplementary File “reaction_procedures.json” (refer to the “Fitness assessment” section for availability) presents a comprehensive inventory of reaction procedures in a structured JavaScript object notation (JSON) format. Each entry encapsulates the class of the monomer being manipulated alongside the corresponding reaction procedures. These procedures comprise a sequence of reactions, with each step delineated by its reaction SMARTS and a succinct description of its intended outcome. Additionally, the requisite reactant SMILES are delineated, along with specifications regarding which reactants must be queried from the database. These reactants are characterized by lists of essential and prohibited substructures, each defined by their respective SMARTS notation. The presence of necessary substructures within the reactants is restricted to be between a minimum and maximum value, while no such restriction is provided for unacceptable substructures. For a reactant to be deemed suitable, it must possess all requisite substructures within the provided range while excluding any flagged as unacceptable.

Predictive models for polymer properties: Gaussian process regression and multitask neural network

The prediction of polymer properties outlined in Fig. 1 relied on two distinct types of machine learning (ML) models. The first type employs GPR to predict the ΔH for ROP polymers. The second type uses MTNNs to predict the remaining polymer properties.

The GPR model was trained using a dataset consisting of experimental ΔH values from ROP polymers, supplemented with DFT data. This hybrid dataset effectively predicts experimental ΔH values for ROP polymers²⁵. The

choice of GPR was motivated by its proven efficacy in accurately predicting polymer properties, particularly with small datasets.

The model employs a Matern kernel, selected via leave-one-out cross-validation (LOOCV) among Rational Quadratic, RBF, and Matern kernels. A white noise kernel was added to account for inherent data noise, with the noise level determined by the standard deviation of the training data. The alpha parameter, which ensures a positive definite matrix and accounts for measurement noise, was tuned using LOOCV¹⁹.

Multiple MTNN models were trained using a diverse array of homopolymer and copolymer property data. These models were trained based on distinct classes of correlated properties, including thermal, mechanical, gas permeability, thermodynamic & physical, electronic, and optical & dielectric properties. The focus of this study, depicted in Fig. 1a, primarily involved thermal, mechanical, and thermodynamic properties. For comprehensive details on the training and testing methodologies, refer to the original MTNN paper¹⁸.

Fingerprinting techniques for polymers and molecules

Both the GPR model and MTNN models rely on the molecular features of polymers for predictions, with the GPR model also requiring monomer molecular features.

Polymer fingerprinting entails three hierarchical levels of descriptors. The initial level quantifies atomic triplets (e.g., H1-C4-H1, denoting two one-fold coordinated hydrogens and a four-fold coordinated carbon). The subsequent level encapsulates predefined chemical building blocks (e.g., -C6H4-, -CH2-, -C(=O)-). The third level encompasses Quantitative Structure-Property Relationship (QSPR) descriptors, incorporating molecular features like molecular quantum numbers and molecular connectivity chi indices, alongside other descriptors such as non-hydrogen atom count and molecular weight. These features are then normalized by the number of atoms in the polymer. Additional details can be found at^{69,70}.

For copolymers, each feature is derived through a linear combination of homopolymer features, with weights determined by the fraction of the respective homopolymer in the copolymer^{18,60}.

Molecule fingerprinting involves the same three hierarchical levels of descriptors as polymer fingerprinting. However, unlike polymers, these descriptors are not normalized based on the number of atoms. Additionally, certain descriptors, like the length of the longest side chain, are disregarded. Crucially, ROP-specific features such as the size of the ring being opened and the valence electron differences in the broken bond of the ring were integrated into the fingerprinting process^{19,71}. This resulted in approximately 200 chemical descriptors, which were then down-selected using feature permutation for the GPR enthalpy model¹⁹.

Uniform manifold approximation and projection (UMAP)

UMAP is a widely employed technique for reducing high-dimensional fingerprints to a lower-dimensional space, referred to as an embedding. Its methodology is predicated on the premise that high-dimensional data can be conceptualized as a Riemannian manifold. By undertaking the task of mapping this manifold to a lower-dimensional embedding while conserving the relative distances between data points, UMAP endeavors to adequately represent the data in fewer dimensions. This approach ensures that the structure of the original data is preserved as faithfully as possible in the reduced space, facilitating efficient visual analysis and interpretation.

The embeddings generated by UMAP are stochastic and influenced by key hyperparameters: 'n_neighbors', 'min_dist', and 'metric'. The 'n_neighbors' parameter determines the number of neighboring data points considered by the algorithm when constructing a topological representation of the data, affecting the preservation of global versus local structure. A larger 'n_neighbors' value emphasizes global topology preservation, while a smaller value emphasizes local structure. On the other hand, 'min_dist' regulates the minimum distance between points

in the embedding; a higher value prevents points from clustering too closely together for effective visualization, while a lower value can result in more tightly packed points. Lastly, the 'metric' parameter specifies the distance metric used in the high-dimensional space to compute distances between data points, which influences how neighboring data points are identified. UMAP was employed to visualize polymer features in two dimensions, facilitating analysis of polymer similarity graphically⁴⁴.

Fitness assessment

The evaluation of hypothetical polymers against the screening criteria depicted in Fig. 1b, c entailed a three-step procedure:

1. Enthalpy transformation: The target range for ΔH was set between -10 kJ/mol and -20 kJ/mol. To ensure uniform weighting between values less than -20 kJ/mol and greater than -10 kJ/mol, ΔH values were adjusted using Eq. (1):

$$\Delta H_i^t = \begin{cases} \Delta H_i + 30, & \text{if } \Delta H_i < -20 \\ 10, & \text{if } -20 \geq \Delta H_i \geq -10 \\ \Delta H_i * -1, & \text{if } \Delta H_i > -10 \end{cases} \quad (1)$$

Here, ΔH_i^t stands for the transformed ΔH value for polymer i , and ΔH_i stands for the predicted ΔH value for polymer i .

2. Clipping of predicted properties: Predicted properties exceeding the target thresholds were clipped to the targets. This process is mathematically represented by Eq. (2):

$$k'_i = \min(k_i, k_{target}) \quad (2)$$

In the equation, k_i represents the predicted value of property k for polymer i , where k can be any of the following: T_g , T_d , σ_b , E , C_p , or ΔH . k_{target} denotes the minimum threshold that a property must meet, as defined in Fig. 1b, c. k'_i signifies the clipped predicted value of polymer i : it remains unchanged if the value is below the target, or becomes the target if it exceeds it. This calculation prioritizes polymers that satisfy all criteria over those that excel in only a few aspects.

3. Normalization and fitness calculation: The adjusted property values were normalized within the range of 0–1 using a Min-MaxScaler. A composite fitness value for each polymer was then calculated by multiplying these normalized properties, as described in Eq. (3):

$$\theta_i = \prod_{k=T_g, T_d, E, \sigma_b, C_p, \Delta H^t} \frac{k'_i - k'_{min}}{k_{target} - k'_{min}} \quad (3)$$

Here, k'_{min} represents the minimum clipped predicted value for all polymers in the dataset. θ_i represents the fitness score for polymer i .

Data availability

All relevant resources, including polymer designs, detailed reaction procedures, and accompanying code for generating polymers from monomers and reaction procedures, are publicly available in our [polyVERSE repository on GitHub](#).

Received: 4 February 2025; Accepted: 2 June 2025;

Published online: 15 June 2025

References

1. Ozbay, G., Jones, M., Gadde, M., Isah, S. & Attarwala, T. Design and operation of effective landfills with minimal effects on the environment and human health. *J. Environ. Public Health* **2021**, 1–13 (2021).

2. Shahul Hamid, F. et al. Worldwide distribution and abundance of microplastic: how dire is the situation? *Waste Manag. Res.* **36**, 873–897 (2018).
3. Barboza, L. G. A. Marine microplastic debris: an emerging issue for food security, food safety and human health. *Mar. Pollut. Bull.* **13**, 336–348 (2018).
4. Campanale, C., Massarelli, C., Savino, I., Locaputo, V. & Uricchio, V. F. A. Detailed review study on potential effects of microplastics and additives of concern on human health. *Int. J. Environ. Res. Public Health* **17**, 1212 (2020).
5. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **3**, e1700782 (2017).
6. Schyns, Z. O. G. & Shaver, M. P. Mechanical recycling of packaging plastics: a review. *Macromol. Rapid Commun.* **42**, 2000415 (2021).
7. Kuenneth, C. et al. Bioplastic design using multitask deep neural networks. *Commun. Mater.* **3**, 96 (2022).
8. Frequent questions regarding EPA's facts and figures about materials, waste and recycling <https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/frequent-questions-regarding-epas-facts-and> (2018).
9. Foundation, E. M. *The New Plastics Economy: Rethinking the Future of Plastics & Catalysing Action*. Tech. Rep. <https://www.ellenmacarthurfoundation.org/publications/the-new-plastics-economy-rethinking-the-future-of-plastics-catalysing-action> (Ellen MacArthur Foundation, 2017).
10. Miao, Y., von Jouanne, A. & Yokochi, A. Current technologies in depolymerization process and the road ahead. *Polymers* **13**, 449 (2021).
11. Foundation, E. M. *The Global Commitment 2021 Progress Report*. Tech. Rep. <https://www.unep.org/resources/report/global-commitment-2021-progress-report> (Ellen MacArthur Foundation, 2021).
12. U.S. Department of Health and Human Services. Report on Carcinogens. 12th ed., Public Health Service, National Toxicology Program, p. 6 (2011).
13. Nugnes, R., Lavorgna, M., Orlo, E., Russo, C. & Isidori, M. Toxic impact of polystyrene microplastic particles in freshwater organisms. *Chemosphere* **299**, 134373 (2022).
14. Hwang, J. et al. Potential toxicity of polystyrene microplastic particles. *Sci. Rep.* **10**, 7391 (2020).
15. Fredi, G. & Dorigato, A. Recycling of bioplastic waste: a review. *Adv. Ind. Eng. Polym. Res.* **4**, 159–177 (2021).
16. Martin, O. & Avérus, L. Poly(lactic acid): plasticization and properties of biodegradable multiphase systems. *Polymer* **42**, 6209–6219 (2001).
17. Overview of materials for polystyrene, extrusion grade. <https://www.matweb.com/search/DataSheet.aspx?MatGUID=1c41e50c2e324e00b0c4e419ca780304> (2011).
18. Kuenneth, C. & Ramprasad, R. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. *Nat. Commun.* **14**, 4099 (2023).
19. Toland, A. et al. Accelerated scheme to predict ring-opening polymerization enthalpy: simulation-experimental data fusion and multitask machine learning. *J. Phys. Chem. A* <https://doi.org/10.1021/acs.jpca.3c05870> (2023).
20. Coates, G. W. & Getzler, Y. D. Y. L. Chemical recycling to monomer for an ideal, circular polymer economy. *Nat. Rev. Mater.* **5**, 501–516 (2020).
21. Zhang, X., Guo, W., Zhang, C. & Zhang, X. A recyclable polyester library from reversible alternating copolymerization of aldehyde and cyclic anhydride. *Nat. Commun.* **14**, 5423 (2023).
22. Phan, H. et al. Functional initiators for the ring-opening polymerization of polyesters and polycarbonates: an overview. *J. Polym. Sci.* **58**, 1911–1923 (2020).
23. Tian, H., Tang, Z., Zhuang, X., Chen, X. & Jing, X. Biodegradable synthetic polymers: preparation, functionalization and biomedical application. *Prog. Polym. Sci.* **37**, 237–280 (2012).
24. Capasso Palmiero, U., Sponchioni, M., Manfredini, N., Maraldi, M. & Moscatelli, D. Strategies to combine ROP with ATRP or RAFT polymerization for the synthesis of biodegradable polymeric nanoparticles for biomedical applications. *Polym. Chem.* **9**, 4084–4099 (2018).
25. Tran, H. et al. Toward recyclable polymers: ring-opening polymerization enthalpy from first-principles. *J. Phys. Chem. Lett.* **13**, 4778–4785 (2022).
26. Pirok, G. et al. Making “real” molecules in virtual space. *J. Chem. Inf. Model.* **46**, 563–568 (2006).
27. Kim, S., Schroeder, C. M. & Jackson, N. E. Open macromolecular genome: generative design of synthetically accessible polymers. *ACS Polym. Au* **3**, 318–330 (2023).
28. Ohno, M., Hayashi, Y., Zhang, Q., Kaneko, Y. & Yoshida, R. SMiPoly: generation of a synthesizable polymer virtual library using rule-based polymerization reactions. *J. Chem. Inf. Model.* **63**, 5539–5548 (2023).
29. Gurnani, R. et al. Ai-assisted discovery of high-temperature dielectrics for energy storage. *Nat. Commun.* <https://doi.org/10.1038/s41467-024-50413-x> (2024).
30. Sterling, T. & Irwin, J. J. ZINC 15 – ligand discovery for everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
31. Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
32. Vwr.eMolecules.com | Chemical Structure Drawing Search - eMolecules. <https://vwr.emolecules.com/index.php> (2014).
33. eMolecules. Buy Research Compounds | Search CAS Number | eMolecules. <https://www.emolecules.com> (2005).
34. Daylight theory: SMARTS - a language for describing molecular patterns. <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> (2011).
35. RDKit: Open-source cheminformatics. <https://doi.org/10.5281/zenodo.10633624> (2023).
36. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1**, 8 (2009).
37. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
38. Kuenneth, C. Ramprasad-Group/canonicalize_psmiles. Ramprasad Group https://github.com/Ramprasad-Group/canonicalize_psmiles (2024).
39. Liu, H. et al. Dynamic remodeling of covalent networks via ring-opening metathesis polymerization. *ACS Macro Lett.* **7**, 933–937 (2018).
40. Cama, G., Mogosanu, D., Houben, A. & Dubruel, P. Synthetic biodegradable medical polyesters. In *Science and Principles of Biodegradable and Bioresorbable Medical Polymers* (ed. Zhang, X.) 82–83 (Elsevier, 2016).
41. Sathe, D. et al. Olefin metathesis-based chemically recyclable polymers enabled by fused-ring monomers. *Nat. Chem.* **13**, 743–750 (2021).
42. Tu, Y.-M. et al. Biobased high-performance aromatic–aliphatic polyesters with complete recyclability. *J. Am. Chem. Soc.* **143**, 20591–20597 (2021).
43. Li, L.-G., Wang, Q.-Y., Zheng, Q.-Y., Du, F.-S. & Li, Z.-C. Tough and thermally recyclable semiaromatic polyesters by ring-opening polymerization of benzo-thia-caprolactones. *Macromolecules* **54**, 6745–6752 (2021).
44. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).

45. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 20 (2015).
46. Chung, T.-S., Cheng, S.-X. & Jaffe, M. Liquid crystalline polymers, main-chain. In *Encyclopedia of Polymer Science and Technology* 3rd edn (ed. Mark, H. F.) (Wiley, 2003).
47. Yu, X. et al. Unraveling substituent effects on the glass transition temperatures of biorenewable polyesters. *Nat. Commun.* **9**, 2880 (2018).
48. Scheelje, F. C. M., Destaso, F. C., Cramail, H. & Meier, M. A. R. Nitrogen-containing polymers derived from terpenes: possibilities and limitations. *Macromol. Chem. Phys.* **224**, 2200403 (2023).
49. van der Sman, R. G. M. Predictions of glass transition temperature for hydrogen bonding biomaterials. *J. Phys. Chem. B.* **117**, 16303–16313 (2013).
50. Ma, X., Zheng, F., van Sittert, C. G. C. E. & Lu, Q. Role of intrinsic factors of polyimides in glass transition temperature: an atomistic investigation. *J. Phys. Chem. B.* **123**, 8569–8579 (2019).
51. Li, W., Ma, J., Wu, S., Zhang, J. & Cheng, J. The effect of hydrogen bond on the thermal and mechanical properties of furan epoxy resins: molecular dynamics simulation study. *Polym. Test.* **101**, 107275 (2021).
52. Huang, X., Nakagawa, S., Houjou, H. & Yoshie, N. Insights into the role of hydrogen bonds on the mechanical properties of polymer networks. *Macromolecules* **54**, 4070–4080 (2021).
53. Sugiyama, F. et al. Effects of flexibility and branching of side chains on the mechanical properties of low-bandgap conjugated polymers. *Polym. Chem.* **9**, 4354–4363 (2018).
54. Cowie, J. M. G., Reid, V. M. C. & McEwen, I. J. Effect of side chain length on the glass transition of copolymers from styrene with *n*-alkyl citraconimides and with *n*-alkyl itaconimides. *Br. Polym. J.* **23**, 353–357 (1990).
55. Zhu, J.-B., Watson, E. M., Tang, J. & Chen, E. Y.-X. A synthetic polymer system with repeatable chemical recyclability. *Science* **360**, 398–403 (2018).
56. Cywar, R. M. et al. Redesigned hybrid nylons with optical clarity and chemical recyclability. *J. Am. Chem. Soc.* **144**, 5366–5376 (2022).
57. Su, Y. et al. Reprocessable and recyclable materials for 3d printing via reversible thia-Michael reactions. *Angew. Chem. Int. Edn.* **64**, 1–6 (2025).
58. Kern, J., Chen, L., Kim, C. & Ramprasad, R. Design of polymers for energy storage capacitors using machine learning and evolutionary algorithms. *J. Mater. Sci.* <https://doi.org/10.1007/s10853-021-06520-x> (2021).
59. Atasi, C., Kern, J. & Ramprasad, R. Design of recyclable plastics with machine learning and genetic algorithm. *J. Chem. Inform. Model.* <https://doi.org/10.1021/acs.jcim.4c01530> (2024).
60. Kuenneth, C., Schertzer, W. & Ramprasad, R. Copolymer informatics with multitask deep neural networks. *Macromolecules* **54**, 5957–5961 (2021).
61. Shukla, S. S., Kuenneth, C. & Ramprasad, R. Polymer informatics beyond homopolymers. *MRS Bull.* **49**, 17–24 (2024).
62. Illy, N., Puchelle, V., Le Luyer, S. & Guégan, P. Alternating copolymerization of bio-based *N*-acetylhomocysteine thiolactone and epoxides. *Eur. Polym. J.* **153**, 110490 (2021).
63. Kumar, K. & Goodwin, A. P. Alternating sulfone copolymers depolymerize in response to both chemical and mechanical stimuli. *ACS Macro Lett.* **4**, 907–911 (2015).
64. Takebayashi, K., Kanazawa, A. & Aoshima, S. Cationic ring-opening copolymerization of a cyclic acetal and γ -butyrolactone: Monomer sequence transformation and polymerization–depolymerization control by vacuuming or temperature changes. *Polym. J.* <https://www.nature.com/articles/s41428-023-00847-9> (2023).
65. Kim, C., Batra, R., Chen, L., Tran, H. & Ramprasad, R. Polymer design using genetic algorithm and machine learning. *Comput. Mater. Sci.* **186**, 110067 (2021).
66. Venkatram, S. et al. Predicting crystallization tendency of polymers using multifidelity information fusion and machine learning. *J. Phys. Chem. B.* **124**, 6046–6054 (2020).
67. Patra, A. et al. A multi-fidelity information-fusion approach to machine learn and predict polymer bandgap. *Comput. Mater. Sci.* **172**, 109286 (2020).
68. Batra, R., Paliana, G., Uberuaga, B. P. & Ramprasad, R. Multifidelity information fusion with machine learning: a case study of dopant formation energies in hafnia. *ACS Appl. Mater. Interfaces* **11**, 24906–24918 (2019).
69. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer genome: a data-powered polymer informatics platform for property predictions. *J. Phys. Chem. C.* **122**, 17575–17585 (2018).
70. Doan Tran, H. et al. Machine-learning predictions of polymer properties with Polymer Genome. *J. Appl. Phys.* **128**, 171104 (2020).
71. Kern, J., Venkatram, S., Banerjee, M., Brettmann, B. & Ramprasad, R. Solvent selection for polymers enabled by generalized chemical fingerprinting and machine learning. *Phys. Chem. Chem. Phys.* **24**, 26547–26555 (2022).
72. Brandrup, J., Immergut, E. H. & McDowell, W. *Polymer Handbook* 2nd edn (Wiley, New York, 1975).
73. Pfäffli, P., Zitting, A. & Vainio, H. Thermal degradation products of homopolymer polystyrene in air. *Scand. J. Work, Environ. Health* **4 Suppl 2**, 22–27 (1978).

Acknowledgements

This work is supported by the Office of Naval Research through a Multidisciplinary University Research Initiative (MURI) Grant (N00014-20-1-2586). J.K. gratefully acknowledges support through the National Defense Science and Engineering (NDSEG) Fellowship Program from the Department of Defense (DoD). J.K. is grateful to Aubrey Toland for his support on the paper.

Author contributions

J.K. wrote the original draft of the manuscript, developed the software, and performed the digital analyses. Y.S. designed and performed the physical experiments, reviewed candidate polymers for synthesizability, and contributed figures and text for the Supplementary Section. W.G. assisted with the physical experiment procedures and reviewed candidate polymers for synthesizability. R.R. contributed to the theoretical framework and result analysis. All authors reviewed and edited the manuscript.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01683-6>.

Correspondence and requests for materials should be addressed to Rampi Ramprasad.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025